

Ab initio simulations of the degradation of biodegradable
batteries

Benjamin Deacon

First Supervisor: Dr Samuel Murphy

Second Supervisor: Dr John Hardy

Lancaster University, Bailrigg, Lancashire, LA1 4YW

This thesis is submitted in partial fulfilment
of the requirements for the degree of MSc by Research

December 2020

Declaration of Authorship

I, Benjamin Deacon, declare that this thesis titled 'Ab initio simulations of the degradation of biodegradable batteries' and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree while at Lancaster University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations this thesis is entirely my own work.
- I have acknowledged all main sources of help
- Where the thesis is based on work done by myself jointly with others, I have made clear the individual contributions.

Signed: 

Date: 4.1.2021

Abstract

Transient implantable medical bionics (TIMBs), such as biodegradable batteries that disappear after their operation are gaining attention. They potentially facilitate the deployment of novel instructive biomaterials for regenerative medicine. Implantable, biodegradable and biocompatible batteries may be capable of satisfying the power requirements of some biomedical devices before harmlessly degrading. One material of particular interest for the construction of biodegradable batteries is Bombyx Mori silk. Lancaster University is developing a biodegradable battery that will utilise silk both in the electrolyte and to encase the battery. Using the silk offers the battery a degree of protection that enables the device to operate for several days before it harmlessly degrades. Key to tuning the lifetime of the battery is understanding how the structure of the silk changes under different operating conditions and how this changes the diffusivity of the cations (i.e. Mg^{2+}) and other species such as choline nitrate used as the ionic liquid in the electrolyte. This project will aim to further this understanding through the use of quantum mechanical methods.

This project quantifies the behaviour of various molecules in the presence of SF, including water, choline and Mg ions. This helps to see how the biocompatible and biodegradable batteries will behave when made from SF. This is completed via DFT simulation as to perform the experiment is unfeasible. For example, the diffusion pathway of water can not be experimentally generated. Furthermore this project has generated ramachandran plots via DFT for silk fibroin which have not been carried out on this material previously. This allows for a detailed comparison with classical mechanical data.

This understanding will allow for further work to elucidate and exploit the properties of SF. Further understanding will allow for fine-tuning of how long the SF biodegradable battery will take to break down; this can be changed for the required use. This will help to understand the ions contribution to the effect on the decay rate of the electrode.

Acknowledgements

I want to extend my special thanks to Dr Samuel Murphy and Dr John Hardy for the supervision and provided support throughout this project. To Mathew Haskew who, provided constant support, help with errors and difficulties that arise with a project. This project would not have been possible without the use of ARCHER and HEC computational facilities, and therefore, I would extend my gratitude to Lancaster Physics Department and the ARCHER community. Moreover, finally to any other lecturer who has offered guidance and help throughout, especially Dr Andrew Kerridge, Dr Michael Peach and Dr Marcella Iannuzzi, in a range of ways from checking input scripts to computational advice.

Publication

This work has been submitted in part for publication as part of the paper "Evaluating a Bombyx mori silk fibroin model using a range of atomistic simulation tools". Mathew Haskeew is the first author, and Benjamin Deacon is the second author on account of the provision of DFT data. At the end of this thesis is the paper.

Copyright

For all reproduced pictures copyright has been gained where required. For this thesis the only image that has required a license is Figure 1, all other reproduced images are allowed without a license due to how they are published. The copyright license has been included in the supporting information.

List of Figures

1	An image showing a voltaic pile battery, taken from [18]. The copyright license has been included in the supporting evidence.	8
2	An image showing the standard battery structure, taken from [21].	9
3	An image showing the suitable battery materials, taken from [21].	10
4	An image showing a schematic of an SF battery. a) View x of the battery. b) View z of the battery. c) a labelled diagram of the SF battery. Produced by Mathew Haskew (unpublished).	13
5	An image showing the simulated unit in SF, the green cell indicates glycine and the red cell shows alanine. Created by the Author using Ovito [42]. . . .	14
6	An image to show the difference in silk structure. Created by the Author. a) shows the simplified silk structure. b) shows the complex silk structure. . . .	15
7	An image showing the antiparallel and parallel β sheets in the schematic form [50]. a) Shows the parallel β sheets. b) Shows the antiparallel β sheets. Made by the Author.	18
8	Figure showing the IR spectrum of natural spun SF, Glycine and Alanine, reproduced by the Author using data from [57].	19
9	Figure showing a torsional angle, ϕ , in a silk chain. Created by the Author.	20
10	Ramachandran plot of the allowed torsion angles of alanine in SF, as determined by solid-state NMR. Taken from [61].	21
11	Ramachandran plot of the allowed torsion angles of glycine in SF, as determined by solid-state NMR. Taken from [61].	21
12	Ramachandran plot for alanine molecule, taken from [60].	21
13	Ramachandran plot for glycine molecule, taken from [60].	22
14	Ramachandran plot for serine molecule, taken from [60].	22
15	Ramachandran plot for alanine molecule, taken from [59]. Where: α -helix (α), 310-helix (3), π -helix (π), left-handed α -helix (α L), 2.27 ribbon (2), polyproline-ii (II), collagen (C), parallel β -sheet ($\uparrow\uparrow$), and antiparallel β -sheet ($\uparrow\downarrow$). Taken from [60].	23
16	An image to show the comparison of two Nyquist plots of two different ions in the Silk cell. Where a is the Nyquist plot for the magnesium ion taken from [5], and b is the Nyquist plot for the lithium-ion taken from [37].	25

17	An image to show the change in hydrogen bonds for D3 and no D3 calculation in alanine. Created by the Author using Vesta [79]. a) shows two hydrogen bonds formed in the D3 calculation and b) shows the singular hydrogen bond in the no D3 calculation.	32
18	An image to show the label nomenclature for alanine. Created by the Author using Vesta [79].	36
19	An image to show the label nomenclature for glycine. Created by the Author using Vesta [79].	36
20	An image to show how a basis set is used. Taken from [84].	37
21	An image showing the SF unit-cell. Created by the Author using Ovito [42].	40
22	An image showing the box size comparison between the $1 \times 1 \times 1$ cell (shown in pink) and $2 \times 2 \times 2$ cell (shown in black). Created by the Author using Ovito [42].	41
23	A graph to show the energy convergence of 3000 steps of a cell optimisation run.	44
24	A graph to show the number of steps completed by NVT and NPT at a range of temperatures.	46
25	An image to show the change in silk structure when different basis sets are used. Created by the Author using Vesta [79]. a) shows the 6-31G basis set, b) shows DZVP and c) shows TZVP.	52
26	An image to show the change in silk structure after undergoing 3000 cell optimisation steps. Created by the Author using Vesta [79] and Ovito [42]. Where a and b shows the change in box size with a) being pre-cell optimisation and b) being post cell optimisation. C and d show the change in hydrogen bonding before and after cell optimisation with c) being pre-cell optimisation and d) being post cell optimisation.	53
27	An image to show the change in silk structure after undergoing 20000 MD steps, both taken from the z view. Created by the Author using Vesta [79]. Where a) shows the silk cell pre-MD and b) shows the silk cell post MD.	54
28	A figure showing the labelled areas of an empty Ramachandran plot. Created by the Author using VMD [101].	55

29	An image to show the Ramachandran plots at 10 K. Created by the Author using VMD [101]. a) Hydrated cell run at 10 K with NPT. b) Non-Hydrated cell run at 10 K with NPT. c) Hydrated cell run at 10 K with NVT. d) Non-Hydrated cell run at 10 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).	56
30	An image to show the Ramachandran plots at 298 K. Created by the Author using VMD [101]. a) Hydrated cell run at 298 K with NPT. b) Non-Hydrated cell run at 298 K with NPT. c) Hydrated cell run at 298 K with NVT. d) Non-Hydrated cell run at 298 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).	56
31	An image to show the Ramachandran plots at 310 K. Created by the Author using VMD [101]. a) Hydrated cell run at 310 K with NPT. b) Non-Hydrated cell run at 310 K with NPT. c) Hydrated cell run at 310 K with NVT. d) Non-Hydrated cell run at 310 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).	57
32	An image to show the Ramachandran plots at 673 K. Created by the Author using VMD [101]. a) Hydrated cell run at 673 K with NPT. b) Non-Hydrated cell run at 673 K with NPT. c) Hydrated cell run at 673 K with NVT. d) Non-Hydrated cell run at 673 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).	57
33	An image to show the Ramachandran plots at 310 K in the $2 \times 2 \times 2$ cell. a) Hydrated cell run at 310 K with NPT $2 \times 2 \times 2$ cell compared to DFT. b) Hydrated cell run at 310 K with NVT $2 \times 2 \times 2$ cell compared to DFT. c) Non-Hydrated cell run at 310 K with NPT $2 \times 2 \times 2$ cell compared to DFT. d) Non-Hydrated cell run at 310 K with NVT $2 \times 2 \times 2$ cell compared to DFT. The black dots show the DFT data points, and the contour plot shows the classical mechanical data.	59
34	An image to show the Ramachandran plots at 310 K in the $4 \times 4 \times 4$ cell. a) Hydrated cell run at 310 K with NPT $4 \times 4 \times 4$ cell compared to DFT. b) Hydrated cell run at 310 K with NVT $4 \times 4 \times 4$ cell compared to DFT. c) Non-Hydrated cell run at 310 K with NPT $4 \times 4 \times 4$ cell compared to DFT. d) Non-Hydrated cell run at 310 K with NVT $4 \times 4 \times 4$ cell compared to DFT. The black dots show the DFT data points, and the contour plot shows the classical mechanical data.	59

35	A graph to show the time dependency of the thermal expansion at 150 K of the hydrated system, the last 1500 frames are shown.	61
36	Images to show the thermal expansion of a hydrated cell. a) A graph to show the change in x vector of a hydrated cell at various temperatures. b) A graph to show the change in y vector of a hydrated cell at various temperatures. c) A graph to show the change in z vector of a hydrated cell at various temperatures. d) A graph to show the change in volume of a hydrated cell at various temperatures.	62
37	Images to show the thermal expansion of a non-hydrated cell. a) A graph to show the change in x vector of a non-hydrated cell at various temperatures. b) A graph to show the change in y vector of a non-hydrated cell at various temperatures. c) A graph to show the change in z vector of a non-hydrated cell at various temperatures. d) A graph to show the change in volume of a non-hydrated cell at various temperatures.	63
38	A figure showing the generated IR spectra of simplified SF, via DFT using geometry optimisation. Produced using Molden [94].	65
39	A figure showing the generated IR spectra of alanine. Produced using Molden [94].	65
40	A figure showing the generated IR spectra of glycine. Produced using Molden [94].	66
41	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 10 K. Produced using Molden [94].	67
42	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 150 K. Produced using Molden [94].	68
43	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 298 K. Produced using Molden [94].	68
44	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 310 K. Produced using Molden [94].	69
45	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 673 K. Produced using Molden [94].	69
46	An image showing the various water positions.	71
47	An image to show the most and least favourable water positions. Created by the Author using Vesta [79]. a) Most favourable position for water - W_6 . b) Least favourable position for water - W_5	72

48	A figure showing the water diffusion in the x-direction. Hydrogen atoms have been omitted.	75
49	A figure showing the water diffusion in the y-direction. Hydrogen atoms have been omitted.	75
50	A figure showing the water diffusion in the z-direction. Hydrogen atoms have been omitted.	76
51	An image showing the various magnesium positions.	77
52	An image to show the most and least favourable magnesium positions for the hydrated system. Created by the Author using Vesta [79]. a) Most favourable position for magnesium - M ₇ . b) Least favourable position for magnesium - M ₁₀	79
53	An image to show the most and least favourable magnesium positions for the non-hydrated system. Created by the Author using Vesta [79]. a) Most favourable position for magnesium - M ₁₀ . b) Least favourable position for magnesium - M ₁	79
54	A graph showing the comparison of the energy of the hydrated and non-hydrated system.	81
55	A figure showing the water and Mg diffusion in the x-direction. Hydrogen atoms have been omitted.	82
56	A figure showing the water and Mg diffusion in the y-direction. Hydrogen atoms have been omitted.	83
57	A figure showing the water and Mg diffusion in the z-direction. Hydrogen atoms have been omitted.	83
58	A figure showing the electron density of the charged cell.	85
59	A figure showing the electron density map of the charged cell. Where the key is shown on the left-hand side, the colours are arbitrary, with red showing a greater area of electron density and the blue showing an area of low electron density.	85
60	A figure showing the choline structure	86
61	A figure showing the simulated choline nitrate structure; this is not the experimental structure of choline nitrate.	87
62	A figure showing the choline nitrate in the xy plane.	88
63	A figure showing the choline nitrate in the zx plane.	89
64	A figure showing the first frame of the full simulation.	92

65	A figure showing the last frame of the full simulation.	93
66	An overview of the number of jobs run on ARCHER.	95
67	An image to show the size difference in silk cells. Created by the Author using Vesta [79].	100
68	An image to show the Ramachandran plots at 50K. Created by the Author using VMD [101]. a) Hydrated cell run at 50K with NPT. b) Non-Hydrated cell run at 50K with NPT. c) Hydrated cell run at 50K with NVT. d) Non-Hydrated cell run at 50K with NVT.	124
69	An image to show the Ramachandran plots at 150K. Created by the Author using VMD [101]. a) Hydrated cell run at 150K with NPT. b) Non-Hydrated cell run at 150K with NPT. c) Hydrated cell run at 150K with NVT. d) Non-Hydrated cell run at 150K with NVT.	125
70	An image to show the Ramachandran plots at 273K. Created by the Author using VMD [101]. a) Hydrated cell run at 673K with NPT. b) Non-Hydrated cell run at 673K with NPT. c) Hydrated cell run at 673K with NVT. d) Non-Hydrated cell run at 673K with NVT.	126
71	An image to show the Ramachandran plots at 373K. Created by the Author using VMD [101]. a) Hydrated cell run at 373K with NPT. b) Non-Hydrated cell run at 373K with NPT. c) Hydrated cell run at 373K with NVT. d) Non-Hydrated cell run at 373K with NVT.	127
72	An image to show the Ramachandran plots at 473K. Created by the Author using VMD [101]. a) Hydrated cell run at 473K with NPT. b) Non-Hydrated cell run at 473K with NPT. c) Hydrated cell run at 473K with NVT. d) Non-Hydrated cell run at 473K with NVT.	128
73	An image to show the Ramachandran plots at 500K. Created by the Author using VMD [101]. a) Hydrated cell run at 500K with NPT. b) Non-Hydrated cell run at 500K with NPT. c) Hydrated cell run at 500K with NVT. d) Non-Hydrated cell run at 500K with NVT.	129
74	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 50 K. Produced using Molden [94].	130
75	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 273 K. Produced using Molden [94].	130
76	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 373 K. Produced using Molden [94].	131

77	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 473 K. Produced using Molden [94].	131
78	A figure showing the generated IR spectra of simplified SF, via DFT using MD at 500 K. Produced using Molden [94].	132

List of Tables

1	Table showing the difference in length of the hydrogen bonding in alanine. . .	31
2	Table showing the number of steps produced for various functionals.	34
3	Table showing the comparison of alanine bond lengths to published data (all numbers are rounded to 4dp).	34
4	Table showing the comparison of glycine bond lengths to Published data (all numbers are rounded to 4dp).	34
5	Table showing the comparison of single-point energy calculations to experimental values.	35
6	Table showing the number of steps produced for various basis sets.	38
7	Table showing the comparison of single-point energy calculations to experimental values for various basis sets.	38
8	Table showing the various cell sizes considered for this project	40
9	Table showing the computational cost for various cell sizes.	41
10	Table showing the comparison of choline bond lengths to Published data data (all numbers are rounded to 4dp). Hydrogen bonds have been omitted. . . .	42
11	Table showing the computational cost for cell optimisation vs MD.	43
12	Table showing the computational cost for various temperatures.	46
13	Table showing the lattice parameters for the SF as both hydrated and non-hydrated cells.	53
14	Table showing the cell volume at various temperatures.	62
15	Table showing the thermal expansion coefficients at various temperatures. . .	64
16	Table showing the peak assignment for the SF spectra. Not all peaks will have a secondary structure associated with it.	66
17	Table showing the water energy in various positions.	71
18	Table showing the water energy in varying amounts.	73
19	Table showing the water displacement in the one water molecule system. . . .	73
20	Table showing the water displacement in the eight water molecule system. . .	73
21	Table showing the water displacement as a sum of the individual vectors in the eight water molecule system.	74
22	Table showing the magnesium energy in various positions for the non-hydrated cell.	78
23	Table showing the magnesium energy in various positions for the hydrated cell. .	78

24	Table showing the magnesium energy in varying amounts for the hydrated system.	80
25	Table showing the magnesium energy in varying amounts for the non-hydrated system.	80
26	Table showing the water displacement in the charged system.	82
27	Table showing the magnesium displacement in the charged system.	84
28	Table showing the bond lengths of choline. Hydrogen bonds have been omitted.	87
29	Table showing the bond lengths of choline nitrate. Hydrogen bonds have been omitted.	88
30	Table showing the water displacement in the full system.	91
31	Table showing the magnesium displacement in the full system.	91

Contents

Declaration of Authorship	i
Abstract	ii
Acknowledgements	iii
Publication	iv
Copyright	iv
List of Figures	v
List of Tables	xii
Contents	xiv
1 Introduction	1
2 Literature Review	3
2.1 Transient Implantable Medical Bionics (TIMBs)	3
2.1.1 Biocompatibility	3
2.1.2 Cost	4
2.1.3 Detection or Wireless Communication	4
2.1.4 Power Management	5
2.1.5 Size	5
2.1.6 Waste	6
2.2 Batteries	7
2.2.1 History of Batteries	7
2.2.2 Current Technology	8
2.2.3 Future of Batteries	11
2.2.4 Polymers in Batteries	11
2.2.5 Biocompatible and Biodegradable Batteries	13
2.3 Silk	13
2.3.1 Composition of Silk	15
2.3.2 Secondary Structure	17
2.3.3 SF Battery	23

2.3.4	SF Computational Simulations	25
3	Methodology	27
3.1	Classical vs Quantum	27
3.2	DFT+D3	28
3.3	Functionals	32
3.4	Basis Sets	36
3.5	Cell Generation	39
3.6	Calculation Methods	42
3.6.1	Optimisation	43
3.6.2	Molecular Dynamics	44
3.6.3	Vibrational Analysis	47
3.7	Ramachandran Plots	48
3.8	Software	49
4	Results and Discussion	51
4.1	Preliminary Investigation of the SF Structure	51
4.2	Ramachandran Plots	54
4.3	Thermal Expansion	60
4.4	Spectra	64
4.5	Behaviour of Water	70
4.6	Behaviour of Magnesium	77
4.6.1	Electron Density	84
4.7	Behaviour of Choline	86
4.8	Behaviour of the complete cell	90
5	Conclusions	94
5.1	Is the generated structure valid?	94
5.2	Is the computational time enough for all calculations to be successfully run?	94
5.3	How do various species behave in SF?	96
5.4	How does the structure of SF behave when undergoing vibrational excitement?	96
5.5	How does the SF material behave under varying temperatures?	97
5.6	How does DFT data compare to classical mechanics and experimental data?	97
6	Future Work	99
	Supporting Information	102

References **133**

Evaluating a Bombyx mori silk fibroin model using a range of atomistic simulation tools **144**

1 Introduction

This project aims to simulate silk fibroin (SF), from the *Bombyx Mori* (*B. Mori*) silkworm, using density functional theory (DFT) [1]. The focus is on SF for use in biodegradable and biocompatible batteries. Where biodegradable batteries will break down to their, ideally and in the case of SF, harmless constituent molecules, this is of importance for the future as it reduces waste, reduces the required number of operations and has a wide range of uses. It is a relatively new material for use in batteries. Few studies have been performed on the material and even fewer on the inclusion of SF in batteries with a view of how other molecules affect the battery. This material is of increasing importance due to its biocompatibility and ability to degrade to harmless substances [2–4]. Additionally, SF has shown to allow diffusion of species through matrices composed of SF [5]. Which makes SF an attractive material for use in biological batteries or more general use in TIMBs. This topic will be discussed in further detail in the literature review.

The overall aim of this project is to simulate a cell that replicates the experimental environment by using a combination of MD and DFT computational techniques, with this project focusing on the DFT calculations. The thesis aims to answer the following questions: How do various molecules behave in SF? How does DFT data compare to classical mechanics and experimental data? Is the generated structure valid? Is the computational time enough for all calculations to be successfully run?

This thesis is organised in the following chapters: Chapter 2 Literature review: An overview of past and future battery technologies and materials, A detailed look into silk with respect to structure of silk, different experiments performed and previous simulations in both classical and quantum mechanics. Chapter 3 Methodology: This section gives a detailed discussion of how the project is to be carried out, it contains computational details of how to perform the required experiments with the given rationale. Chapter 4 Results and Discussions: The results are compiled and discussed with relevance to experimentally published data and classical mechanics data. Chapter 5 Conclusions: This area discusses the impact of this project and the success of the simulations. There are a range of conclusions drawn considering the structure and the cost of these simulations. Chapter 6 Future Work: Ideas for the future of this project have been presented with a focus of further developing the simulations via improving the size of the cell and altering the structure simulated.

Throughout this thesis, the colour scheme for the pictures shall be the same: carbon - brown, oxygen - red, nitrogen - blue, hydrogen - white, magnesium - orange. This project has utilised a large variety of applications for molecule creation, cell visualisation, data processing and simulations. For this project 2500 kAUs of ARCHER time was awarded, the application is in the supporting information. This project will also use the HEC resource provided by Lancaster University Physics.

2 Literature Review

This literature review will investigate and discuss previous and current research in multiple different areas, giving both context, knowledge and a comprehensive basis for this project. The literature review will focus on areas relevant to the use of SF in biodegradable, biocompatible batteries. Such areas of interest include Transient Implantable Medical Bionics, batteries, and silk.

2.1 Transient Implantable Medical Bionics (TIMBs)

TIMBs are devices that can be implanted into the body to achieve a specific function, such as pacemakers, cochlear implants, and blood pressure measurement devices. These materials pave the way for the development of novel biomaterials for regenerative medicine. For more widespread adoption of these materials, there are several issues to overcome, for this specific project biocompatibility and power management are the main issues as discussed below.

2.1.1 Biocompatibility

The biocompatibility of a device depends on the material, and any bacteria on the surface. The standard definition of biocompatibility is taken as the ability of a material to perform with an appropriate host response in a specific situation. The material must be accepted into the body; otherwise, the body and internal systems will fight to get rid of the foreign body. The bacteria on the surface may cause infections or adverse reactions if the bacteria are not generally in the body. There are multiple methods to overcome this bacteria; this is generally completed via sterilisation. Methods for this include heat treatment, pressured vapour sterilisation, radiation and chemical sterilisation - ethylene oxide (EtO), formaldehyde, (H_2O_2) and peracetic acid [6]. There are many biocompatible materials which include titanium and its alloys, some cobalt-based alloys and biocompatible polymers. This is where the choice of silk for this project becomes essential as it is biocompatible due to the monomers in the structure; this is in more detail in the silk section of this literature review.

The biocompatibility of the products has to be considered due to changes in behaviour when the original material degrades [7]. While silk may be biocompatible, the product may be harmful to the body or the environment. The ideal materials to be produced from degradation are carbon dioxide, water vapour, and organic material, which is not harmful to the environment. While this is ideal, it may not always be the case, many plastics (PVC and PTFE) are safe to touch and handle but when burnt will produce chlorine and fluorine which is toxic to humans. In the case of biocompatible materials, the toxicity of the product is the

amount of product present, and it may be possible for small amounts of a toxic product to be present, for example, CO may be produced below 50 ppm in 24 hours and still be safe. SF is biocompatible and has been shown in multiple papers evaluated in terms of cytotoxicity, intracutaneous reactivity, sensitization, acute systemic toxicity, subchronic systemic toxicity, genotoxicity, and hemolysis. This allows for SF to be used in multiple applications including as part of an SF battery.

2.1.2 Cost

The cost of TIMBS relies on the availability of the materials and how easy they are to fabricate. However, for materials to be widely accepted, a few considerations are the cost of TIMBs, how often TIMBs need replacing and the cost of operations with/without TIMBSs. As with the majority of new and specialised materials, the initial cost will be relatively high; however, this will reduce with time as both technology and understanding of the material is improved. Several competitor materials can be used in implants, common metals that are readily available and therefore cheaper are still regularly used, these materials include stainless steel, cobalt-chrome alloy, titanium, and nickel-titanium alloy. The metals used in implants are long-lasting, biocompatible and non-biodegradable, this will mean that a second operation is required to remove the implant; this would not occur in TIMBs. The batteries currently used in implants will tend to consist of lithium metal anodes with cathode systems including iodine, again, this is long-lasting, cheaper materials and will not breakdown [8]. SF for experiments have a high cost at this time, £200 for 20 ml from Sigma Aldrich, this is relatively high but will decrease as a better understanding of the material is gained.

2.1.3 Detection or Wireless Communication

There are several detection and communication methods for implantable bionics. This need for communication with devices inside the body arises from needing to monitor the devices and the body [9]. This requirement is related to several parameters such as rate of drug delivery, or investigation of the condition of the organ and if the condition is becoming better or worse, close monitoring will then allow further steps to be taken at an earlier stage improving future quality of life. There are various issues with communicating with devices in the body, including distance to the detector, transmission rate and frequency of the device. Communication is a developing field, one type of communication, as proposed in [9], which uses a different frequency of 402-405 MHz to increase the distance and data transfer. A second type uses a 433 MHz separate RF link which can transmit up to 9 cm, as opposed to

6 cm in the previous method [10].

2.1.4 Power Management

There are two types of implantable batteries: single-use and rechargeable. Single-use batteries require further operation at the end of the battery life to replace the battery whereas rechargeable batteries can recharge without the need for a second operation. The requirement for an operation will be dependent on if the battery is degradable or not. The rechargeable batteries may require an external battery pack to provide the magnetic field to recharge the internal battery [6]. Rechargeable batteries will only work near the surface of the body; otherwise, the magnetic field will not be strong enough to induce a current in the internal battery. For example, a pacemaker will require a single-use battery, this has a lifespan of around seven years making this more reliable than rechargeable batteries, and due to the position of the implant, it would be very inefficient and difficult to recharge. One issue with power management is the size and capacity of energy storage. A battery must last a sufficient amount of time to complete the job required or long enough not to require invasive surgery regularly. Multiple new methods are being investigated for rechargeable batteries. An example of this is using light and photovoltaic converters to generate the necessary power for the battery [11]. Other technologies of interest are devices which will be self-charged by the energy harvested directly from the human body, and this can be done in a multitude of ways such as using thermal energy and kinetic energy from the movement of the blood or movement of body parts [12, 13].

2.1.5 Size

A limiting factor for the development of TIMBs has been the size of available technology. The size and weight of an implant is an essential factor due to implant cost and the comfort of patients. A large implant under the skin may be uncomfortable to have for extended periods and may reduce the quality of life for the patient - for example, the patient would not be able to partake in particular sports or activities. The size of technology has been decreased dramatically due to improvements in microelectronics, biotechnology, and material science, this has allowed for the continued development of smaller and smaller devices for a broader range of uses [6, 13]. Batteries have a significant effect on the size of the implant; typically, a single-use battery will be larger as will require more charge and power. There has been an investigation into battery-less implants which would decrease the size and weight of the implant. Currently, there have been tests on animals for battery-less pressure sensors which

shows that the technology, in the first trials, is successful and can be developed further for use in humans in the future [14]. Another form of battery-less implants use an outside energy source such as an inductive link, or from the neighbouring environment, however, this has shown to be inefficient with power loss over the inductive link, as technology improves a higher efficiency will enable smaller models [15].

The size of the material will affect the biocompatibility of the decay products. A smaller implant may have a smaller effect due to the material being in a lower concentration. Where a material may become harmful above a certain threshold, e.g. formation of small chain carboxylic acids can lead to increased pH in sufficient concentrations. Keeping the size as small as reasonably possible will decrease the possibility of any issues with concentration [7]. SF can be made in a range of sizes, depending on the application. The smallest published SF structure is quoted at 70-100 μm for a circular structure. This would be suitable for use in the body as a biological battery.

2.1.6 Waste

For the majority of TIMBs, there will either be surgery for removal of the device or the device will be permanent and left in the body. Due to the nature of the devices, these will contribute to WEEE (waste electrical and electronic equipment). This waste has to be dealt with in specialised facilities. This is reduced with biodegradable devices where the body breaks down the device; this can also be completed via natural degradation. This makes for more environmentally friendly implants [16].

In a standard operation using permanent materials, a second operation will remove the device, with the waste of the device, there is also secondary waste associated with standard procedure. This would include gloves, face masks and disposable tools used in operation. Biodegradable materials remove the need for a second operation and the waste that is associated with it.

The waste generated from the TIMBs breaking down will heavily depend on the material chosen. For example, poly(caprolactone) can be used in implantable devices where degradation has been studied [17]. This reaction has shown to occasionally produce highly reactive radicals that will then interact immediately with biological systems. However, it will, in general, produce oligomers and water, which are harmless to the body (and the reason it is FDA approved for a variety of medical applications). An SF battery will break down to constituent molecules which are biocompatible. This means that the body will be able to remove all the molecules and reduce waste going to landfill.

2.2 Batteries

The scientific field of batteries encompasses a wide range of types of batteries, different materials, electrochemical properties and varying uses of batteries. The two types of batteries are primary batteries which are non-rechargeable and secondary batteries which are rechargeable.

2.2.1 History of Batteries

Alessandro Volta first invented batteries in 1799 [18]. The original batteries consisted of discs of two different metals, such as copper and zinc, separated by cardboard soaked in brine. This type of battery was known as the voltaic pile; shown in figure 1. This type of battery had an issue where the metal would corrode, and scientific understanding at the time could not account for this fault. In 1833 the Daniell cell was invented [19], which solved corrosion issues and a hydrogen issue in which bubbles form at interfaces. These types of batteries were regularly improved over the next 100 years as scientific knowledge improved to give a better understanding of the shortcomings of the previous version. Up to 1850 batteries were single-use and could not be recharged; however, with the invention of the lead-acid cell in 1859, rechargeable batteries were made [19]. In the turn of the century, Thomas Edison worked on making cheap and widespread iron-nickel batteries aimed at being used for electric vehicles, while this did not manifest, he did contribute batteries to a wide range of other applications such for trains, backup power for signals and lamps. In 1912 lithium-ion batteries were first experimented with, creating one of the most common battery types of the present day. The lithium-ion battery has undergone drastic changes since 1912 to become the batteries that are recognised today. This occurred in 1997, where pyrolysed silicon-containing polymers were used as high capacity anodes to dramatically reduce the size and increase the specific capacity [20].

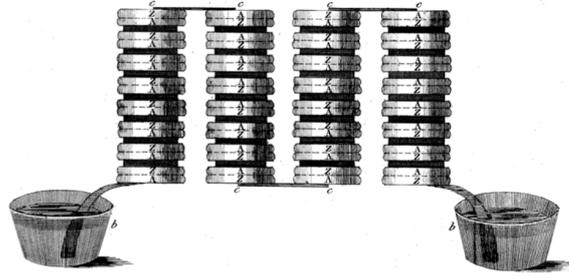


Figure 1: An image showing a voltaic pile battery, taken from [18]. The copyright license has been included in the supporting evidence.

2.2.2 Current Technology

Currently, batteries are used for a wide range of applications utilising various materials and different types of battery. A standard battery consists of an anode, cathode, the current collectors this sits on, electrolyte and separator, some types of battery will vary from this, but in general, this form is correct for modern batteries, this can be seen in figure 2 [21].

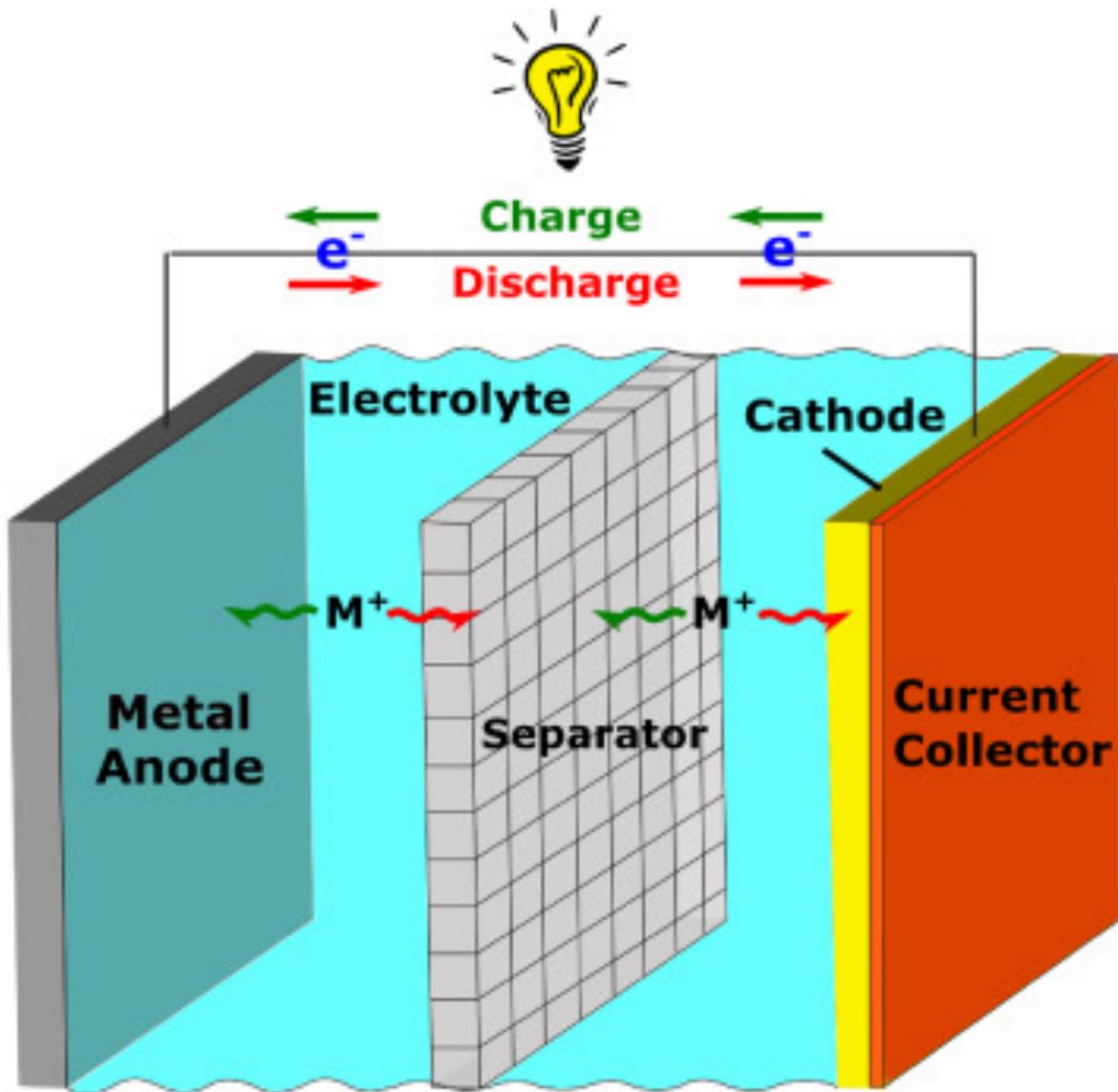


Figure 2: An image showing the standard battery structure, taken from [21].

There are two main types of batteries available; within each category, there are multiple types of battery. Primary batteries include:

- Alkaline batteries
- Coin cell batteries

For secondary batteries:

- Lead- Acid batteries
- Ni-Cd batteries
- Ni-MH batteries
- Li-ion batteries

- Li-Po batteries

These lists are not exhaustive but provide a range of examples. Each type of battery has its' advantages and disadvantages that are not associated with the battery being primary or secondary. For example, Li-ion and Li-Po batteries can be explosive if set up incorrectly, whereas Ni-Cd batteries are non-explosive with a simple installation method, but Ni-Cd batteries do contain toxic material. One of the most common battery materials currently is lithium. Lithium has many advantages that other materials do not have, such as lightweight, very high power density and high voltage production. Primary batteries are, in general, made of zinc, manganese dioxide and potassium hydroxide. An analysis of the periodic table and how suitable each element is shown in figure 3.

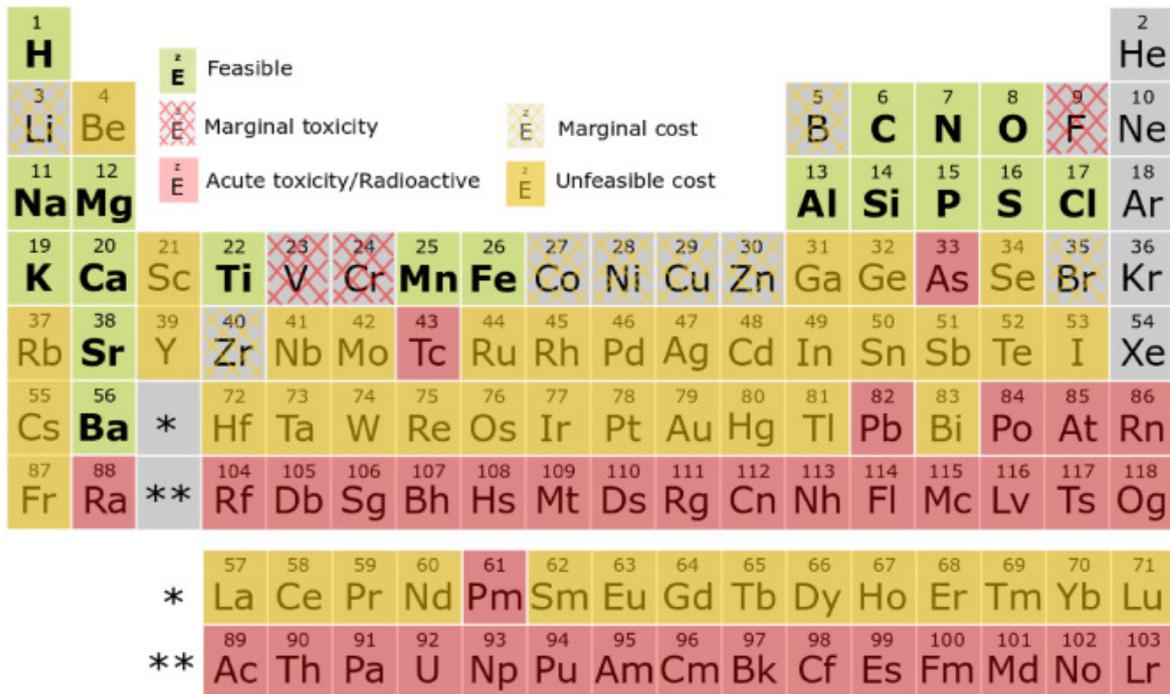


Figure 3: An image showing the suitable battery materials, taken from [21].

There is a wide range of applications for different batteries. In general, primary batteries are used for everyday objects and small appliances that require a low voltage to run. Typical uses for primary batteries include torches, remotes, wall clocks, small portable gadgets, watches and other miniature electronic products. These require small batteries with low voltages, and have a long battery life. Secondary batteries have a wide range of uses both in small and large appliances. Lead-acid batteries are used in electric vehicles; however, they are relatively large and have a low power density.

2.2.3 Future of Batteries

The field of batteries is a continually evolving subject; the future of batteries will be heavily linked with environmental concerns, safety and cost. With constant research into new materials for both the electrodes and the electrolyte, which results in new types of batteries. The main aims of the future are to reduce the cost to make the batteries - either through cheaper materials or simpler production techniques, increased capacity and higher voltages. One suggested improvement would be the removal of cobalt; cobalt is a rare metal and expensive which is present in cathodes in lithium-ion batteries, the removal will make the batteries cheaper and produce a higher energy density [22]. Another material of scientific interest is the use of silicon in place of graphite for the anode. Silicon has ten times the capacity of graphite and will lead to longer-lasting batteries [23]. The use of silicon in [23] can be sustainably produced and environmentally friendly from using barley husks, whereas graphite is mined by open pit and underground methods, which is more harmful to the environment [24]. Faster charging is also of interest as this removes the need for larger and more powerful batteries if it can be recharged in a short time frame. [25] describes how faster charging can be gained by increasing the temperature of the battery with a high voltage. [26] has also demonstrated that higher temperatures conserve the battery life when charging, the higher temperatures reduce the lithium batteries from plating.

2.2.4 Polymers in Batteries

For both current and future batteries, polymers play an essential role. Polymer batteries have been around since 1987 when poly(aniline) was used [27]. Since then a range of other polymers have been explored for use in batteries. Specific polymers have been designed to allow for stable artificial interfaces and improve battery safety which is required for future batteries. Several polymers being developed to improve batteries [28], such as poly(ethylene oxide)-based networks which allow the battery and polymer to be recycled as it becomes malleable at higher temperatures, [29]. Some innovative polymers include self-healing binders, which will allow for longer life span batteries as the polymer will not degrade as quickly. Another area that has undergone intensive study are solid polymer electrolytes, which aim to be low cost and easy to integrate with current technology [28]. Polymers have multiple benefits over standard metals, such as better charging speed and cycling stability [30]. These benefits occur for several reasons; for example, a reduced ion diffusion path will improve the rate capabilities of most batteries, this is due to the increased number of pores found in polymer materials [31]. Polymers can be designed for a range of uses which can benefit the battery

industry. A new polymer, poly(2,5-dihydroxyaniline), has been synthesised with a quoted energy storage capacity of 443 mAhg^{-1} whereas NCA, Co-Al co-doped Lithium Nickel Oxide cathode, has an energy storage capacity of 200 mAhg^{-1} [32, 33]. Another novel polymer is polyaniline polysulfide which has a quoted energy storage capacity of 980 mAhg^{-1} , this is taken from an initial test and may decrease on repeated cycles [34]. However, there is little literature referencing polyaniline polysulfide since the first study, but there has been research on polyaniline, which has shown to be effective. Polyaniline has been shown to have high energy storage and retains the storage after repeated cycles - initial value of 785 mAhg^{-1} and drops to 472 mAhg^{-1} after 100 cycles [35].

SF has a discharge capacity of 133 mAhg^{-1} [36]. This, however, is dependent on the size of the pores present in the material. The capacity will increase as the pore size increases. For pores of $38\text{-}22 \mu\text{m}$, the discharge is 109.5 mAhg^{-1} , whereas, at $250\text{-}106 \mu\text{m}$, the discharge is 133 mAhg^{-1} [36]. This is in agreement with other publications where a range of $100\text{-}130 \text{ mAhg}^{-1}$ has been quoted [37].

For SF in batteries, there are two areas in which SF is used, the electrodes and the electrolyte. The electrolyte is made from SF mixed with choline nitrate, the inclusion of choline nitrate increases the diffusivity of ions through the material. Choline nitrate is a molten salt, which will conduct ions. The electrodes are formed by coating a piece of magnesium or gold in SF to create the anode and cathode, respectively. The anode may be made of other metals, but experimentally magnesium has been used. The silk layers encapsulating the battery will determine how long the battery will generate a current and how quickly the battery will then decompose [5]. A diagram of an SF battery has been included in figure 4

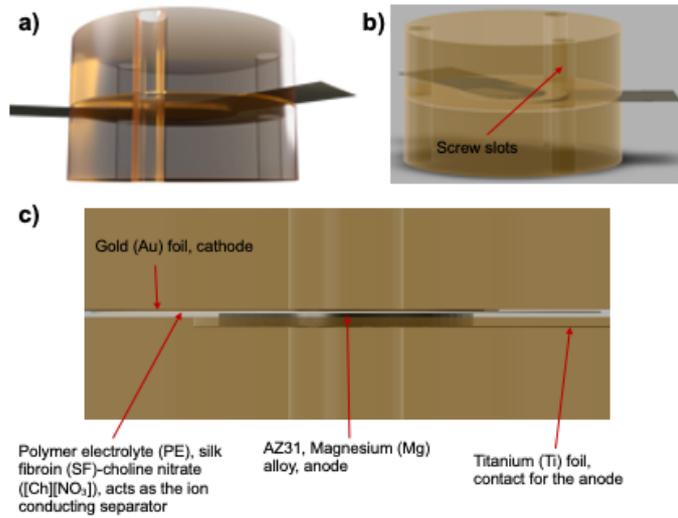


Figure 4: An image showing a schematic of an SF battery. a) View x of the battery. b) View z of the battery. c) a labelled diagram of the SF battery. Produced by Mathew Haskeew (unpublished).

2.2.5 Biocompatible and Biodegradable Batteries

Biocompatible batteries are batteries for use in the body. This biocompatibility is especially crucial for using batteries in the body, in a medical capacity, for personalised advanced healthcare (both wearable and implantable) devices [38,39]. Breakdown of the biodegradable battery produces the original materials. This breakdown can occur in a few ways such as bacteria performing decomposition on the material, the material breaking down over time or breakdown from an electric current. This breakdown has benefits such as reducing the electronic waste generated, as covered previously, it also reduces the number of operations required, only one is needed to implant the battery, and none will be required to remove the battery. One issue with biocompatible batteries is that the battery must work optimally at body temperature, or higher if present on the skin rather than inside the body [40].

2.3 Silk

Silk fibroin from the *B. Mori* silkworm is the primary material used throughout this project. The structure of SF is $(-Ala-Gly-Ser-Gly-Ala-Gly-)_n$. This Ala-Gly rich material is spun by silkworms to produce strong and tough fibres [1], the monomers are shown in figure 5. *B. Mori* is one of the most widely studied strains of silk due to the availability and biocompatibility of it. SF is biodegradable and biocompatible, as shown in [4,41]. This has led to

many proposed uses such as clothing, composite materials, sutures and a wide range of engineering and biomedical applications. However, synthetic silk has not successfully replicated the natural silk properties as of yet. Additionally, there has not been a full computational model of SF, using $(-\text{Ala}-\text{Gly}-\text{Ser}-\text{Gly}-\text{Ala}-\text{Gly}-)_n$ with experimentally accurate ratios of water, choline and magnesium ions. For this project, a significantly simplified structure of silk has been used $(-\text{Ala}-\text{Gly}-)_n$. SF will be modelled with various other molecules (water, magnesium ions and choline) that may be used in biological batteries.

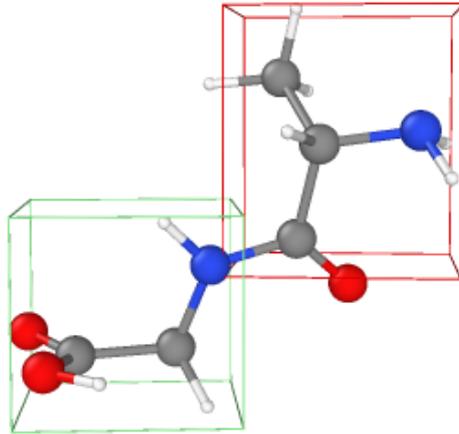


Figure 5: An image showing the simulated unit in SF, the green cell indicates glycine and the red cell shows alanine. Created by the Author using Ovito [42].

The SF simulated material repeat structure is $(-\text{Ala}-\text{Gly}-)_n$, as shown in figure 5 [43]. For this project, the size of the unit-cell is $(-\text{Ala}-\text{Gly}-)_{16}$, as discussed in the methodology. The structure shown, figure 5, has 20 atoms but still contains the capping atoms, which, when removed, form water to create longer chains such as $(-\text{Ala}-\text{Gly}-)_2$ or higher. Individually alanine contains 13 atoms, having two conformational isomers with only one structure found naturally. Glycine has ten atoms, with glycine being the most simple amino acid. This project does not consider the addition of Ser, as it does not affect the ^{13}C cross-polarisation magic angle spinning (CP-MAS) NMR chemical shifts. This has little effect on the simulated

structure however from an experimental view without Ser the SF will be soft whereas the addition of the Ser acts like glue and holds the SF together. The computational model is discussed in further detail in the methodology.

2.3.1 Composition of Silk

Experimental SF is a complicated structure. The full structure contains Gly (42.9 %), Ala (30.0 %), Ser (12.2 %), and Tyr (4.8 %) [44] and an assortment of other amino acids. These amino acids often occur in one of three repeat sequences: $(-Ala-Gly-Ser-Gly-Ala-Gly-)_n$, $(-Gly-Ala-Gly-Ala-Gly-Val-Gly-Tyr)_n$ and $(-Gly-Ala-Ala-Gly-)_n$ [44] shown in figure 6. The oligomer $(-Ala-Gly-Ser-Gly-Ala-Gly-)_n$ can create a crystalline region, the amorphous region will be formed of $(-Gly-Ala-Ala-Gly-)_n$ and negatively charged, polar, bulky hydrophobic, and aromatic residues [44]. There are different ways of showing the structure of SF, [45] describes the structure as repeat $(-Gly-X-)_n$ where X can take the form of Ala, Ser, Tyr, Val and Thr. However, Tyr, Val and Thr have not been mentioned as a part of the structure in either [44, 46]. This disparity may occur due to the low amounts Tyr, Val and Thr are found in the structure ($\leq 5\%$).

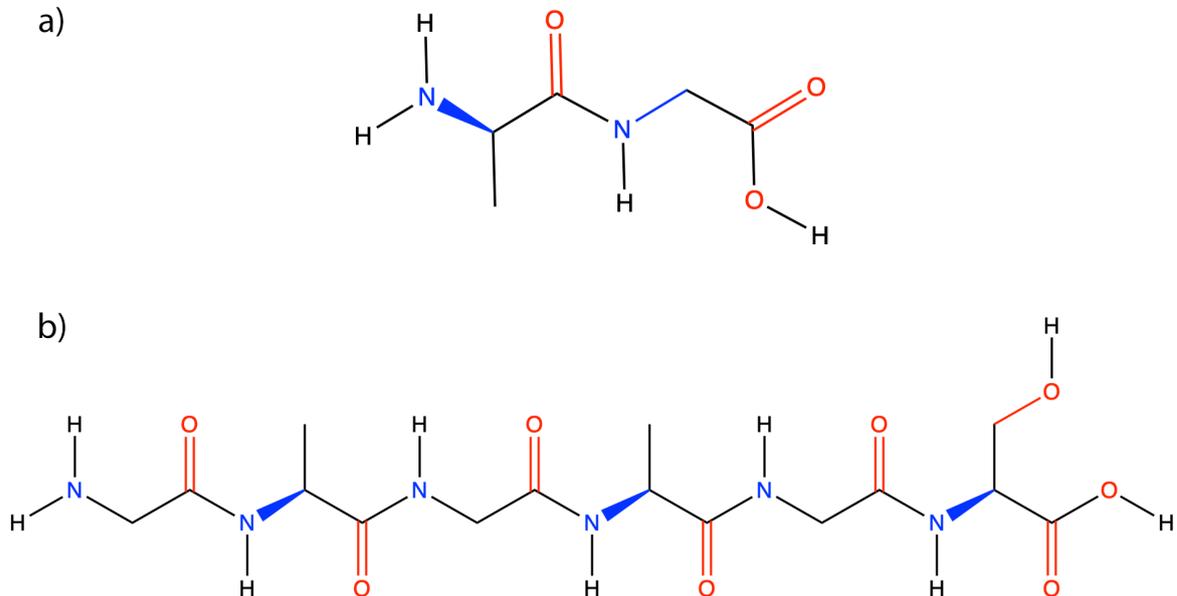


Figure 6: An image to show the difference in silk structure. Created by the Author. a) shows the simplified silk structure. b) shows the complex silk structure.

The structure of the silk comprises of a heavy (H) chain and a light (L) chain connected by a disulphide bond [44]. Various papers agree with the presence of a heavy and light chain, but the weights of the two are different, such as in [45] the heavy chain has a weight of 391 kDa and the light chain of 26 kDa. Whereas [44] has shown the weights to be 390 kDa and 26 kDa of the heavy and light chain. The difference in weight is minimal and has not changed for the light chain. [45] also confirms both the weight and the connecting disulphide bond but then goes on to say that the light chain has little influence on the structure and properties of SF. There is also a third chain which appears in aqueous silk - a glycoprotein named P25 with a weight of 30 kDa. The ratio of these appearing is 6:6:1, as shown in [45].

SF from *B. Mori* can assume two different forms, type I and type II. Silk I is the name given to silk before it undergoes spinning. Where spinning is the process of conversion from a random coil and alpha helical soluble conformations to beta sheet fibres. The silk I is formed of repeated β -turn type II [47]. This paper has performed this investigation on non-crystalline silk and thus may change at different (higher) temperatures and with different repeat sequences. The main sequence used to determine the type I silk has been $(-\text{Ala}-\text{Gly}-)_{15}$, which will be accurate for the model used throughout this research, where the primary structure used for calculations are $(-\text{Ala}-\text{Gly}-)_{16}$ [47]. The method for elucidating the type I structure uses solid-state two dimensional spin-diffusion nuclear magnetic resonance and rotational-echo double-resonance [48]. Type 2 silk occurs after spinning of the silk. It differs from silk I in structure and strength. The structure of silk II contains a significant portion of antiparallel β sheets [1]. An acceptable representation of what a parallel and antiparallel β sheets are in SF is in figure 7 [49, 50].

The silk sheets are held together by hydrogen bonds. For the silk material, the hydrogen bonds are formed between $i \rightarrow i + 4$. This notation describes the residues between which the hydrogen bonds form. Other hydrogen bonds can occur but $i \rightarrow i + 4$ are the most prominent bonds in this structure, especially in the α -helix formation. The hydrogen bonds in silk occur between the NH and CO groups, where the CO is double bonded. The CO group of residue i of a polypeptide is hydrogen bonded to the NH group of residue $i + 4$. The hydrogen bonds occur between the same two groups in the β sheets. $i \rightarrow i + 4$ intramolecular hydrogen bonds stabilise the repeat β -turn type II structure (the yellow dashed lines represents hydrogen bond interactions). The overall planar sheets are held together by intermolecular hydrogen bonding interactions, involving the central amide-bond of the β -turn, perpendicular to intramolecular interactions.

2.3.2 Secondary Structure

The structure and secondary structure can be investigated via multiple different pathways including infrared spectroscopy (and various forms of this), nuclear magnetic resonance spectroscopy, scanning electron microscope (SEM) and x-ray diffraction; this is a small subset of the available methods to investigate the structure. Each method provides slightly different information, for example, infrared spectroscopy (IR) shows the bond frequency when interacting with infrared radiation, this then allows for the type of bond and bond group to be deduced, and secondary structure information may also be present. NMR allows for functional group identification, mainly used for proteins. SEM allows for a close look at the structure; in general, the sample must be liquid or crystal depending on the available resources.

The silk structure can contain antiparallel β sheets and parallel β sheets, additionally, α helices can be found in the silk structure. α helices are a spiral structure; the most concise description is that the main-chain or polypeptide chains are twisted into a spiral, commonly found as a right-hand spiral but can occur as a left-hand spiral. The left-hand spirals occur when a large portion of the molecule is formed of glycine; this is true for the structure considered in this project [51]. This allows for the close packing of chains and accounts for the protein structure of the molecule. The β sheets do not allow for large side chains otherwise the structure would become unstable.

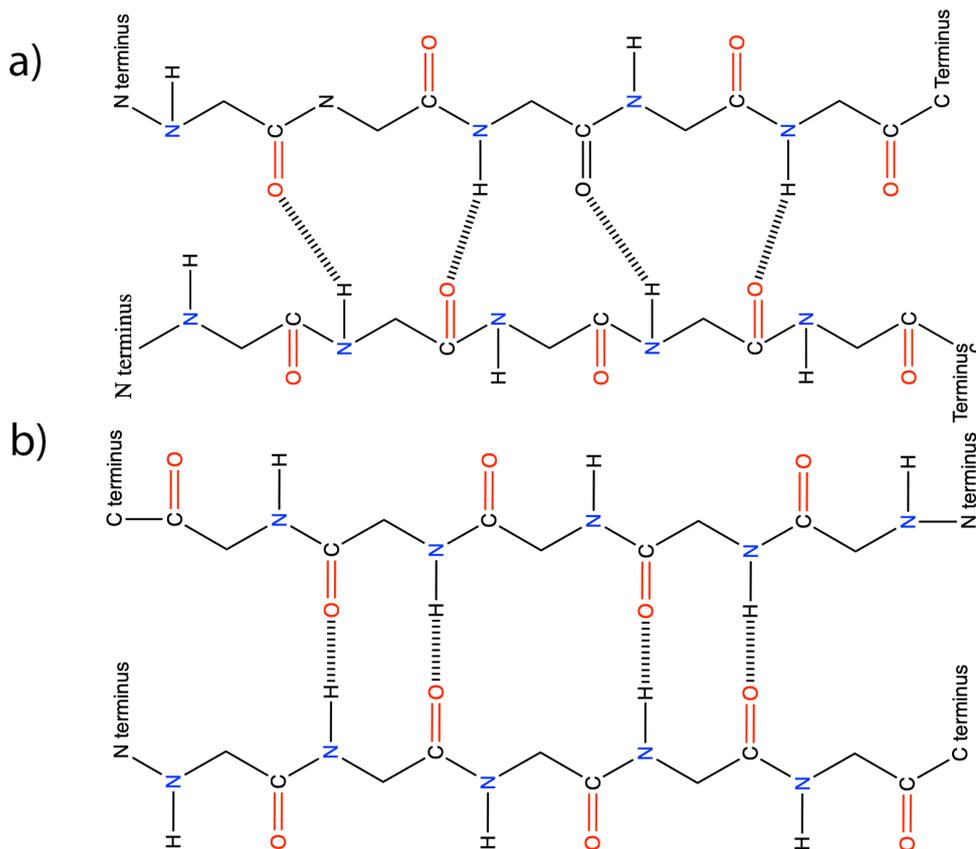


Figure 7: An image showing the antiparallel and parallel β sheets in the schematic form [50]. a) Shows the parallel β sheets. b) Shows the antiparallel β sheets. Made by the Author.

Infrared spectroscopy has been extensively performed for the silk material. Three main types of IR spectroscopy are considered in this report: IR, FTIR and Raman spectroscopy. The experiments have been performed for multiple types of silk; this allows for a comparison of their structure [52]. The comparison has been made on naturally occurring silk material rather than synthetic silk, which will have a unique IR spectrum. The IR spectra allows for some details on the structure of a simplified version of *B. Mori* silk it produces a peak at 1395 cm^{-1} which is consistent with the greater amount of sericin in the outermost layer for silks such as *B. Mori* [52]. A difference between the inner and outer layers of silk has also been demonstrated via the use of IR [52]. [53] has used FTIR to analyse the secondary structure and estimate the percentage of α helix and β sheets present in the silk material. [54] has performed FT-IR on synthetic SF fibres, with the results agreeing strongly with those presented in [53]. Both papers have shown that the most substantial peak occurs at $1652/1657\text{ cm}^{-1}$, which is a good agreement between the two. However [54] has only provided the data between 1600 and 1700 cm^{-1} and the rest of the spectrum is not visible to the reader this means a direct comparison can not be performed and relative absorbance can

not be compared. [55] has also performed FT-IR; however, this has used a slightly different silk model utilising *Antheraea pernyi* SF instead of *B. Mori*. The main peak produced for this silk occurs at 1659cm^{-1} , very similar to the previous data generated by FT-IT data. The IR of *A. Pernyi* and *B. Mori* are shown to be similar in [52], indicating that the structure of the two are similar and would have mostly the same proteins present. This reduced range has also removed the fingerprint region, which is a useful tool for identifying unique structures. The third type of IR that can be used is Raman spectroscopy, the shape of the graph and the positions of the peaks are again very similar to FT-IR [56]. There are a few differences in the spectra, but this may be a result of [56] using liquid silk rather than SF solid. As the monomers of SF, the IR spectra of alanine and glycine have been included to compare to the SF IR spectrum, figures 8.

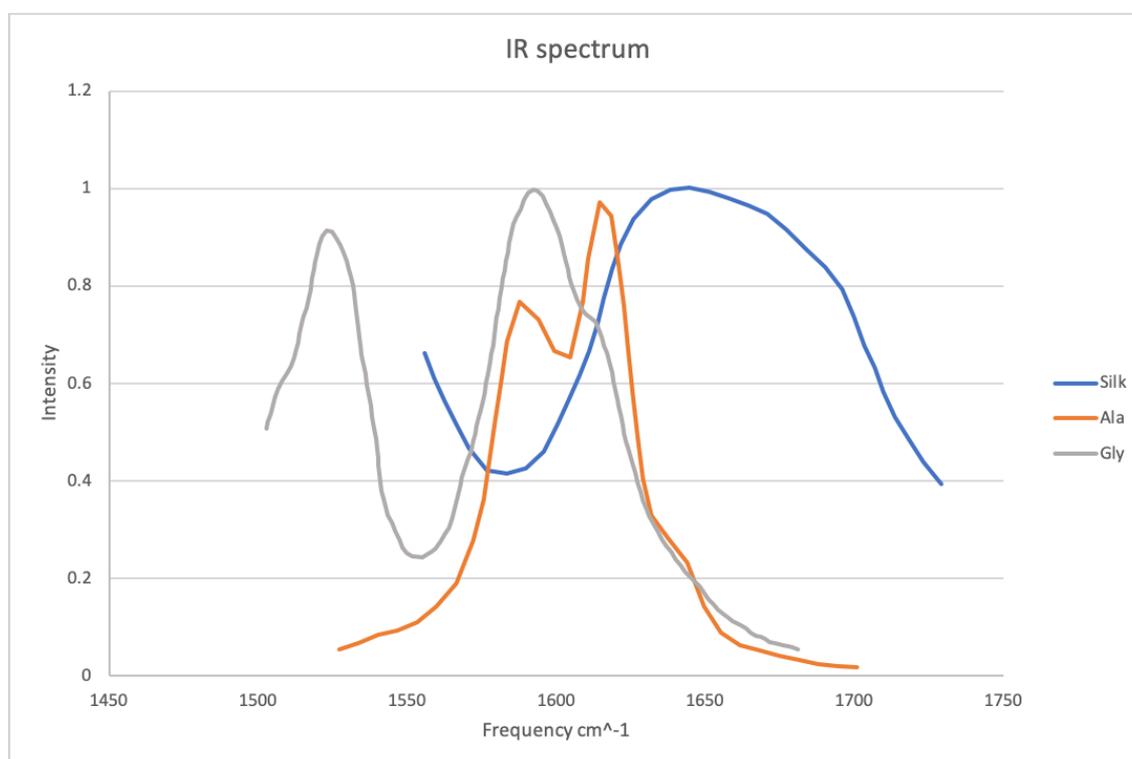


Figure 8: Figure showing the IR spectrum of natural spun SF, Glycine and Alanine, reproduced by the Author using data from [57].

Another critical tool for the analysis of the secondary structure of *B. Mori* is the use of Ramachandran plots. Ramachandran plots plot ϕ vs ψ torsional angles, shown in figure 10. A torsional angle, more commonly known as a dihedral, is the angle between two intersecting planes, formed by three consecutive bonds in a molecule, demonstrated in figure 9. This tool is mainly used for studying the secondary structure of proteins [58].

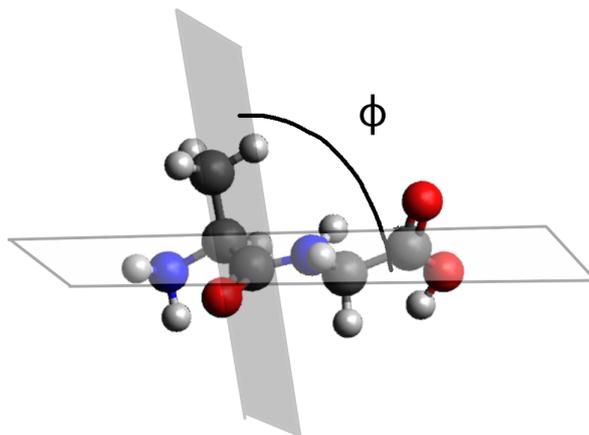


Figure 9: Figure showing a torsional angle, ϕ , in a silk chain. Created by the Author.

Ramachandran plots can be determined both experimentally and computationally, which lends itself to being a useful tool for verification, to see how accurate the computational model is compared to expected values. Furthermore, Ramachandran plots are unique for each molecule and thus provides a method for characterisation. The experimentally determined plots that are of importance for this project are Ala and Gly Ramachandran plots; this is due to these proteins being the building blocks for the silk structure. There are silk Ramachandran plots available which are discussed below, these are from experimental structures of *B. Mori* silk, as shown in figures 10 and 11. There is a description of the proteins and associated Ramachandran plot - including a breakdown of the Ramachandran plots for each specific region. Below the Ramachandran plots for Gly, Ala and Ser have been included due to the importance and relevance. The plots will allow a direct comparison later in the project; these are in figures 12, 13 and 14. The comparison of these three will provide the allowed regions for ϕ and ψ for the full silk structure. The use of ala and gly will provide the allowed regions for the simplified silk structure, as discussed later in the thesis. [59,60] have discussed types of structure present in Ramachandran plots and the nomenclature that corresponds to this, shown in figure 15. [60] has presented an argument for renaming one such region the turn region as it encompasses so many different secondary structures that it

is hard to justify using one name to represent the area.

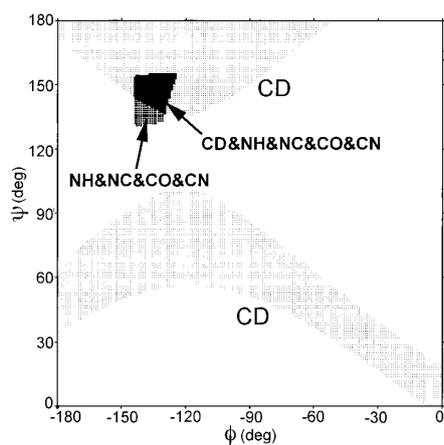


Figure 10: Ramachandran plot of the allowed torsion angles of alanine in SF, as determined by solid-state NMR. Taken from [61].

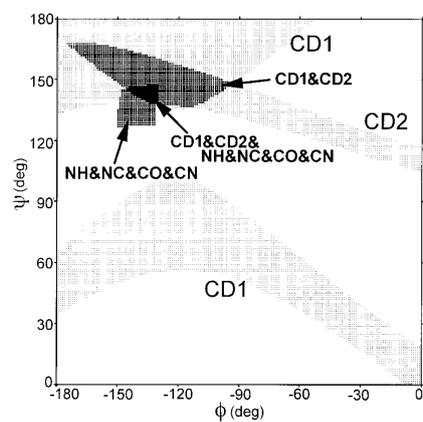
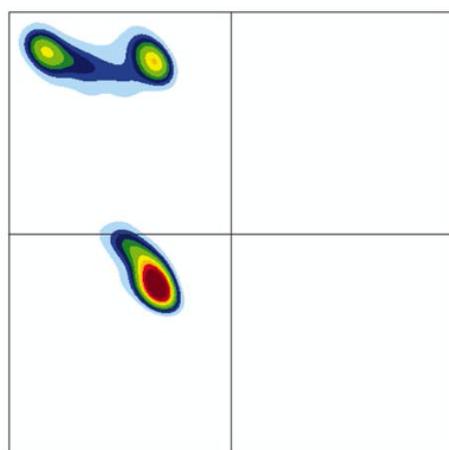
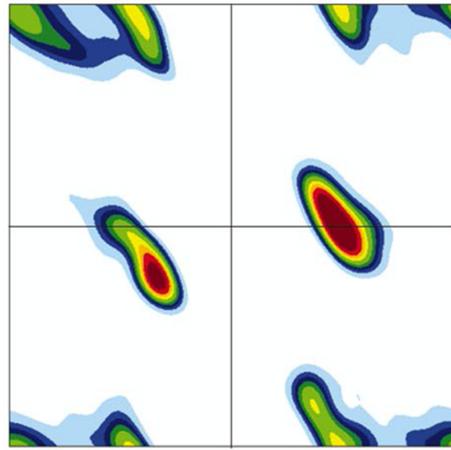


Figure 11: Ramachandran plot of the allowed torsion angles of glycine in SF, as determined by solid-state NMR. Taken from [61].



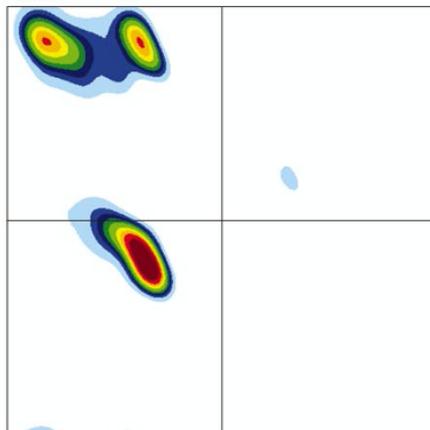
Ala

Figure 12: Ramachandran plot for alanine molecule, taken from [60].



Gly

Figure 13: Ramachandran plot for glycine molecule, taken from [60].



Ser

Figure 14: Ramachandran plot for serine molecule, taken from [60].

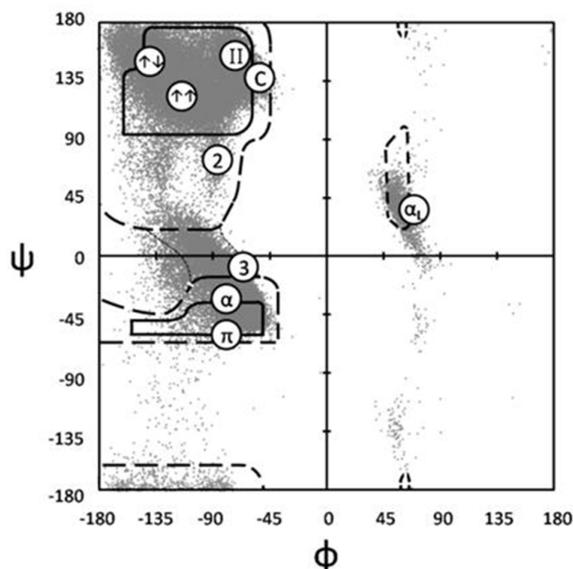


Figure 15: Ramachandran plot for alanine molecule, taken from [59]. Where: α -helix (α), 310-helix (3), π -helix (π), left-handed α -helix (α_L), 2.27 ribbon (2), polyproline-ii (II), collagen (C), parallel β -sheet ($\uparrow\uparrow$), and antiparallel β -sheet ($\uparrow\downarrow$). Taken from [60].

2.3.3 SF Battery

In addition to the charged ions, choline or choline nitrate will also appear in the cell. Choline nitrate is an ionic liquid which improves the conductivity of the cell, as thus very important for inclusion into an SF battery. Due to the size of the choline, it will have a significant effect on the structure of the cell. Choline nitrate is two separate species with cationic choline and anionic/zwitterionic glycine [62]. There is an agreement that choline nitrate will be present as two separate species; however, there is little agreement or investigation into the behaviour of choline nitrate in polymers, looking specifically at silk. [63] suggests that the choline nitrate or any form of choline will statically bind to polymer chains via hydrogen bonding. This computational paper also shows that the choline molecule will sit in between the polymer chains. [64] has shown that as choline nitrate is introduced into the system, the tensile strength is reduced. This is a direct result of choline nitrate reducing the number of hydrogen bonds that can form; this is supported by [5] which shows that choline increases the distance between the chains and thus reduces the hydrogen bonding strength between Ala and Gly. This is not in agreement with [63], where it shows the inclusion of the ionic liquid

increases the bonding between adjacent chains, and thus an increase in tensile strength. The two papers differ in the type of ionic liquid used, [63] has used choline glycine, whereas [5,64] have used choline nitrate. The same Author writes both [5,64]. Furthermore [63,65] both agree that that the inclusion of choline ions stabilises the structure due to the molecule binding to the polymer, it is assumed the binding is formed from hydrogen bonds, but this is not explicitly stated. Choline nitrate has been shown, computationally, to be mobile in specific polymers where the charge allows it to travel through the porous structure [66]. Three SF batteries with varying amounts of choline have been created [5]. The amounts are 1:1, 1:3 and 1:5 with the ratio of silk to choline by molecular weight. The number of water molecules added will not affect the choline ratio. The structure of the silk will vary when additional molecules are added into the cell. The most likely molecules that would interact with the silk are water, charged ions and choline [5]. This paper has suggested sodium and magnesium ions to be the cations to diffuse through the cell. However, multiple papers have suggested that various other ions could be used, and it appears that we are not limited to a specific ion for use with silk [37]. The papers have comparable Nyquist plots for the two different ions, Mg and Li. The two graphs show that the Mg has a higher Z'' than the lithium-ion. The lithium-ion experiment has been performed at 25 degrees celsius. However, the temperature is unknown for the magnesium ion, there is a suggested range of 20 to 150 degrees celsius, but this is not a definitive temperature, [5,37]. The temperature will have a direct effect on the resistance of the battery. The two plots have been shown in figure 16.

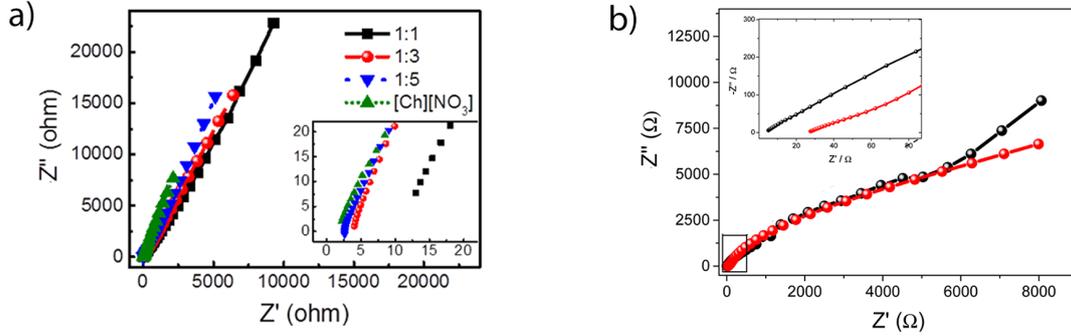


Figure 16: An image to show the comparison of two Nyquist plots of two different ions in the Silk cell. Where a is the Nyquist plot for the magnesium ion taken from [5], and b is the Nyquist plot for the lithium-ion taken from [37].

2.3.4 SF Computational Simulations

For simulations classical mechanics or density functional theory can be used. There are a few limited papers that have utilised quantum mechanics to simulate silk structures [67]. This has used DFT to create an optimised structure, and then a classical mechanic's molecular dynamics run has been performed. The classical mechanics methods have used a variety of force fields which will alter the results. Specific force fields will have been optimised for different calculations. The CHARMM27 force field has been used in both [68, 69]; the force field has been previously used to simulate proteins and earlier silk models accurately. [67] has used the parm99 forcefield to simulate the silk for a molecular dynamics run. This is a result of additional ions in the system and thus a requirement for a different force field, however, little reasoning has been offered for this choice.

The paper that has performed a range of calculations is a study into the change of silk structure via the use of molecular dynamics [1]. A more in-depth study of the silk material has been performed in [68], where deformations under tensile and shear stress have been

investigated. The same paper has investigated both amorphous and crystalline regions of silk which very few other papers have completed. There are a few differences between the methods taken by both papers - the size of the cell, the structure used and the method used in the molecular dynamics. For the size of the cell [68] has carried out a further investigation into the optimal size for simulations to recreate experimental results. The model should have at least six to eight residues to recreate the β sheet structure accurately. Other papers have shown the size of their boxes but not the reason behind this [1,69]. The published box sizes will impact the generated structure for the project. Four boxes have been published for the SF structure: $23.64 \text{ \AA} \times 19.2 \text{ \AA} \times 22.4 \text{ \AA}$ [68], $20 \text{ \AA} \times 20 \text{ \AA} \times 20 \text{ \AA}$ [68], $20 \text{ \AA} \times 20 \text{ \AA} \times 20 \text{ \AA}$ [69] and $55 \text{ \AA} \times 38 \text{ \AA} \times 71 \text{ \AA}$ [67]. The box size from [67] is exceptionally large compared to the others due to the investigation considering how SF behaves at interfaces with water and thus has additional space to include the water.

The structure used will have a significant impact on the material. The type of structure simulated will also change; it appears that modelling only the crystalline region is the most adopted method, [1,67,69]. It is possible to simulate the amorphous region of silk via molecular dynamics [68]; however, this is not regularly undertaken, due to the complex structure and more expensive computational cost. The structure allows for differences between two models to be compared; however, few studies have been performed on the amorphous region.

3 Methodology

This methodology aims to discuss and rationalise the method taken to generate the results for this project. The reasons for any choice made has been shown with supporting calculations and graphs to rationalise and validate the method.

3.1 Classical vs Quantum

Classical mechanics and quantum mechanics can both be used to simulate energies, atomistic structures and trajectories of atoms. The main differences between the two methods lies in how the energies are calculated. In classical mechanics the energy is calculated numerically solving Newton's equations of motion for a system of interacting particles, where forces between the particles and potential energies are calculated using user defined molecular mechanics force fields. Where as KS-DFT solves the the one-electron Schrödinger equation as discussed in more detail below. This is an important distinction as the lack of force field allows for DFT to simulate the breaking and making of bonds between any atoms in a specified distance. This is a result of the additional electronic information included in the electronic structure. This is not a possibility with the use of force fields as only user-defined interactions can occur, which limits bonds forming and breaking but shows how bonds act in given environments. Quantum mechanics has been chosen over classical mechanics for this project for a variety of reasons. Quantum mechanics allows for a deeper understanding of the electronic behaviour and thus will simulate the making and breaking of hydrogen bonds more rigorously than compared to classical. This is important for SF due to the number of hydrogen bonds that can be formed in the structure. Part of the motivation of this project is to provide a comparison to classical mechanics data and thus quantum mechanics will be required for this.

Although [70] was published in 2004 and does not include the latest developments in either classical or quantum simulations, it discusses the differences in accuracy between the two types of simulations. It has concluded that quantum calculations are more accurate even when classical mechanics use a three-body term. However [71] has shown good agreement between the two methods of calculation, this may be due to improvements of computational resources or those simulations performed in [71] are more straightforward, and thus a similar accuracy is expected. DFT can simulate up to a few thousand atoms in a structure, given sufficient computational power whereas classical mechanics can simulate up to a million atoms, again dependent on the computational power available. This system size will affect the applications DFT can be applied to, biological viruses will require more atoms than DFT

is capable of, but a small crystal structure will be simulated to a high degree of accuracy by DFT. This shows DFT is more computationally expensive compared to classical mechanics; this cost is mainly dependent on basis sets and functionals chosen.

3.2 DFT+D3

Density functional theory is a modelling method used to investigate electronic structure [72]. This can be performed on a range of systems from single atoms up to multibody systems. Kohn-Sham DFT (KS-DFT) is a variation of DFT where the total energy of the investigated system is given as a functional of the charge density, similar to DFT; this aims to solve the one-electron Schrödinger equation, [73]. The particles in the Kohn-Sham system are non-interacting fermions which gives rise to a density similar to a system of interacting particles. There are other DFT methods including SDFT (spin-DFT) and RDFT (relativistic-DFT). The variations are concerned with adding components to make DFT more accurate. For example SDFT looks at molecules behaviour when a magnetic field is applied. This can be completed relatively easily by adding a suitable index to the densities.

The equations used in KS-DFT are given in equations 1 to 3. For equation 1 $E[p]$ is the total energy, $T_s[p]$ is the kinetic energy of the system, v_{ext} is the external potential acting on the interacting system, $E_H[p]$ is the Hartree energy, and $E_{xc}[p]$ is the exchange-correlation energy. For equation 2 where $T_s[p]$ is the kinetic energy of the system, \hbar is the reduced Planck constant, ∇^2 is the Laplacian operator. For equation 3 $p(r)$ is the electron density and r is a given radius.

$$E[p] = T_s[p] + \int dr v_{ext}(r)p(r) + E_H[p] - E_{xc}[p] \quad (1)$$

$$T_s[p] = \sum_{i=1}^N \int dr \sigma_i^*(r) \left(-\frac{\hbar}{2m} \nabla^2\right) \sigma_i(r) \quad (2)$$

$$E_H = \frac{e^2}{2} \int dr \int dr' \frac{p(r)p(r')}{|r-r'|} \quad (3)$$

The exchange-correlation energy is determined through the use of exchange-correlation functionals. This is due to the actual form of E_{xc} is not known; approximate functionals based upon the electron density are used to describe this term. $E_{xc}[p]$ can contain empirical parameters where the values have been fitted to experiments or non-empirical parameters. For example, BLYP contains empirical parameters, and PBE does not. The correct functional for each project will change depending on the required calculations and must be determined

via comparison calculations.

Density functional theory + dispersion correction (DFT+D3) has been utilised throughout this project for all calculations [72]. DFT functionals, as standard, do not recreate the London dispersion interactions accurately [74]. KS-DFT poorly recreates long-range interactions; this is due to KS-DFT works for one-electron systems and thus will not contain any terms for electron-electron interactions, [73]. However, the addition of D3 accounts for long-range dispersion interactions which include hydrogen bonds, this is essential for modelling large systems of polymers [72]. The D3 is an empirical correction for the DFT model, as generally, all DFT models neglect the long-range interactions and hydrogen bonds due to the computational cost and low efficiency of generating the interaction [75]. The D3 term improves accuracy without being significantly more computationally expensive and more precisely describes the electrostatic and exchange-repulsion interactions [75]. The DFT requires a basis set and an exchange-correlation function. For this project DZVP and BLYP have been chosen. This is discussed in greater detail in section 3.3 and 3.4.

The main equations for DFT are equations 4 to 11.

Equation 4 is the total DFT+D3 energy created by the addition of atom pairwise dispersion correction to the KS-DFT energies [72]. The C_9 term models this long-range interaction. The dispersion energy is made up of the two and three-body interaction energies, as shown in equation 5. The most important two-body term is calculated at long range in equation 6. Where the C_n^{AB} term denotes the n-th order dispersion coefficient for atom pair AB. r_{AB} is the internuclear distance between any two atoms, A and B, where $A \neq B$. For equation 7, s_n is a functional dependent scaling factor which matches the long and midrange correlation of D3 with the semi-local correlation computed by the functional, in this case, the functional is BLYP. With $f_{n,d}$ is an adequate damping function which is included in all calculations, with all other parameters previously appearing in other equations. R_0^{AB} and $f(R_0^{AB})$ in equations 8 and 9, [74, 76, 77]. In addition to the two-body term, a three-body term can be calculated for DFT-D3 by Axildor-Teller-Muto (ATM) model, which is more complex as this includes angles and bond lengths rather than just the equilibrium distance between two atoms [78], this model is shown in equation 10. The angles used in equation 10 are the internal angles formed by the triangle of three bodies: A, B and C. Perhaps the most important equation for this DFT+D3 is given in equation 11, this term accounts for the long-range dispersion forces generated by the London dispersion forces and hydrogen bonding [72].

Equation 4 gives the total energy of the DFT-D3 corrected system, with E_{KS-DFT} representing the Kohn-Shan DFT as discussed above and E_{disp} is the energy of the dispersion

correction.

$$E_{DFT-D3} = E_{KS-DFT} + E_{disp} \quad (4)$$

Equation 5 is the dispersion correction as a sum of two- and three-body terms, E^2 and E^3 respectively.

$$E_{disp} = E^2 + E^3 \quad (5)$$

Equation 6 shows the long-range dispersion correction. Equation 7 shows the two-body term. C_n^{AB} denotes the averaged isotropic nth order dispersion coefficient for atom pair AB, where r_{AB}^n is their internuclear distance [72]. S_n is a scaling factor, functional dependent, to match the long and midrange correlation of D3 with the semi-local correlation computed by the functional, an adequate damping function, $f_{n,d}$, should be included.

$$E_{disp} = -\frac{1}{2} \sum_{A \neq B} \sum_{n=6,8} \frac{s_n C_n^{AB}}{r_{AB}^n} \quad (6)$$

$$E^2 = -\frac{1}{2} \sum_{A \neq B} \sum_{n=6,8} s_n \frac{C_n^{AB}}{r_{AB}^n + f(R_0^{AB})^n} \quad (7)$$

The internuclear distance is given by equation 8, C_8^{AB} , C_6^{AB} is a specific-order dispersion coefficient, in this case, $n = 6,8$ [72].

$$R_0^{AB} = \sqrt{\frac{C_8^{AB}}{C_6^{AB}}} \quad (8)$$

The damping function is in equation 9. Where R_0^{AB} is the cutoff radius, and a_1 is the steepness parameter which has been set [72].

$$f(R_0^{AB}) = a_1 R_0^{AB} + a_1 \quad (9)$$

In equation 10, θ is an angle, r is the distance between two atoms and C_9 is a specific order dispersion coefficient.

$$E^3 = -\frac{1}{6} \sum_{A \neq B \neq C} \frac{C_9^{ABC} 3 \cos(\theta_a) \cos(\theta_b) \cos(\theta_c) + 1}{(r_{AB} r_{BC} r_{CA})^3} f_{dmp}(r_{ABC}) \quad (10)$$

For equation 11, all three terms are the dispersion coefficient between two atoms.

$$C_9^{ABC} \approx -\sqrt{C_6^{AB}C_6^{BC}C_6^{CA}} \quad (11)$$

A brief comparison of simulations with and without the D3 correction have been performed. This was performed on single alanine and glycine molecules with an additional water molecule to see how the hydrogen bond is affected and the effect on the energy of the cell. Figure 17 shows that without the D3 correction, a second hydrogen bond has been unable to form. This is accurately represented in table 1, where the D3 calculation has lower energy by 0.02 au. The second bond has not formed as an exclusion radius of 2.5 Å has been set.

Type	Geometry Optimisation	
	Length of Hydrogen Bond (Å)	Length of Hydrogen Bond 2 (Å)
D3	1.6913	2.0242
No D3	1.5986	NA

Table 1: Table showing the difference in length of the hydrogen bonding in alanine.

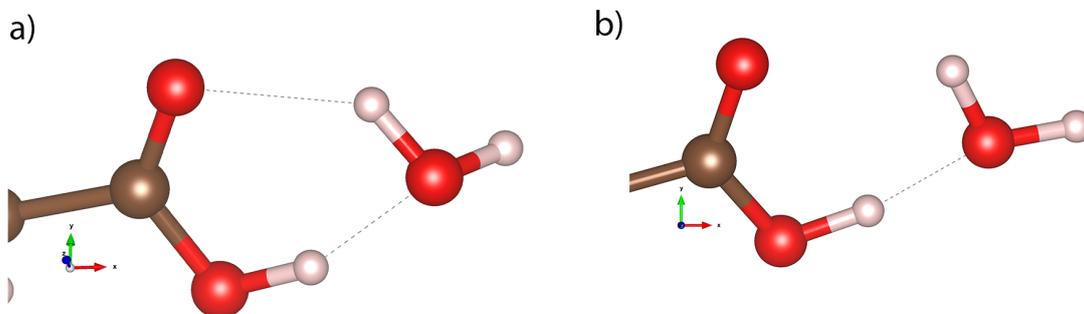


Figure 17: An image to show the change in hydrogen bonds for D3 and no D3 calculation in alanine. Created by the Author using Vesta [79]. a) shows two hydrogen bonds formed in the D3 calculation and b) shows the singular hydrogen bond in the no D3 calculation.

The DFT script has been included in supporting information.

3.3 Functionals

For this project, three functionals have been considered:

- PADE
- PBE
- BLYP

PADE is the Local-density approximations (LDA) CP2K equivalent. Perdew, Burke and Ernzerhof (PBE) and Becke, Lee, Yang, and Parr - BLYP is an exchange-correlation functional [76, 80, 81].

The equation for LDA/PADE is given in equation 12 [81]. Where $n(r)$ is the density, r is the radius, $E_{xo}[n]$ is the exchange and correlation energy per electron, ν_{xc} is the exchange-correlation, ϵ is the correlation energy per particle.

$$E = \sum_1^N \epsilon - \frac{1}{2} \int \int \frac{n(r)n(r')}{|r-r'|} dr dr' + E_{xo}[n] - \int \nu_{xc}(r)n(r)d(r) \quad (12)$$

The equation for the functional, BLYP is given in equations 13 and 14 [74,76,77]. Where E_X^{LDA} is the energy of the LDA approximation, ρ_σ is the spin density, γ and β are chosen coefficients, SE denotes semi-empirical, and x_σ is the dimensionless ratio given in equation 14. With ∇ being the Laplacian operator, $p(\sigma)$ is the electron density.

$$E_X^{SE} = E_X^{LDA} - \beta \sum_\sigma \int \rho_\sigma^{\frac{4}{3}} \frac{x_\sigma^2}{(1 + \gamma x_\sigma^2)} d^3 r \quad (13)$$

$$x_\sigma = \frac{\nabla \rho_\sigma}{\rho_\sigma^{\frac{4}{3}}} \quad (14)$$

The equation for PBE is given in equation 15 [80]. Where $\epsilon_X^{unif}(n)$ is the exchange correlation energy per particle of a uniform electron gas, ζ is the relative spin polarisation, and s is a dimensionless density gradient, and F_{XC} is an enhancement factor over a local range.

$$E_{XC}^{GGA}[n_\uparrow, n_\downarrow] = \int d^3 r n \epsilon_X^{unif}(n) F_{XC}(r_s, \zeta, s) \quad (15)$$

To determine which functional to use a comparison has been made, both to literature and each other. The literature comparison involved running small simulations of the monomers of SF with each functional and comparing to literature values. The calculations that have been performed analysed bond lengths of alanine, glycine and choline and energy calculations of constituent atoms. To gain a comparison the same basis set, DZVP, has been used throughout. Comparing the speed and computational cost of functionals have been shown in table 2, these simulations have been performed on SF. A second method for choosing which functional is best suited for this project is comparing bond lengths to experimentally calculated values; however, there are no experimental values for the simplified SF structure, these simulations have been performed on alanine and glycine to make up for this lack of data, this comparison is shown in table 3 and 4. The published data is from [82]. These bond lengths have been calculated using the geometry optimisation method.

Basis Set	AIMD			
	Cores	Cost (kAUs)	Steps run	Steps per kAU
BLYP	96	34.6	1451	41.9
PBE	96	34.6	1427	41.2
PADE	96	34.6	1662	48.0

Table 2: Table showing the number of steps produced for various functionals.

Atoms		PADE		PBE		BLYP		Published data
Atom 1	Atom 2	Bond length (Å)	Error to published data (%)	Bond length (Å)	Error to published data (%)	Bond length (Å)	Error to published data (%)	Bond length (Å)
C1	H1	1.0799	-1.7011	1.1101	1.0366	1.0987	0.0027	1.0986
C1	H2	1.1172	-1.9799	1.1086	1.1967	1.0992	0.3332	1.0955
C1	H3	1.1317	2.9352	1.1177	2.9352	1.1025	0.2847	1.0994
C1	C2	1.5334	0.2307	1.5432	0.8680	1.5301	0.8510	1.5299
C2	H4	1.1230	1.9862	1.1275	2.3957	1.1124	1.0244	1.1011
C2	N1	1.4728	-0.2715	1.4962	1.3129	1.4980	1.4362	1.4768
C2	C3	1.4948	-3.3881	1.5484	0.0801	1.5439	0.4020	1.5472
N1	H5	1.0437	2.4520	1.0307	1.1681	1.0268	0.7892	1.0188
N1	H6	1.0451	2.5714	1.0357	1.6469	1.0266	0.7498	1.0189
C3	O1	1.2228	1.0270	1.2234	1.0741	1.2227	1.0179	1.2104
C3	O2	1.3534	0.4996	1.3857	2.9006	1.3967	3.7204	1.3466
O2	H7	0.9675	-3.2172	0.9786	-2.1008	0.9808	-1.8807	0.9996

Table 3: Table showing the comparison of alanine bond lengths to published data (all numbers are rounded to 4dp).

Atoms		PADE	PBE		BLYP		Published data
Atom 1	Atom 2	Bond length (Å)	Bond length (Å)	Error to Published data (%)	Bond length (Å)	Error to Published data (%)	Bond length (Å)
N1	H4	N/A	1.0338	-1.1304	1.0243	-2.0361	1.0456
N1	H5	N/A	1.0296	-1.6187	1.0255	-2.0096	1.0465
N1	C1	N/A	1.4642	0.2671	1.4770	1.1429	1.4603
C1	H2	N/A	1.1217	0.8397	1.1143	0.1780	1.1123
C1	H3	N/A	1.1170	0.5238	1.1042	-0.6228	1.1112
C1	C2	N/A	1.5325	1.7076	1.5423	2.3607	1.5068
C2	O1	N/A	1.2223	0.0458	1.2214	-0.0246	1.2217
C2	O2	N/A	1.3768	2.0751	1.3991	3.7291	1.3488
O2	H1	N/A	0.9839	1.6899	0.9790	1.1865	0.9675

Table 4: Table showing the comparison of glycine bond lengths to Published data (all numbers are rounded to 4dp).

The atom references can be seen in figures 18 and 19.

From tables 3 and 4, it can be seen that PADE accurately describes C-O bonds in alanine but does not describe other bonds as well as other functionals. However, when PADE has been used to simulate glycine the structure has broken and has lead to PADE being removed from further calculations (except the basis sets comparison calculations as they were performed before this result). BLYP appears to replicate the majority of the bonds at a lower level of error compared to PBE for alanine however both functionals perform equally when applied to glycine, with PBE seemingly recreating the nitrogen environment better for all cases of N-X bonding.

Further calculations have been performed to further differentiate between PBE and BLYP. Single point energy calculations have been performed on the constituent atoms and then compared these to experimental data, shown in table 5. The experimental data is included in the potential file provided by CP2K [83]. There has been a lack of data for hartree energies for alanine and glycine.

Molecule	PBE		BLYP		Published data
	Energy (hartree)	Error to published data (%)	Energy (hartree)	Error to published data (%)	Energy (hartree)
H	-0.4957	-0.8690	-0.4945	-1.1088	-0.5
C	-5.4020	0.8393	-5.3441	-0.2416	-5.3570
N	-9.7306	0.9186	-9.7271	0.8820	-9.6420
O	-15.7418	-0.2033	-15.7516	-0.1410	-15.7739
Ala	-63.1767	-	-63.0611	-	-
Gly	-56.3018	-	-56.2108	-	-

Table 5: Table showing the comparison of single-point energy calculations to experimental values.

In table 5, BLYP has performed better for all atoms in comparison to experimental results except the hydrogen atom, this can be seen by looking at the percentage error for each atom.

This has lead to the decision to use BLYP for all further calculations, except vibrational calculations which use PBE functional, this is explained in section 3.6.3.

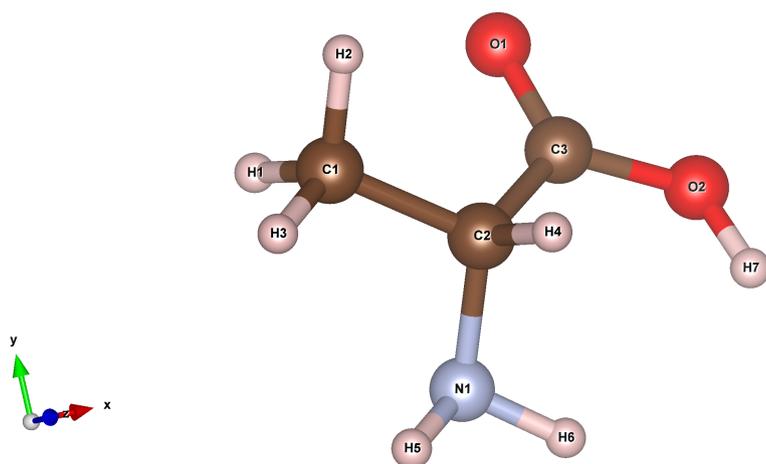


Figure 18: An image to show the label nomenclature for alanine. Created by the Author using Vesta [79].

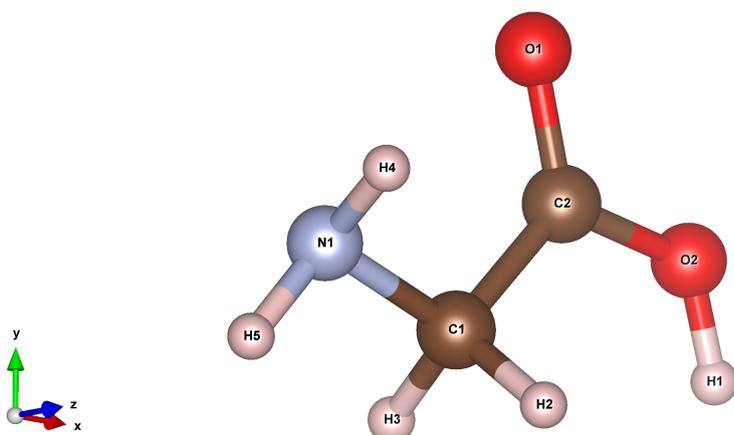


Figure 19: An image to show the label nomenclature for glycine. Created by the Author using Vesta [79].

3.4 Basis Sets

A basis set is a set of one particle functions used to create molecular orbitals for the system, these functions are typically atomic orbitals centered on atoms, but can theoretically be any function. A diagram of how a basis set is used is shown in figure 20. For this project, three basis sets have been considered:

- 6-31G*

- DZVP
- TZVP

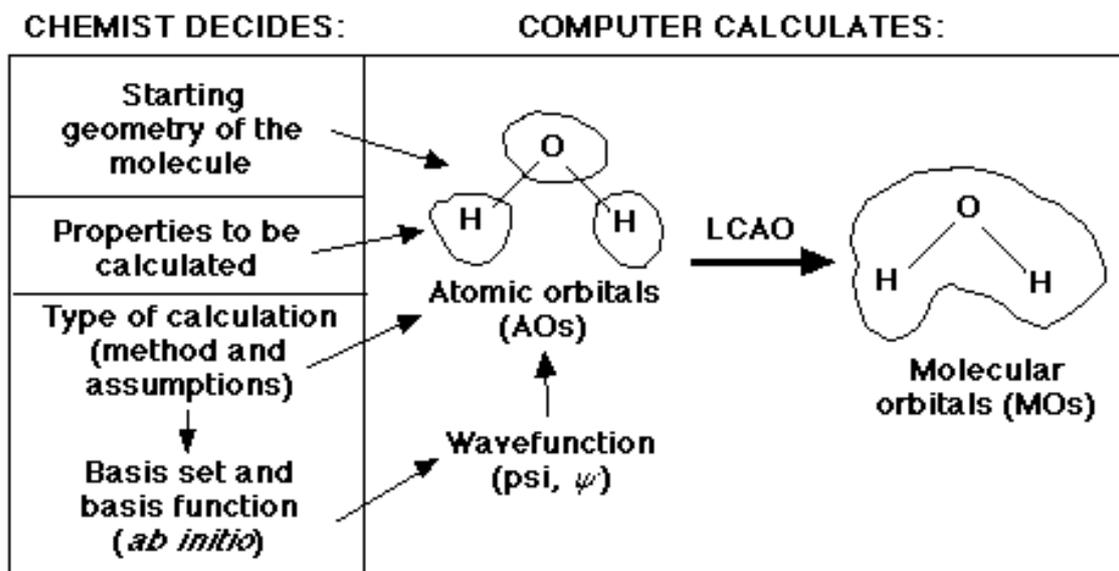


Figure 20: An image to show how a basis set is used. Taken from [84].

Double zeta valence polarised basis set (DZVP) contains two contracted functions per orbital [85]. DZVP is the CP2K equivalent for 6-31G*. Triple zeta valence polarised basis set (TZVP) has three contracted functions, and Quad zeta valence polarised basis set (QZVP) has four and so on. One consideration that has to be made is that as the number of functions included per orbital increases, the computational cost will increase as well. Both DZVP and TZVP have one additional set of polarisation functions.

Other basis sets in CP2K that are worth mentioning are SZV - single zeta valence which is the smallest basis set available and should not be used for any calculation; and TZV2P - triple zeta valence two polarisation this has not been included in the comparison table due to the restricted elements available for this basis set and the high computational cost. However, TZV2P has been used for IR calculations due to the requirement of having a tighter convergence and require a better starting structure than other calculations.

Similar to the comparison of potentials above comparisons of speed and cost of three basis sets have been made. The three have been chosen due to the varying size where 6-31G is the smallest basis set and TZVP being the most extensive set considered here. All calculations have been run with the fastest potential, PADE. The choice of PADE was made, so the time dependency is on the basis set rather than the potential. All calculations have been run on $1 \times 1 \times 1$ cell, and again this is to reduce the computational cost.

Basis Set	Cores	MD		
		Cost (kAUs)	Steps run	Steps per kAU
6-31G*	96	34.6	1920	55.5
DZVP	96	34.6	1466	42.4
TZVP	96	34.6	1004	29.0

Table 6: Table showing the number of steps produced for various basis sets.

Table 6 shows that 6-31G is significantly faster than the other two basis sets considered for this project. To differentiate between the accuracy of DZVP and TZVP, single-point energy calculations have been performed as above, this time the functional, PADE, will be kept constant, to allow for the two basis sets to be compared. The data shown in table 7 reinforces the above statement of PADE poorly recreates the experimental values for this system and will not be used for this project.

Molecule	DZVP		TZVP		Published data Energy (hartree)
	Energy (hartree)	error to published data (%)	Energy (hartree)	error to published data (%)	
H	-1.3274	165.48	-1.3271	165.43	-0.5
C	-23.6917	342.26	-23.7674	343.67	-5.3570
N	-38.6175	300.51	-38.6692	301.05	-9.6420
O	-23.6917	50.20	-23.1996	47.08	-15.7739

Table 7: Table showing the comparison of single-point energy calculations to experimental values for various basis sets.

While the error is substantial, this is to compare the two basis sets, the absolute value of the error has no impact on the results here. Both basis sets have similar errors produced, with the PADE functional accounting for the size of the error. This shows both basis sets recreate the experimental results with similar accuracy and thus has little effect on the choice of basis set.

From [85], DZVP has been shown to perform well with energy and geometry calculations, providing accurate results. The same paper has also suggested there is little difference between the performance of DZVP and TZVP. This agrees with the results presented in this thesis.

Taking these results and paper [85] into account the basis set chosen for this project is DZVP, this reduces computational cost while increasing the data generated in the same amount of time. Due to CP2K implementing a PWGTO method a cutoff has been introduced

for the plan wave section, this has taken a value of 300 Ry to produce a fine grid. This was taken from CP2K documentation and has not been chosen by the user. There has been little evidence to suggest that TZVP is worth the additional computational cost for this system when performing geometry optimisations and molecular dynamics. This choice will change when performing vibrational calculations.

3.5 Cell Generation

Construction of the (Ala-Gly)_n SF unit-cell followed the methodology described by Yamane et al [1] and has been described further in the attached paper "Evaluating a Bombyx mori silk fibroin model using a range of atomistic simulation tools". The initial unit-cell was created by arranging four Ala-Gly chains with repeated β -turns according to information from experimental and computational experiments on SF [1,45,46,68,86–88]. To simulate the bulk system of the repeated polymer chain, a periodic boundary condition was implemented, where nitrogen and carbon-terminals were connected to mirror images of themselves and the resulting structure is illustrated in figure 21. The lattice parameters of the unit-cell, shown in figure 21, are orthorhombic: $a = 17.8 \text{ \AA}$, $b = 15.7558 \text{ \AA}$, $c = 11.4904 \text{ \AA}$, the box cell size has been considered in conjunction with the cells published in [67–69], as mentioned in the literature review.

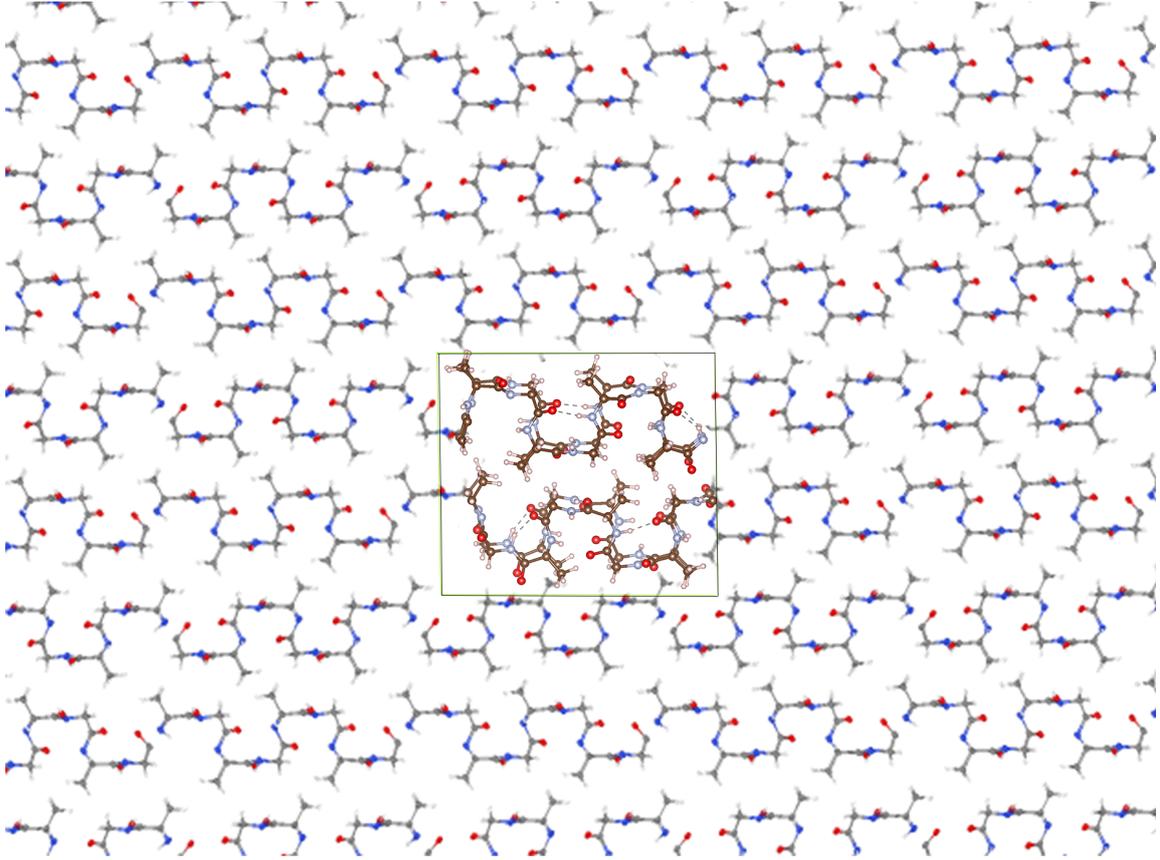


Figure 21: An image showing the SF unit-cell. Created by the Author using Ovito [42].

For this project, multiple cell sizes were investigated to check the best cell size to run on; the cell sizes increase from 272 atoms up to 2176 atoms. The most steps have determined the 'best' cell size ran vs computational cost. There was also a consideration towards which cell will provide accurate results in comparison to the literature and classical mechanics calculations. The smallest possible cell size is shown to be greater than $(\text{Ala-Gly})_6$ [68]. The various cell sizes that are considered are listed in table 8.

Cell label	Number of repeated units			Number of atoms
	x	y	z	
1	1	1	1	272
2	1	1	2	544
3	1	2	1	544
4	2	1	1	544
5	1	2	2	1088
6	2	1	2	1088
7	2	2	1	1088
8	2	2	2	2176

Table 8: Table showing the various cell sizes considered for this project

Figure 22 shows the change in the size of the $1 \times 1 \times 1$ cell compared to the $2 \times 2 \times 2$ cell.

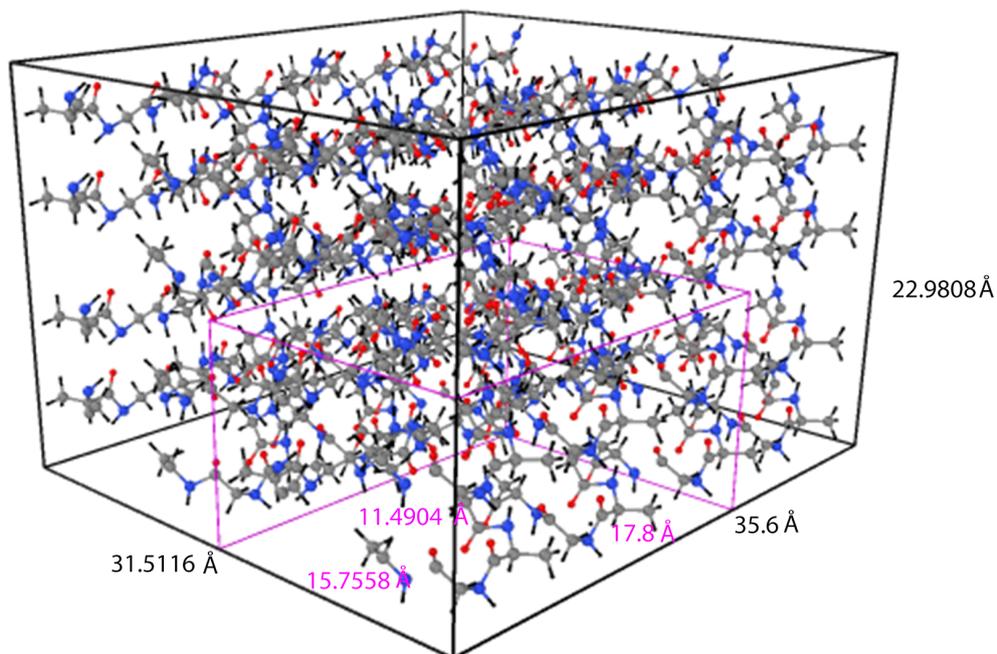


Figure 22: An image showing the box size comparison between the $1 \times 1 \times 1$ cell (shown in pink) and $2 \times 2 \times 2$ cell (shown in black). Created by the Author using Ovito [42].

Each of these cells underwent an AIMD (Ab initio molecular dynamics) calculation, using the method discussed later. Computational cost is considered by how many steps of molecular dynamics can be performed in 24 hours and how many nodes are required. The results are in table 9.

Cell label (refer to table 8)	AIMD			
	Cores	Cost (kAUs)	Steps run	Steps per kAU
1	96	34.6	3183	91.9
2	120	43.2	550	12.7
3	120	43.2	806	20.4
4	120	43.2	746	17.3
5	168	60.48	126	2.1
6	168	60.48	145	2.4
7	168	60.48	146	2.4
8	480	172.8	42	0.2

Table 9: Table showing the computational cost for various cell sizes.

From tables 8 and 9, it can be seen that as the cell size increases the computational cost increases. The cost is determined using the ARCHER cost calculator, based on the number of nodes used and for how long. As the cell size increases, the number of MD steps completed in 24 hours is reduced. Choosing a cell containing only 272 atoms allows for a greater number of calculations to be run. MD calculations are one of the faster methods chosen in this project; the cell optimisation calculations take an increased amount of time and will generate fewer steps in the same period. This further supports the choice to choose the $1 \times 1 \times 1$ unit-cell to perform the majority of the calculations.

Further to the cell size, select cells will be hydrated and seeded with magnesium and choline. An exclusion radius of 1.4 Å has been utilised for both water and magnesium. Extra space has to be created in the cells containing choline, due to the size of the molecule. Both structures have the experimentally accurate ratios of silk to water to choline. The experiment has used values between 7% and 12% for water, this project has used 7.5% and 9% water of the SF molecular weight, depending on the calculation. There has been a lack of literature around the amount of magnesium present in the cell. The experimental ratios of silk to choline are included in [5]. The script to generate the extra space in the cell is included in the supporting information. Choline has been generated from literature in Avogadro then underwent classical minimisation to gain a good starting guess then use DFT to minimise the structure further, with a comparison to literature on the bond lengths as shown in table 10.

Atoms		PBE		BLYP		Published data
Atom 1	Atom 2	Bond length (Å)	Error to published data (%)	Bond length (Å)	Error to published data (%)	Bond length (Å)
O1	C1	1.3747	-4.5340	1.3775	-4.3375	1.4400
C1	C2	1.4140	-3.2190	1.4208	-2.7515	1.4610
C2	N1	1.5419	-1.0962	1.5580	-0.0616	1.5590
N1	C3	1.4115	-5.3300	1.4217	-4.6499	1.4910
N1	C4	1.5617	3.6959	1.5711	4.3207	1.5060
N1	C5	1.4713	-2.4997	1.4740	-2.3174	1.5090

Table 10: Table showing the comparison of choline bond lengths to Published data data (all numbers are rounded to 4dp). Hydrogen bonds have been omitted.

3.6 Calculation Methods

CP2K offers a multitude of calculation methods; the most important for this project are geometry optimisation, molecular dynamics (MD) and vibrational analysis. Each method

will have specific parameters and convergence criteria. In general the basis set and functional will be the same for all calculations (Optimisation, MD and single point energy) however typically for IR calculations the criteria have to be more accurate, and thus more extensive basis sets such as TZV2P will have a beneficial effect.

3.6.1 Optimisation

Geometry optimisation has been completed for each system before any other calculations being performed on it [89]. This method aims to find the lowest energy structure, by optimising/minimising the forces on each atom. This lowest energy structure is the most likely structure to appear experimentally; however, all optimisation methods will find local minima rather than global minima. The optimisation method chosen for this project is the conjugate gradient optimiser. For all minimisations that have been run, there are given criteria which show if the optimisation has been successful or not. The force criteria requires all forces on atoms to change by less than 1×10^{-x} between two steps for the convergence to be successful, where the value of x is chosen for the simulation. A higher value of x produces a more expensive calculation but a more experimentally accurate structure.

It has been shown that the optimisation step is a more computationally expensive method to run, but the benefit of this is it only has to be completed once per structure, i.e. one geometry optimisation produces the optimal structure on which ten different temperatures, NVT or NPT, of MD can be performed. The optimised structure can also be used for the IR spectrum; however, IR spectrum uses a tighter force convergence as IR is sensitive to changes in structure, and the calculation becomes expensive.

Atoms	Time (hours)	Cell optimisation steps	AIMD steps
272	24	96	3183

Table 11: Table showing the computational cost for cell optimisation vs MD.

One method of checking if the cell optimisation has completed is to check the energy convergence; this is shown in figure 23. From this graph, the energy has converged around 350 steps. This energy criterion was set to 1×10^{-3} eV/A, this would require more steps if this convergence criteria were decreased to 1×10^{-6} eV/A or lower. The value of 1×10^{-3} eV/A was chosen due to the faster computational time while still providing accurate results. The self-consistent field, scf, convergence has been set to 1×10^{-5} . This convergence criterion is the same for all cell sizes.

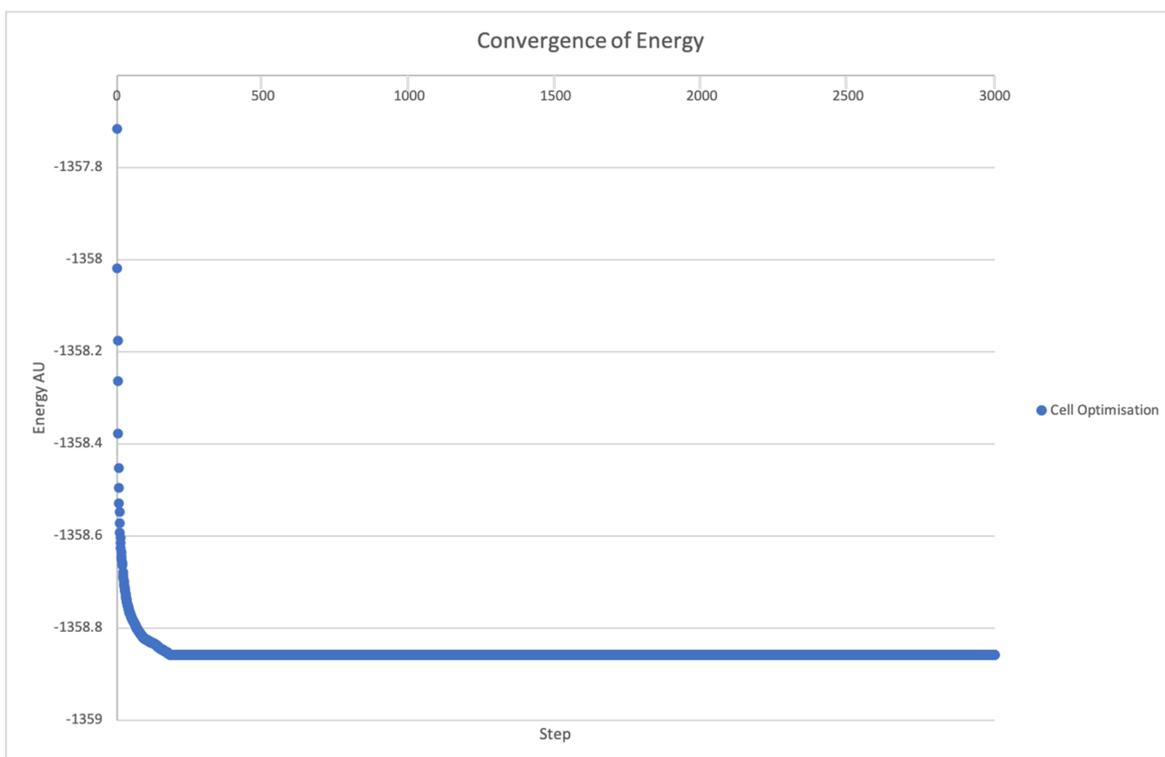


Figure 23: A graph to show the energy convergence of 3000 steps of a cell optimisation run.

The cell optimisation has to be run for each system where there is a change in composition. This has been run for hydrated systems and charged systems with the addition of Mg^{2+} . The energy convergence for both shows the change in the required number of steps. Cell optimisations have been run for all changed systems, but images have not been included here. Due to the complexity of Mg ions, the number of steps performed in the same time frame has been reduced from 3000 to 25. The reduction in the number of steps, shows increasingly complex electron shells have to be accounted for in calculations.

An example geometry optimisation script has been included in the supporting information.

3.6.2 Molecular Dynamics

Molecule Dynamics (MD) is a method to analyse the movement and interaction of atoms, molecules and electrons [90]. The MD is performed by numerically solving Newton's equations of motion, for a system of interacting particles. The forces between the particles and their potential energies are calculated using interatomic potentials or force fields depending on if the calculation is performed by quantum or classical mechanics. In ab initio MD

the Born Oppenheimer approximation is applied. Here the nuclei are fixed in position, the KSDFE equations are solved, the forces are calculated and then the nuclei are moved to the new updated position. This process is repeated until the minimum energy position is gained. Classical MD occurs slightly differently: numerically solving Newton's equations of motion for a system of interacting particles, where forces between the particles and potential energies are calculated using user defined molecular mechanics force fields.

This method allows the simulation to generate trajectories for each atom as-well-as position and force data at every time step. For this method, a time step of 0.05 fs is used, which is universal regardless of the ensemble and temperature used. The short time step was chosen to allow for a detailed view of the atomistic movement. For this project, two separate ensembles have been used: NVT and NPT. NVT is a simulation performed under constant temperature and volume, whereas NPT is performed under constant pressure and temperature. Each ensemble has been run with ten different temperatures: 10 K, 50 K, 150 K, 273 K, 298 K, 310 K, 373 K, 473 K, 500 K, 673 K. The temperatures that are of importance are 273 K - 373 K as these are most likely to include the temperatures used in laboratories or the test environment. The pressure of all NPT calculations is 1 bar; this has not been changed as the temperature dependence has been investigated. Outside of this range, the structure of SF would be predicted to undergo phase transitions - at 10 K the material would be a crystal and at 673 K SF would most likely be degraded; however, all temperatures have been run to provide a complete comparison for the generated Ramachandran plots. For NVT a Nose Hoover thermostat has been used, and for NPT the standard barostat has been used [91], in CP2K, there is no choice of which barostat is implemented. The relaxation time for both the NVT and NPT simulations is 60 fs.

Before MD is run on any system, 3000 steps (or until convergence has been reached) of cell optimisation has to be run to give the lowest energy system and a good starting guess for MD. The number of MD steps that are performed on any single system varies depending on the required information. If the calculation is used to generate Ramachandran plots 3000 steps of MD are run, but for the diffusion path of water a longer time frame is required, between 20000 steps and 50000 steps giving a time frame of between 1000 fs and 2500 fs. A comparison of the effect of temperature on the computational cost has been made but has provided no substantial evidence for a positive or negative correlation. This can be seen in both table 12 and figure 24.

Temperature (K)	AIMD					
	Cores	Cost (kAUs)	NVT steps run	Steps per kAU	NPT steps run	Steps per kAU
10 K	96	34.6	3184	92.1	2541	73.5
50 K	96	34.6	3472	100.5	2584	74.8
150 K	96	34.6	2620	75.8	3083	89.2
273 K	96	34.6	3207	92.8	2781	80.5
298 K	96	34.6	3183	92.1	3080	89.1
310 K	96	34.6	4834	139.9	3085	89.3
373 K	96	34.6	2975	86.1	3080	89.1
473 K	96	34.6	2768	80.1	3073	89.9
500 K	96	34.6	3164	91.6	2374	68.7
673 K	96	34.6	3253	94.1	3106	89.9

Table 12: Table showing the computational cost for various temperatures.

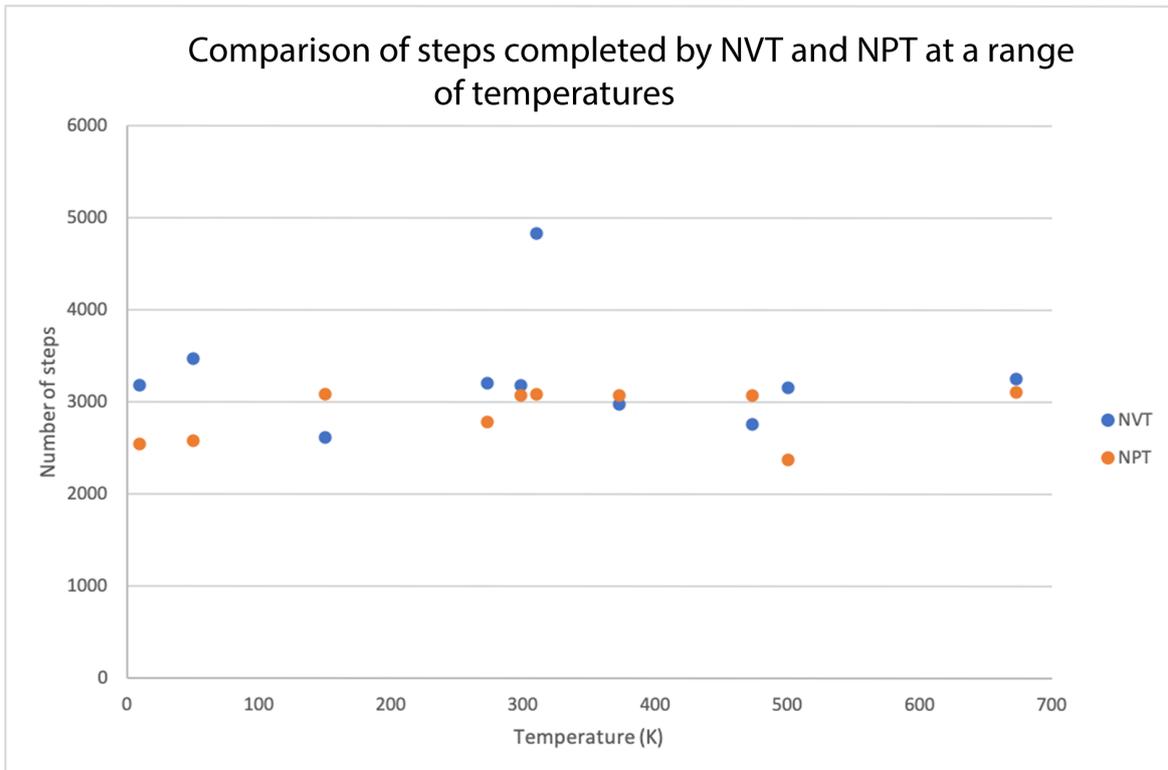


Figure 24: A graph to show the number of steps completed by NVT and NPT at a range of temperatures.

From figure 24, it can be seen that NVT regularly completes more steps than NPT and hence is, in general, less computationally expensive. However, there appears to be no correlation between temperature and computational cost for either NVT or NPT. The

calculations have used the same DFT+D3, BLYP, DZVP method as described above.

For MD, the print level can be specified. This controls the amount of information generated at each step; a higher print level requires greater computational power. Electron density analysis is performed at the highest level, which is very computationally expensive. This cost can be reduced by performing the analysis on the last frame of an MD calculation. Generating position data is the standard method and is relatively cheap. An example MD script has been included in the supporting information.

3.6.3 Vibrational Analysis

The vibrational analysis looks at the type of motion a molecular bond undergoes - there are three main types -rotation, translation and vibration. The vibrational analysis focuses on the vibration of each bond. This technique consists of measuring the vibration and frequency of each bond; this is referred to as normal modes. Several different techniques can be used, including IR and Raman. The experimental method of IR is carried out by bombarding a sample with light to generate absorption spectra.

Several papers are performing IR and Raman experimentally on the *B. Mori* SF; however, there is little to no computational generation of IR spectra for the simplified SF structure used throughout this project. This project has used DFT to generate the IR spectra [92].

This method to generate IR spectra varies widely from previously discussed geometry optimisation and molecular dynamics; this means the basis set and functional have been re-selected. The input script has been discussed with a leading expert in this field. This has led to TZV2P and PBE being used for all IR spectra. TZV2P has been used as IR relies on how the bonds act when interacting with IR light and hence requires a more sensitive functional. No comparison of basis sets and functionals have been carried out for IR spectra due to the high computational cost.

Through the use of DFT, the IR that can be run through geometry optimisation. This choice occurs as they affect the motion of the bonds throughout the simulation. The starting structure has undergone 3000 steps of cell optimisation to achieve an accurate starting structure. Performing the IR coupled with molecular dynamics allows for a choice in temperature or pressure dependent on the ensemble chosen. This will also show how the structure changes with temperature; there have been studies into this with other polymers but not simplified SF [93]. This will show how the structure changes between the phases if the temperature range is great enough. The geometry optimisation choice will allow for a tighter convergence to be reached, which is not possible through MD. The MD plots have been generated via 3000

steps of geometry optimisation, then performing 3000 steps of MD at a given temperature then 100 steps of IR to produce the IR spectra at the correct temperature. This output is then checked for negative frequencies to determine the validity of the IR spectrum.

For generating the IR plots, from the computational output, Molden has been used [94]. This allows for assigning peaks to bonds and to structures within the silk cell, including β turns and α helices. This will then be compared to Ramachandran plots to show if the data is in agreement; this is discussed in further detail later in the report.

An example script of Vibrational analysis IR has been included in supporting information.

3.7 Ramachandran Plots

The Ramachandran plots are a method of validating the secondary structure of proteins, as covered in the literature review [58]. There are several steps to generate the Ramachandran plot:

Step 1 geometry optimisation is run for 3000 steps or until convergence has been reached to gain an accurate starting structure. This is then used to perform molecule dynamics under multiple different ensembles and temperatures ranging from 10 K to 673 K as this will allow the structure to vary. The MD is run for around 3000 steps or 24 hours to produce a final structure; the cell optimisation and MD are performed as stated above. Once the pdb file has been generated, VMD is used to generate ϕ *vs* ψ data; this data is then plotted on an empty Ramachandran plot and merged to produce a final Ramachandran plot. The Ramachandran plots have been generated via DFT for both hydrated and non-hydrated states (experimentally accurate at water is 9% of the SF molecular weight).

For classical mechanics, Ramachandran plots have been created by Mathew Haskew. A brief overview of his method has been included here. The simulations of periodic (Ala-Gly)_n crystal were visualised using the Visual Molecular Dynamics (VMD) programme. The secondary structure of the periodic (Ala-Gly)_n crystal has been evaluated, the torsion angles of the residues were deduced using DL_ANALYSER. A script was written to fabricate Ramachandran contour plots (using the torsion angles of the Ala and Gly residues) to be then used for comparison with Ramachandran plots of SF/Ala and Gly. Various Ramachandran contour plots were produced for the hydrated and non-hydrated states at a temperature range of 10-473 K. Overlays have been produced with the DFT data on top of the classical mechanical plots, which will be shown later in the results. This will also be compared to experimental data produced for SF.

3.8 Software

There are a multitude of different software packages available to simulate proteins in varying physical states. This section will only focus on quantum mechanical codes due to the classical mechanics have been performed by Mathew Haskew. From looking at the literature of quantum mechanical simulations, three principal codes have been considered for this project - VASP, CP2K and Gaussian [95–97]. The codes are considered for general calculations instead of calculations on *B. Mori* silk due to the availability of literature.

The three codes considered in this review are due to the range of calculations, availability or speed of the code. All three can simulate protein systems. Gaussian can perform simulations of geometry optimisations, IR and Raman spectroscopy [98,99]. The additional benefit of using Gaussian is that it includes a wide range of basis sets and functionals for use to a much higher degree than other software packages. For example, Gaussian can run hybrid functionals such as B3LYP whereas it would not be feasible to run hybrid functionals with CP2K especially for SF.

CP2K was sufficient for this project; the ability to run various temperatures and pressures allowed for a range of states of material to be simulated, which is useful for the material used in the project. One central part of CP2K is that it allows for a variety of methods to be run either quantum or classical mechanics to run geometry optimisation, molecular dynamics and more.

The Vienna Ab initio Simulation Package (VASP) can perform similar to CP2K with both geometry optimisation and molecular dynamics available. However, it does not have the same range of functionals that Gaussian contains. There are more sophisticated options available such as simulating magnetism in materials and the presence of Phonons. However, this goes beyond the requirement of this project.

Considering the software available, this removed VASP as an option for the project, due to VASP requiring a license. CP2K is open source, and the university holds a license for Gaussian. The next consideration was the resources available, the majority of the calculations will require computational resources with access to both Lancaster University HEC and ARCHER. [100] has shown that Gaussian does not scale efficiently on supercomputers as it can not run in parallel across multiple nodes; this was confirmed through conversations with the ARCHER maintenance team. This limited the usability of Gaussian for this project but can be used for validation of CP2K. This made the use of CP2K the right choice due to:

- The flexibility of the software

- Availability of basis sets and functionals
- Ease of use
- Speed of calculation
- Range of calculation types

4 Results and Discussion

This section has collated the results in several different methods, such as graphs and figures. The calculations used to generate these results follow the methods as shown above. There will be a comparison of the generated results to literature and classical mechanical results where possible, but there is a lack of experimental data for a range of the calculations provided here. The results generated support three main aims - if the DFT produces experimentally accurate results, elucidating the secondary structure, and investigating the behaviour of various molecules in SF.

These results aim to elucidate the structure and behaviour of SF. This should go towards helping to understand how the SF will behave when fully hydrated and present in the body. As mentioned previously the silk is used in multiple methods in the battery, for both the electrolyte and electrodes. This project focuses on the electrolyte and how the electrolyte will behave with the addition of choline, water and ions. The structure will have an optimal temperature for it to work in the body, and this can be seen from a range of calculations that are performed. The breakdown of the silk structure is outside of the remit of this project but will be considered in future projects.

4.1 Preliminary Investigation of the SF Structure

The first results generated for this project concern performing comparative simulations to determine the basis set chosen for the project, as discussed in the methodology. The basis set will be chosen based on how well the hydrogen bonds are recreated in the simulation, shown by the dashed black line. The comparison is shown in figure 25

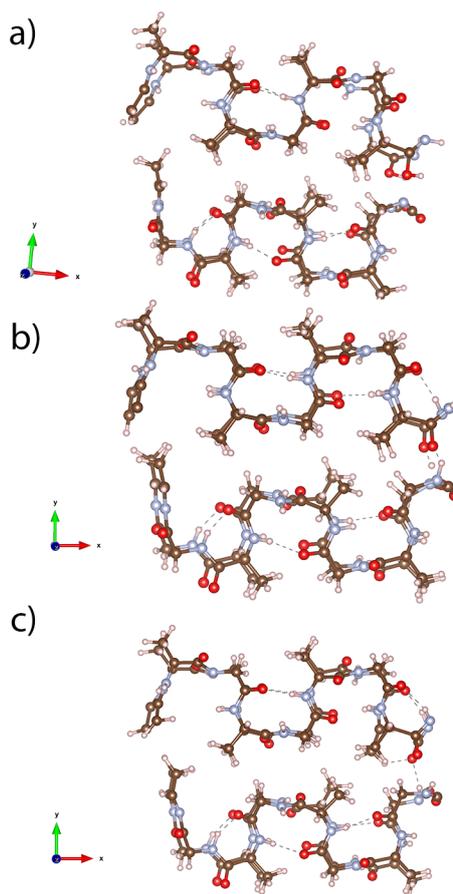


Figure 25: An image to show the change in silk structure when different basis sets are used. Created by the Author using Vesta [79]. a) shows the 6-31G basis set, b) shows DZVP and c) shows TZVP.

When the structures are inspected, figure 25, there is an apparent lack of hydrogen bonds, in 6-31G, in several areas which is insufficient for this project. For the other two basis sets, there are a sufficient number of hydrogen bonds in the expected areas, and thus will not make a considerable difference to the basis set chosen. The hydrogen bonds are not forming in 6-31G due to the incorrect alignment of the H-O and H-N bonds; this can be hard to visualise in the images. Polarisation effects that would be described by d-functions on heavy atoms and p-functions on H atoms are not included in the expansion of the MOs via 6-31G and thus it will not describe hydrogen bonds. As shown in the methodology, the basis set DZVP has been chosen for the majority of calculations.

Next, the unit cell had to undergo geometry optimisation to obtain the structure used for successive calculations; this has been discussed in the methodology. Figure 26 shows how the box size has varied before and after 3000 steps of cell optimisation on $(-\text{Ala-Gly-})_{16}$. The figure also shows how the x,y and z vectors have changed in length, this data is also included in table 13 for both hydrated and non-hydrated states. From the optimised structure it can

be seen that the β -turn type II structure is stabilised by 4 to 1 intramolecular hydrogen bonds.

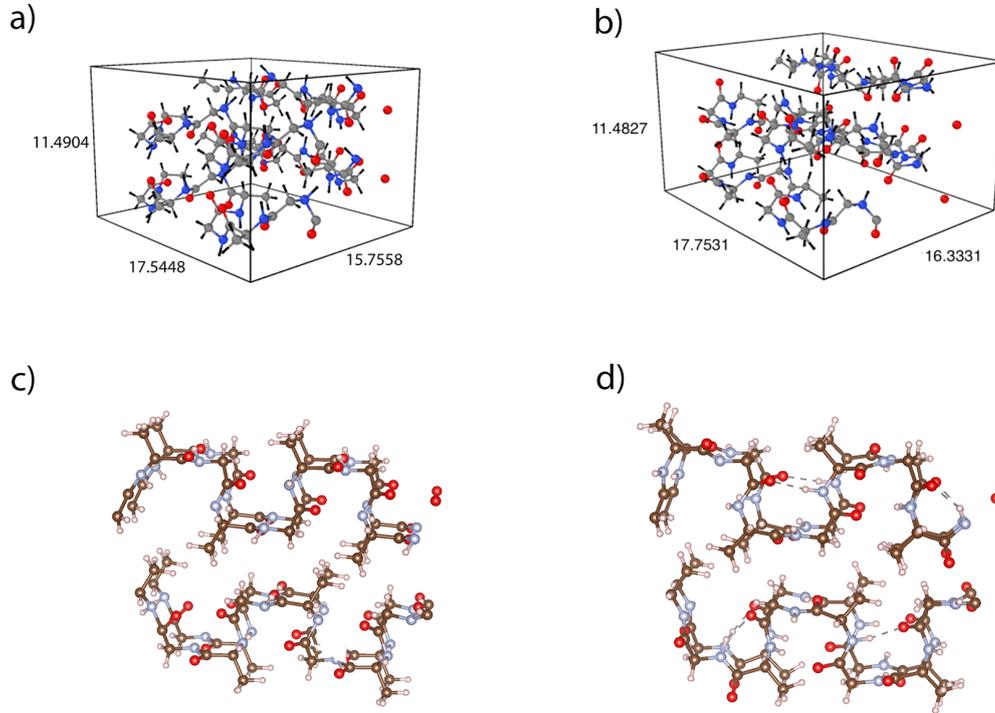


Figure 26: An image to show the change in silk structure after undergoing 3000 cell optimisation steps. Created by the Author using Vesta [79] and Ovito [42]. Where a and b shows the change in box size with a) being pre-cell optimisation and b) being post cell optimisation. C and d show the change in hydrogen bonding before and after cell optimisation with c) being pre-cell optimisation and d) being post cell optimisation.

Cell State	x lattice parameter(\AA)	y lattice parameter(\AA)	z lattice parameter(\AA)
Non-hydrated pre-optimisation	17.8	15.76	11.49
Non-hydrated post optimisation	17.75	16.33	11.48
Hydrated pre-optimisation	18.56	16.12	10.63
Hydrated post optimisation	19.38	16.85	11.66

Table 13: Table showing the lattice parameters for the SF as both hydrated and non-hydrated cells.

Using the optimised structure, MD calculations can be performed. Figure 27 shows how MD affects the structure. This simulation has a single water molecule included to see the behaviour over a short time, this also highlights if further simulations would prove useful or

if the water is simulated incorrectly. The change in the structure after 20000 fs is shown in figure 27.

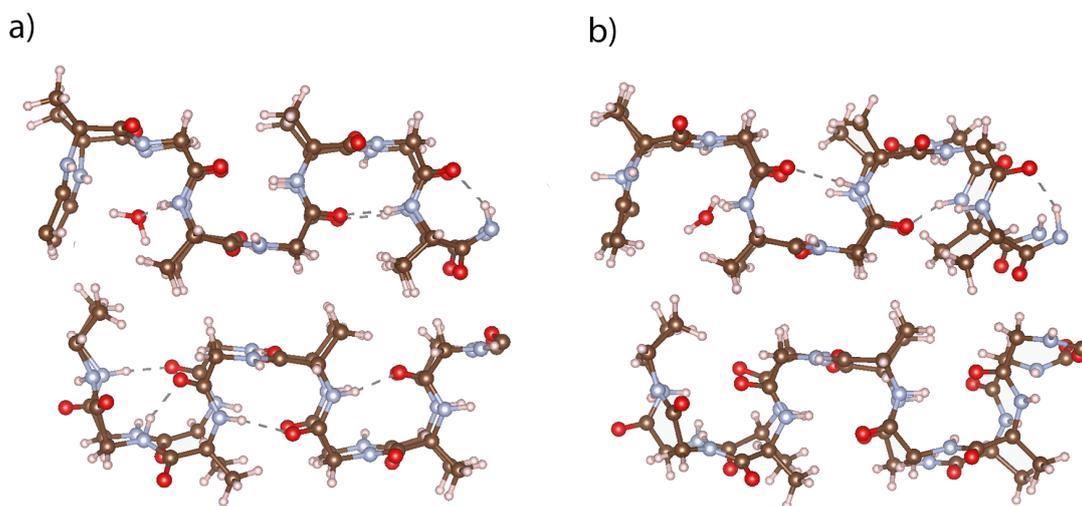


Figure 27: An image to show the change in silk structure after undergoing 20000 MD steps, both taken from the z view. Created by the Author using Vesta [79]. Where a) shows the silk cell pre-MD and b) shows the silk cell post MD.

4.2 Ramachandran Plots

Ramachandran plots have been generated for both DFT and classical methods. The plots have been generated in the method as described above. Where the NVT plots have been generated from the 3000 steps of MD using a geometry optimised structure as the starting structure. A complete set of Ramachandran plots have been generated for the $(-\text{Ala-Gly-})_{16}$ structure, in temperature ranges of 10 K to 673 K, NVT and NPT ensembles and both hydrated and non-hydrated structures. For the $(-\text{Ala-Gly-})_{32}$ structure only two Ramachandran structures have been generated at 298 K, NVT but hydrated and non-hydrated structures have been included. The classical Ramachandran plots have been generated for both $(-\text{Ala-Gly-})_{128}$ and $(-\text{Ala-Gly-})_{1024}$ structures, which have 2176 and 17,408 atoms, respectively. There are Ramachandran plots for temperatures 50 K to 673 K, for both ensembles and non-hydrated and hydrated states. The DFT plots are shown as dot plots, whereas the

classical plots are contour plots, this is a direct result of both the method of generation and the number of data points generated for each cell size.

The Ramachandran plot consists of three main areas β sheets, left-handed α helix and right-handed α helix. β sheets have been shown to occur in the region $\varphi = -60$ degrees, $\psi = 130$ degrees, left-handed α helix's are shown to appear in the region of $\varphi = 70$ degrees, $\psi = 10$ degrees and right-handed α helix in $\varphi = -40$ degrees, $\psi = -60$ degrees. These are shown visually in figure 28.

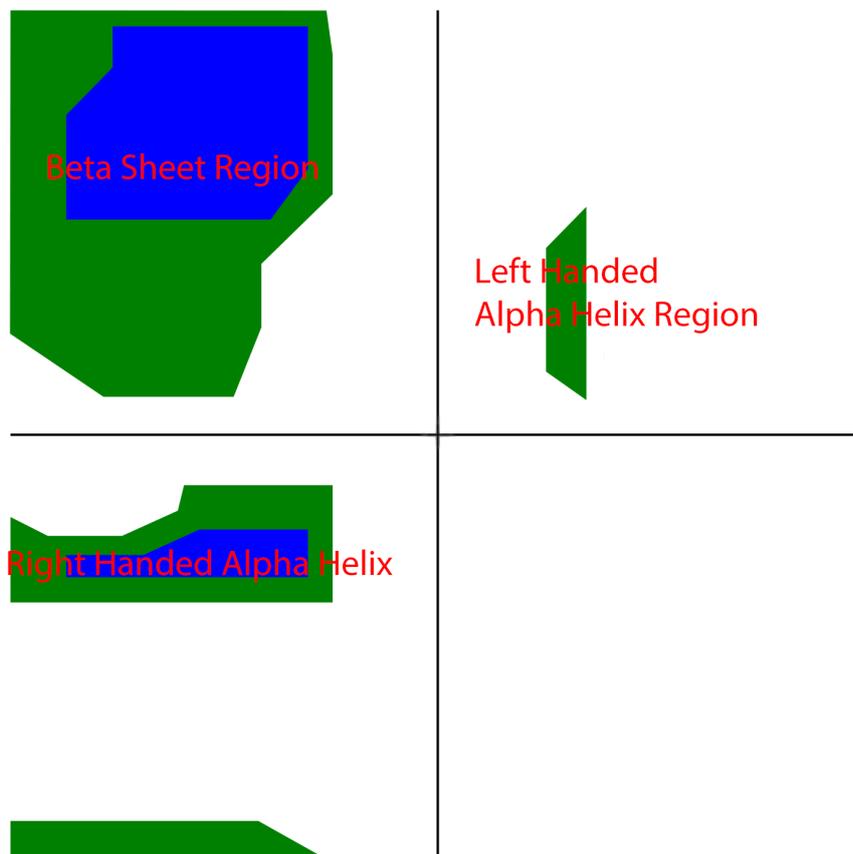


Figure 28: A figure showing the labelled areas of an empty Ramachandran plot. Created by the Author using VMD [101].

For clarity, the DFT Ramachandran plots shown here will include both ensembles, both hydrated and non-hydrated versions but only for 10 K, 298 K, 310 K and 673 K. This reduced data set will allow for an apparent change in structure to be shown with the main focus of the structure around operating temperature (298 K and 310 K). For completeness, all Ramachandran plots are included in the supporting information.

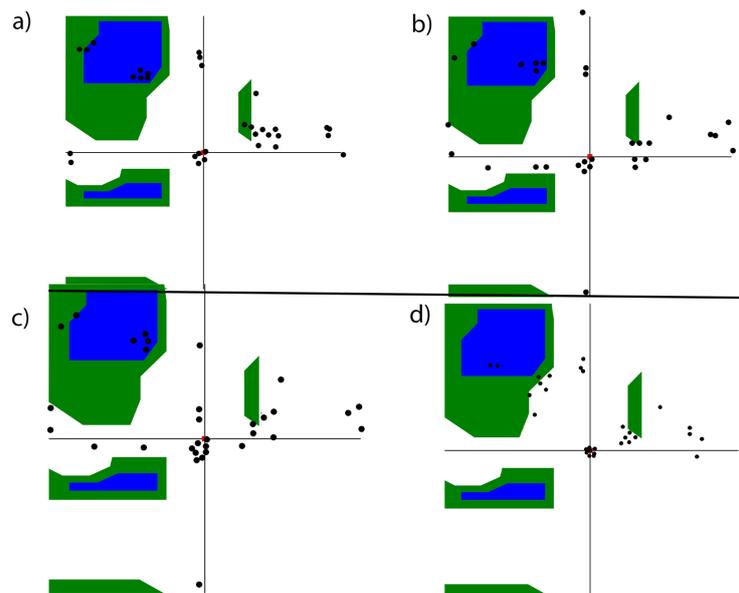


Figure 29: An image to show the Ramachandran plots at 10 K. Created by the Author using VMD [101]. a) Hydrated cell run at 10 K with NPT. b) Non-Hydrated cell run at 10 K with NPT. c) Hydrated cell run at 10 K with NVT. d) Non-Hydrated cell run at 10 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).

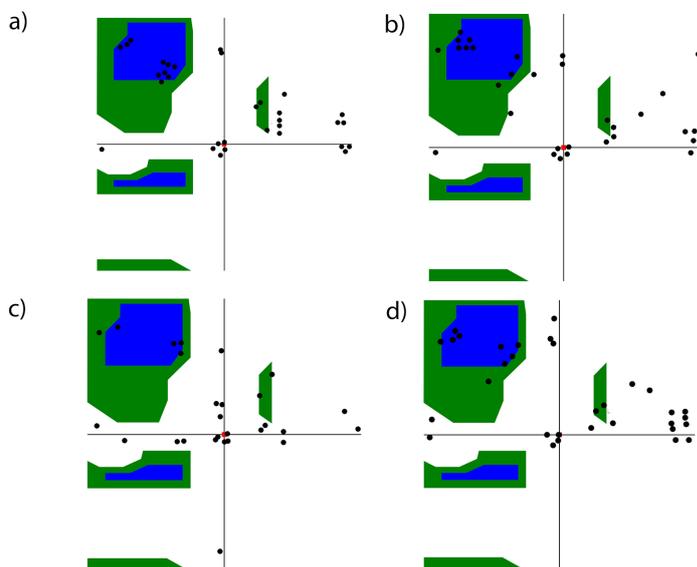


Figure 30: An image to show the Ramachandran plots at 298 K. Created by the Author using VMD [101]. a) Hydrated cell run at 298 K with NPT. b) Non-Hydrated cell run at 298 K with NPT. c) Hydrated cell run at 298 K with NVT. d) Non-Hydrated cell run at 298 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).

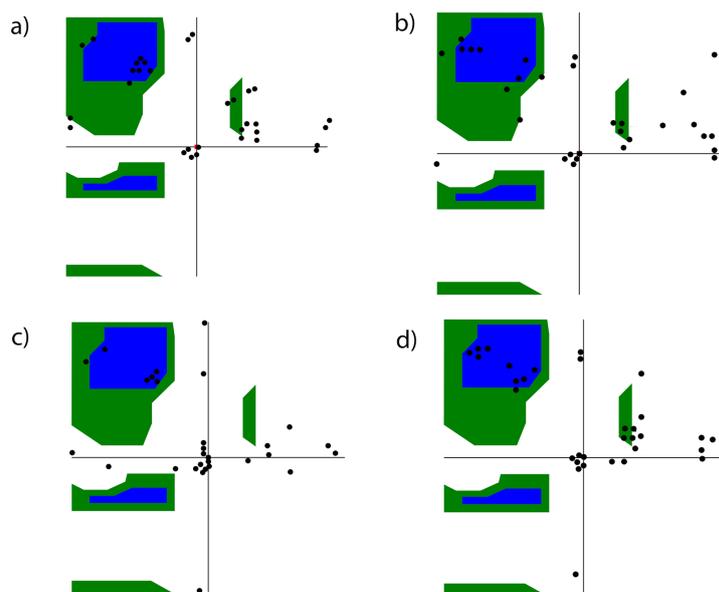


Figure 31: An image to show the Ramachandran plots at 310 K. Created by the Author using VMD [101]. a) Hydrated cell run at 310 K with NPT. b) Non-Hydrated cell run at 310 K with NPT. c) Hydrated cell run at 310 K with NVT. d) Non-Hydrated cell run at 310 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).

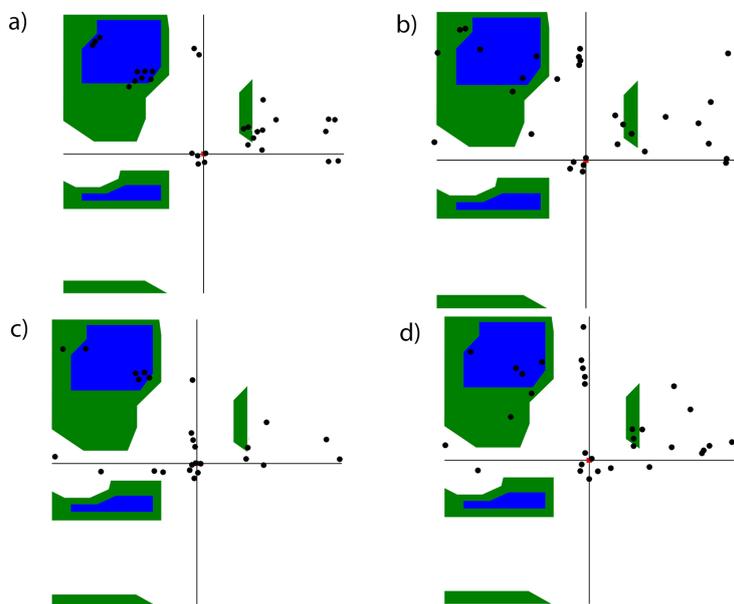


Figure 32: An image to show the Ramachandran plots at 673 K. Created by the Author using VMD [101]. a) Hydrated cell run at 673 K with NPT. b) Non-Hydrated cell run at 673 K with NPT. c) Hydrated cell run at 673 K with NVT. d) Non-Hydrated cell run at 673 K with NVT. The black dots show data points, and the red dot shows the origin (may be covered by a black dot).

It is worth noting that in figures 29 to 32, the red square shows the origin and is not a

data point. For the Ramachandran plots shown in figures 29 to 32, there are several trends shown. As the temperature increases, the torsion angles become more widespread but appear in similar locations indicating the backbone relaxes at a higher temperature but does not break the chain or dramatically alters the structure. This is true for all hydrated and non-hydrated states, the only exception is the hydrated cell run at 673 K with NVT ensemble this has undergone no significant change except a few torsion angles have changed position but in far fewer numbers than seen in the other plots. This is expected behaviour as 673 K will be approaching the vaporisation temperature of the polymer and thus will contain more energy and this will allow for more movement of all particles. This change in structure has been seen experimentally as the difference in the Ala and Gly residue's positions were exacerbated for higher temperatures.

Similarly, as each system is hydrated, the torsion angles become much more regular and localised regardless of the temperature of the simulation. This will be a result of the introduction of water molecules which, will allow for more hydrogen bonds to be formed and thus restricting the degrees of freedom of each atom and therefore reducing the range of angles each chain can form, both between polymer backbone chains and between the polymer and water molecules. The number and strength of hydrogen bonds formed will be a result of the D3 correction, basis set and functionals chosen previously in this project. As the calculations change from NVT to NPT, there is little evidence to show a constant trend which can only be attributed to the change in the ensemble. The main effects appear to be derived from the temperature change and the hydrated state of the cell.

A comparison of DFT to classical mechanics has been performed; only 310 K has been shown in figures 33 and 34.

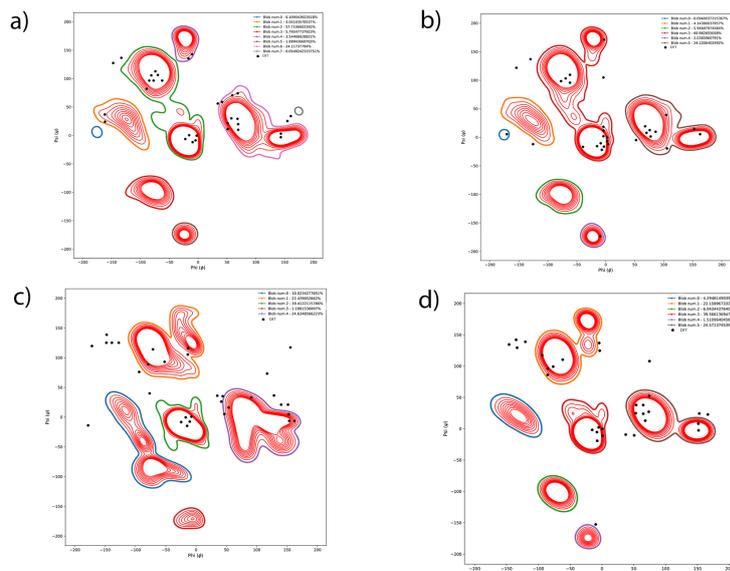


Figure 33: An image to show the Ramachandran plots at 310 K in the $2 \times 2 \times 2$ cell. a) Hydrated cell run at 310 K with NPT $2 \times 2 \times 2$ cell compared to DFT. b) Hydrated cell run at 310 K with NVT $2 \times 2 \times 2$ cell compared to DFT. c) Non-Hydrated cell run at 310 K with NPT $2 \times 2 \times 2$ cell compared to DFT. d) Non-Hydrated cell run at 310 K with NVT $2 \times 2 \times 2$ cell compared to DFT. The black dots show the DFT data points, and the contour plot shows the classical mechanical data.

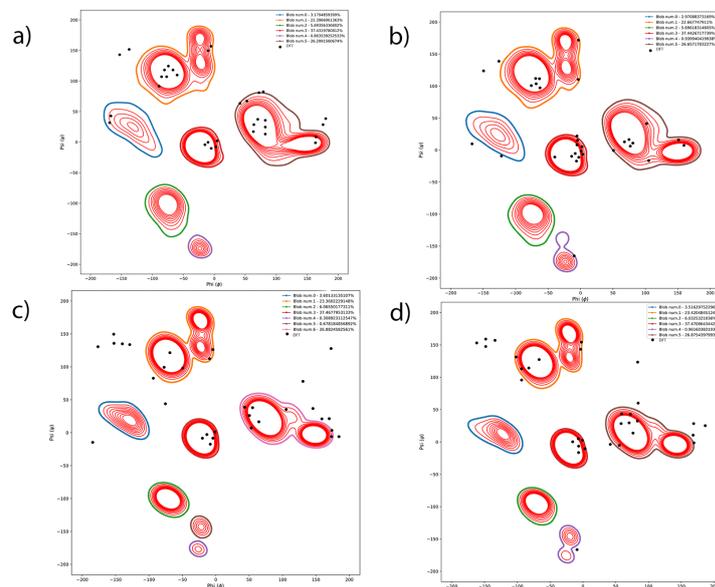


Figure 34: An image to show the Ramachandran plots at 310 K in the $4 \times 4 \times 4$ cell. a) Hydrated cell run at 310 K with NPT $4 \times 4 \times 4$ cell compared to DFT. b) Hydrated cell run at 310 K with NVT $4 \times 4 \times 4$ cell compared to DFT. c) Non-Hydrated cell run at 310 K with NPT $4 \times 4 \times 4$ cell compared to DFT. d) Non-Hydrated cell run at 310 K with NVT $4 \times 4 \times 4$ cell compared to DFT. The black dots show the DFT data points, and the contour plot shows the classical mechanical data.

The DFT and classical are, in general, in agreement, with figures 33 and 34 showing the comparison. DFT lacks a few areas that are in the classical mechanics' method these areas are around $\varphi = -70$ degrees, $\psi = -100$ degrees. In comparison, DFT contains angles in the region of $\varphi = -160$ degrees, $\psi = 150$ degrees. This may be a result of the cell size as DFT has been generated using $1 \times 1 \times 1$ with classical using $2 \times 2 \times 2$ or larger. The larger cells will contain more intra-molecular bonds in the starting position that will be unable to form in the smaller cell. The same trends that have been identified earlier are still relevant in classical mechanics. The other trend that can be seen is that as the cell size is increased, the torsion angles become more localised in each area on the Ramachandran plot. This localisation may be a result of the method of plotting as a larger cell removes some of the torsion angle areas previously seen in the smaller cells. Another difference between DFT and classical is that DFT has been plotted for a single time frame, whereas classical has been plotted over an entire history file. This time frame may have a significant effect on the validity of DFT, but the results shown above look sufficient for this project.

When the Ramachandran plots generated in this work and the classical mechanical plots are compared to literature values, the DFT data recreates the data more accurately. This has been shown where the DFT plots generate data in regions where classical mechanics lacks data. The DFT plot can also be compared to figures 10 and 11. This shows the agreement of the results where the data lines up with the allowed regions.

This agreement between DFT and experiment indicates the simulation method and structure are both valid. The structure is shown to be realistic where no torsion angles are present in forbidden areas, but are mainly grouped in expected β and α regions. Furthermore, the SF crystal models possess a heterogeneous structure, evidenced by a left-handed α -helix, 3_{10} -helix, β -sheet (ca. $\varphi = 70$ degrees and $\psi = 10$ degrees, ca. $\varphi = -40$ degrees and $\psi = -30$ degrees, ca. $\varphi = -60$ degrees and $\psi = 130$ degrees, respectively) and random coil structures, which are consistent with the literature [102]. Hereby, implying the SF model utilised in this work, possesses qualities that have been experimentally observed. The $(\text{Ala-Gly})_{16}$ SF crystal structure is to be considered in the silk I form (i.e. repeated β -turn type II conformation), because β -sheets are not the predominant secondary structure, instead, 3_{10} -helix is the predominant secondary structure (ca. 37%), agreeing with the literature [1,45,46,86–88,102].

4.3 Thermal Expansion

From MD simulations, the thermal expansion can be calculated. This was completed as it goes towards improving the understanding of how SF behaves at various temperatures, this

will also allow for a comparison between hydrated and non-hydrated SF. This will further the understanding of the structure as it may indicate that one structure is favoured over the other. Experiments are performed at temperatures from 10 K to 673 K, which allows for the thermal expansion to be seen over a wide range. This thermal expansion is examined in both non-hydrated and hydrated systems. This has only been performed for the NPT calculations as NVT keeps the volume constant and will not provide reliable data. The comparison cell size for the hydrated system is $19.64 \text{ \AA} \times 16.90 \text{ \AA} \times 11.5107 \text{ \AA}$ and $17.75 \text{ \AA} \times 16.33 \text{ \AA} \times 11.48 \text{ \AA}$ for the non-hydrated system. The volumes for each temperature and the hydrated and non-hydrated cells are shown in table 14. The x,y and z vectors have been generated by omitting the first 1000 steps of MD for each temperature to produce an average vector. This is a more accurate method than using the last frame as the expansion will vary with time, shown in figure 35, only one temperature and lattice parameter has been shown for clarity.

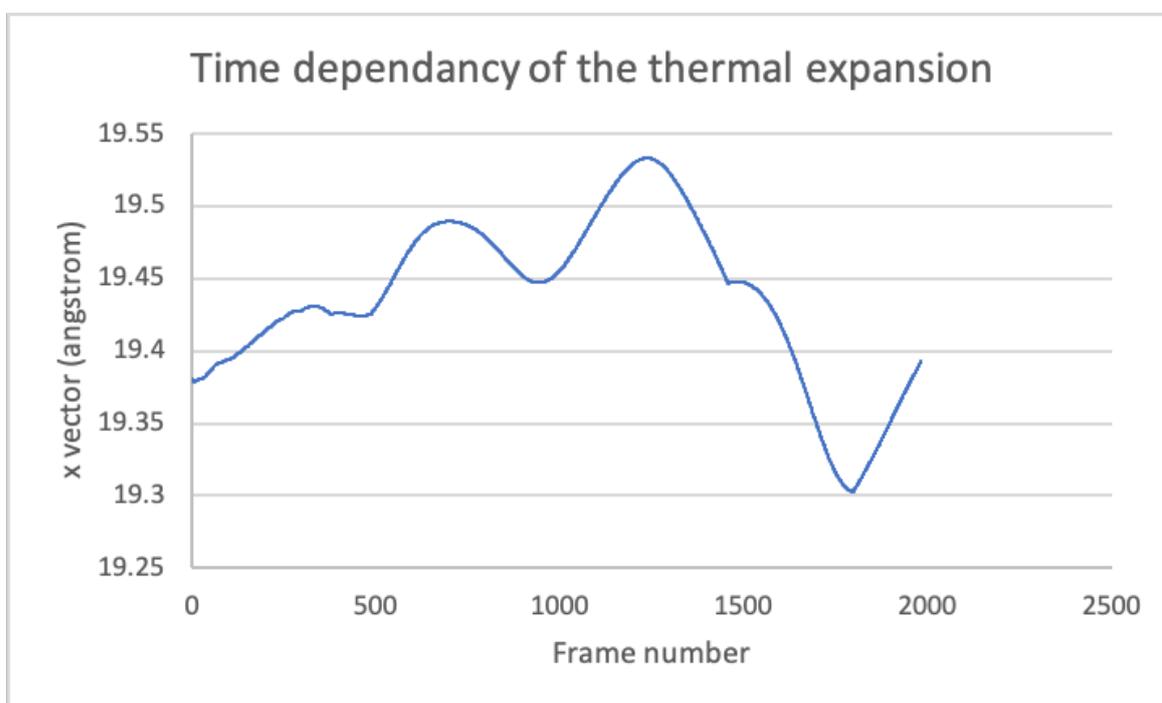


Figure 35: A graph to show the time dependency of the thermal expansion at 150 K of the hydrated system, the last 1500 frames are shown.

Temperature (K)	Cell volume of hydrated cell(\AA^3)	Cell volume of non-hydrated cell(\AA^3)
10 K	4174.69	3460.87
50 K	4580.52	3596.90
150 K	4187.18	3734.16
273 K	3966.73	3725.12
298 K	3964.86	3784.93
310 K	3967.71	3791.91
373 K	3961.37	3775.42
473 K	3913.97	3854.65
500 K	3750.67	3695.40
673 K	4023.21	3939.70

Table 14: Table showing the cell volume at various temperatures.

This is shown in figures 36 and 37. In the graphs showing the change in volume the reference cell is shown by volume equals 0, the other graphs have the reference cell included on the graph.

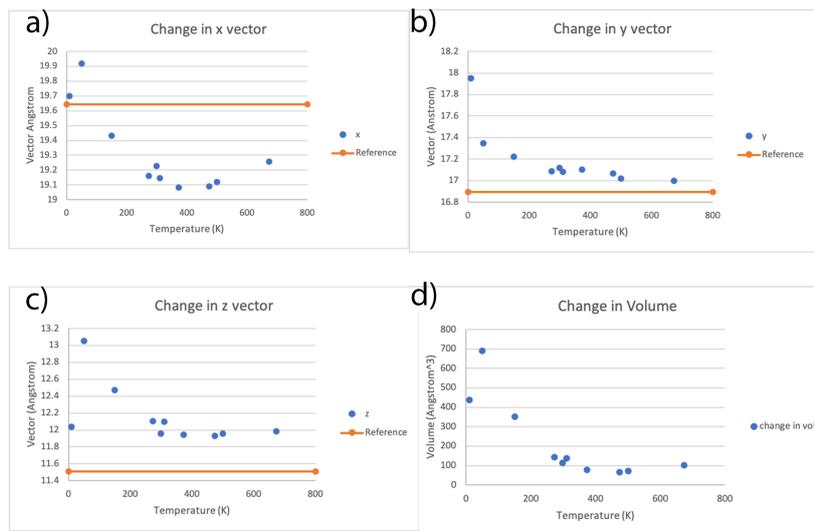


Figure 36: Images to show the thermal expansion of a hydrated cell. a) A graph to show the change in x vector of a hydrated cell at various temperatures. b) A graph to show the change in y vector of a hydrated cell at various temperatures. c) A graph to show the change in z vector of a hydrated cell at various temperatures. d) A graph to show the change in volume of a hydrated cell at various temperatures.

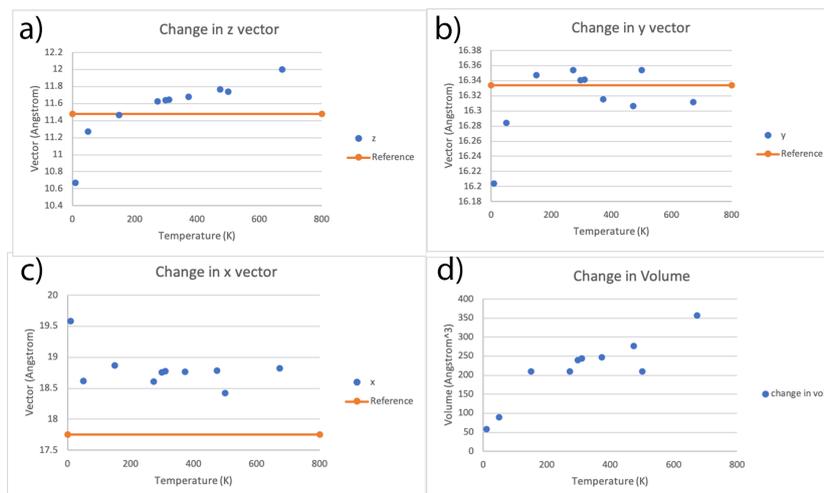


Figure 37: Images to show the thermal expansion of a non-hydrated cell. a) A graph to show the change in x vector of a non-hydrated cell at various temperatures. b) A graph to show the change in y vector of a non-hydrated cell at various temperatures. c) A graph to show the change in z vector of a non-hydrated cell at various temperatures. d) A graph to show the change in volume of a non-hydrated cell at various temperatures.

For the hydrated system, all vectors appear to decrease as the temperature is increased. In contrast, the vectors of the non-hydrated system appear to either increase or have little to no change as the temperature is increased. The volume for the non-hydrated system increases as expected when the temperature increases. However, this is not true for the hydrated cell. For the hydrated system, the volume decreases as the temperature increases; this is unexpected for this system. There are a few reasons for the volume decreasing with the presence of water. The water molecules will form hydrogen bonds which will change the size and restrict the freedom of movement of the silk, whereas fewer hydrogen bonds will be present in the non-hydrated system and this will allow for the silk to expand more freely. Dehydrated SF is not natural and will not be found in nature. Natural silk is likely to undergo thermal contraction as the temperature rises. This is in agreement with [103], where the SF contracts up to 175 degrees celsius or 448.15 K. There is a slight point of inflexion in figure 36.d where the volume at 673 K has increased compared to 473 K and 500 K. The data in [103] has been produced from experiment with a temperature range of 273.12 to 473.15 K; the results produced in this thesis cover a wider temperature range.

The thermal expansion coefficients can be generated using equation 16 [104], are shown in table 15. Where V is the current volume, V_0 is the volume at $t=0$ fs, α is the thermal expansion coefficient, and T is the temperature change in K. The thermal expansion is

temperature-dependent and material dependent, meaning this data can not be compared to experimental values as the experimental structure is more complicated than the simulated SF.

$$V = V_0(1 + 3 \times a \times T) \quad (16)$$

Temperature (K)	Thermal expansion of hydrated cell (1/K)	Thermal expansion of non-hydrated cell (1/K)
10 K	0.003809	0.0005811
50 K	0.001080	0.0001749
150 K	0.0001735	0.0001363
273 K	4.174E-05	7.249E-05
298 K	3.227E-05	7.572E-05
310 K	3.732E-05	7.378E-05
373 K	1.766E-05	6.193E-05
473 K	1.214E-05	5.466E-05
500 K	1.211E-05	3.881E-05
673 K	1.311E-05	5.002E-05

Table 15: Table showing the thermal expansion coefficients at various temperatures.

From table 15, it can be seen that the hydration state of the system has little effect on the thermal expansion coefficient. There is a larger difference seen at lower temperatures; this would be due to the physical state change where water would be crystalline below 273 K, and this will have a large impact on the structure of the SF. At higher temperatures, the difference will become less evident as all thermal expansions are of the same order, E-05.

4.4 Spectra

Infrared (IR) spectra have been calculated for the simplified SF structure. The motivation for this is to see how the structure will change at various temperatures, additionally it provides a comparison to experimental data and the Ramachandran plots calculated in this project. This lets the project determine the validity of the assumed structure. IR should be calculated over a range of temperatures from 10 K to 673 K. This is completed at the same temperatures as the Ramachandran plots to give a comparison between the two data sets. The IR should provide confidence in the Ramachandran plots as it will allow for each peak to be assigned the secondary structure present in the plot. As a further comparison, the IR spectra of alanine and glycine have been generated in the same method. The spectra are shown in figures 38, 39 and 40.

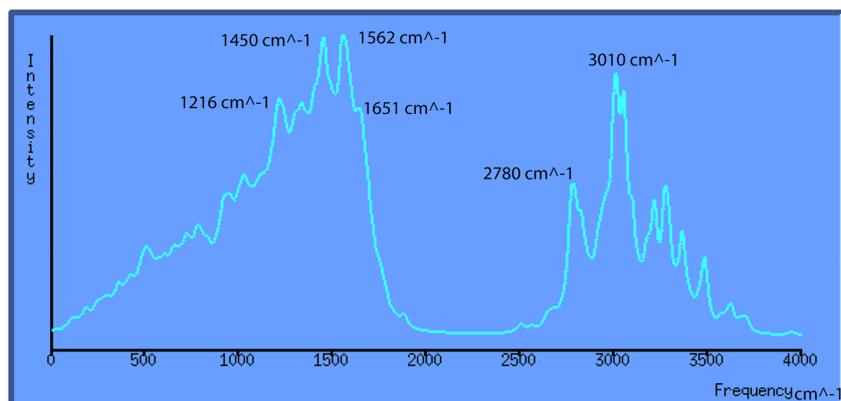


Figure 38: A figure showing the generated IR spectra of simplified SF, via DFT using geometry optimisation. Produced using Molden [94].

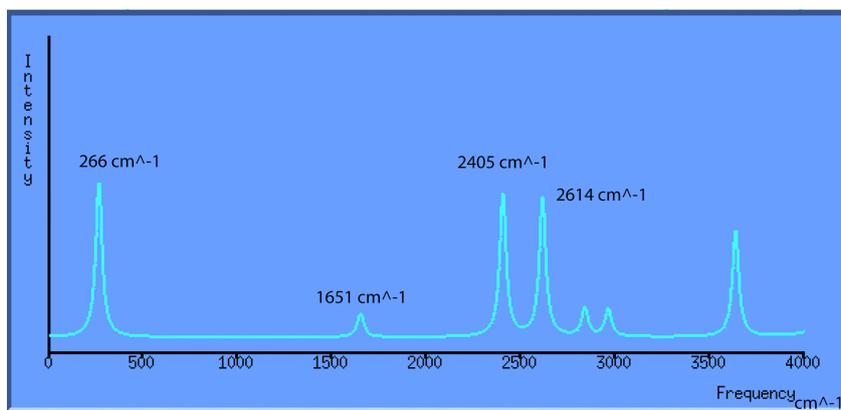


Figure 39: A figure showing the generated IR spectra of alanine. Produced using Molden [94].

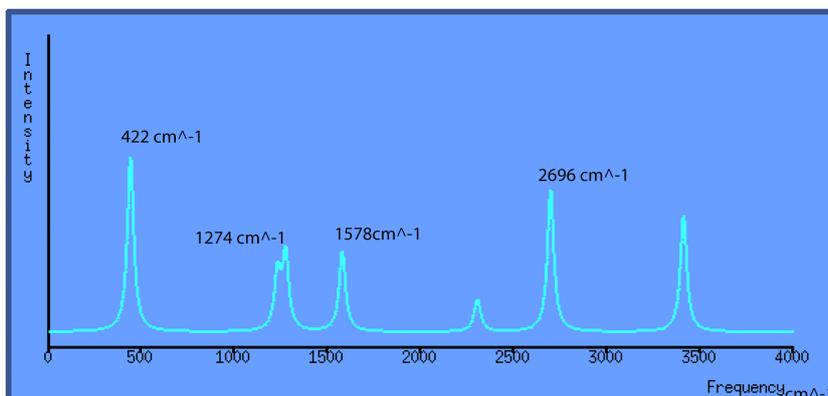


Figure 40: A figure showing the generated IR spectra of glycine. Produced using Molden [94].

From the above figures, it can be seen that the SF spectra contain peaks that can be assigned to alanine and glycine contributions. The 1562 cm^{-1} peak is present in the glycine spectra as 1578 cm^{-1} . The peak at 1651 cm^{-1} is in both the SF and alanine spectra. The SF peak at 1216 cm^{-1} may be attributed to the glycine peak at 1274 cm^{-1} . The prominent peaks of 2614 cm^{-1} and 2696 cm^{-1} present in alanine and glycine respectively can be assumed to relate to the 2780 cm^{-1} peak in the SF structure. All peaks below 1000 cm^{-1} are in the fingerprint region, which is specific to each molecule. The individual peaks for the SF spectra have been assigned in table 16. For the assignments [56,105,106] have been used.

Frequency (cm^{-1})	Assignment group	Assignment secondary structure
1220	CO stretch	NH bend in β sheet
1450	CH bending, alkane	-
1562	C=C stretch	NH deformation in α helix
1651	C=C stretch	CO stretch in β turn
2780	CH stretch	-
3010	OH stretch in carboxylic acid	-
3280	OH stretch	-

Table 16: Table showing the peak assignment for the SF spectra. Not all peaks will have a secondary structure associated with it.

The spectra can be considered in conjunction with Ramachandran plots where both β sheets and α helix should be present for both the IR spectra and Ramachandran plots. The computationally generated spectra can be compared to the experimental data from

figure 8. The two IR plots for alanine contain peaks in the region $1600\text{-}1700\text{ cm}^{-1}$. The computational value has a peak at 1651 cm^{-1} , whereas the experimental data has a peak at 1620 cm^{-1} . This change is due to the experiment being performed on multiple molecules, whereas the computational simulation was performed on a single molecule, which removes all intermolecular interactions. Computational glycine plots have a peak at 1578 cm^{-1} and a peak experimentally at 1590 cm^{-1} . The change in frequency is smaller for glycine than it is for alanine; this is due to glycine being a simpler molecule and having fewer degrees of freedom. This smaller change may also indicate that the glycine molecule has been minimised to a higher degree than the alanine molecule. However, both the computational IR spectra recreate the experimental to a sufficient degree of accuracy. The intensity is of different magnitudes, but this is due to the experimental plots only considering frequencies between 1500 and 1700 cm^{-1} , with the computational frequencies ranging from 0 to 4000 cm^{-1} . The SF plot in figure 8 and 38 both have a peak at 1650 cm^{-1} . This agreement generates a high level of confidence in this structure; this would be increased if the experimental plots were extended over a broader frequency range.

Further to the geometry optimisation IR, the IR have been produced at different temperatures $10\text{ K} - 673\text{ K}$. The plots of 10 K , 150 K , 298 K , 310 K and 673 K are shown below in figures 41 to 45. All other IR graphs have been included in the supporting information

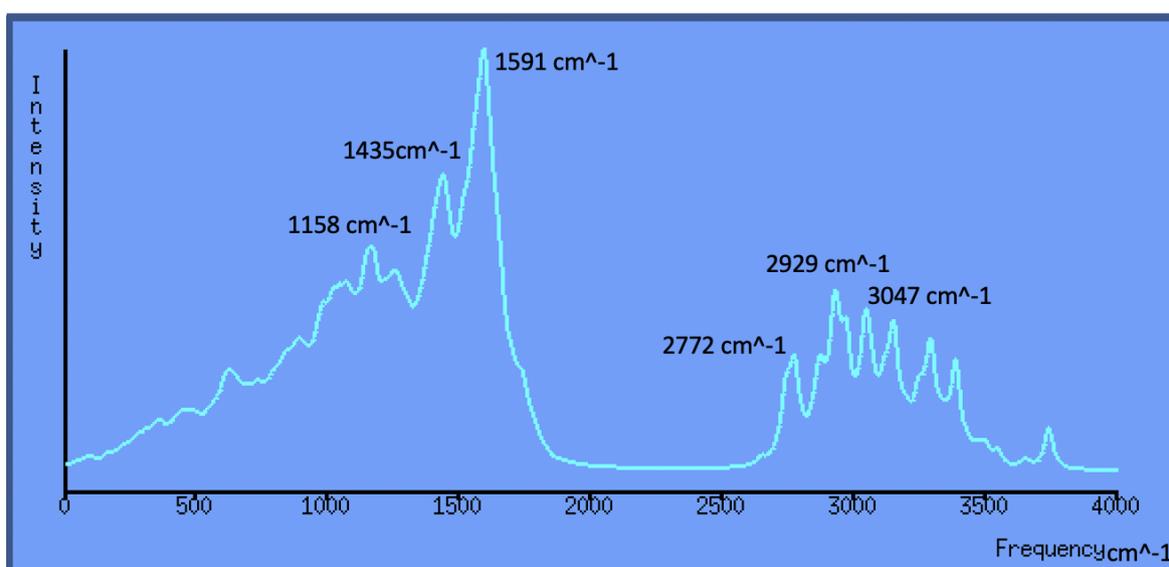


Figure 41: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 10 K . Produced using Molden [94].

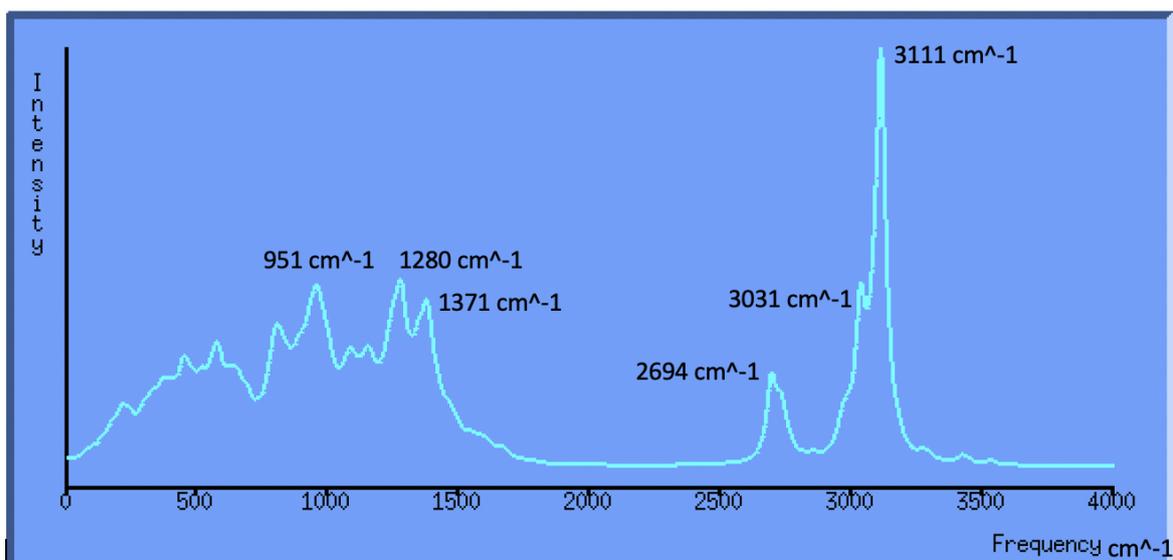


Figure 42: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 150 K. Produced using Molden [94].

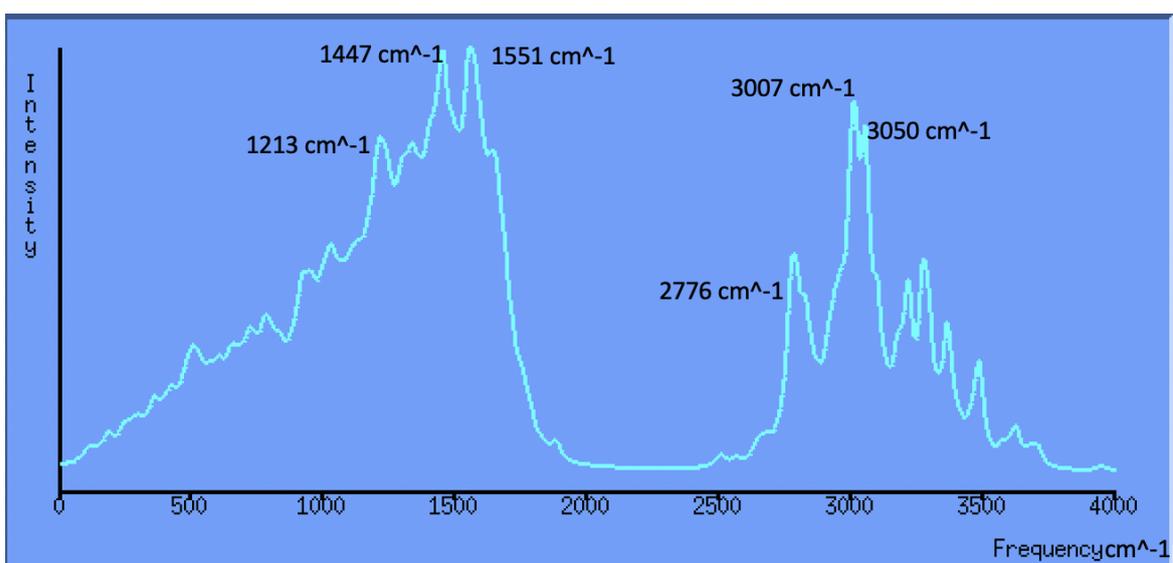


Figure 43: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 298 K. Produced using Molden [94].

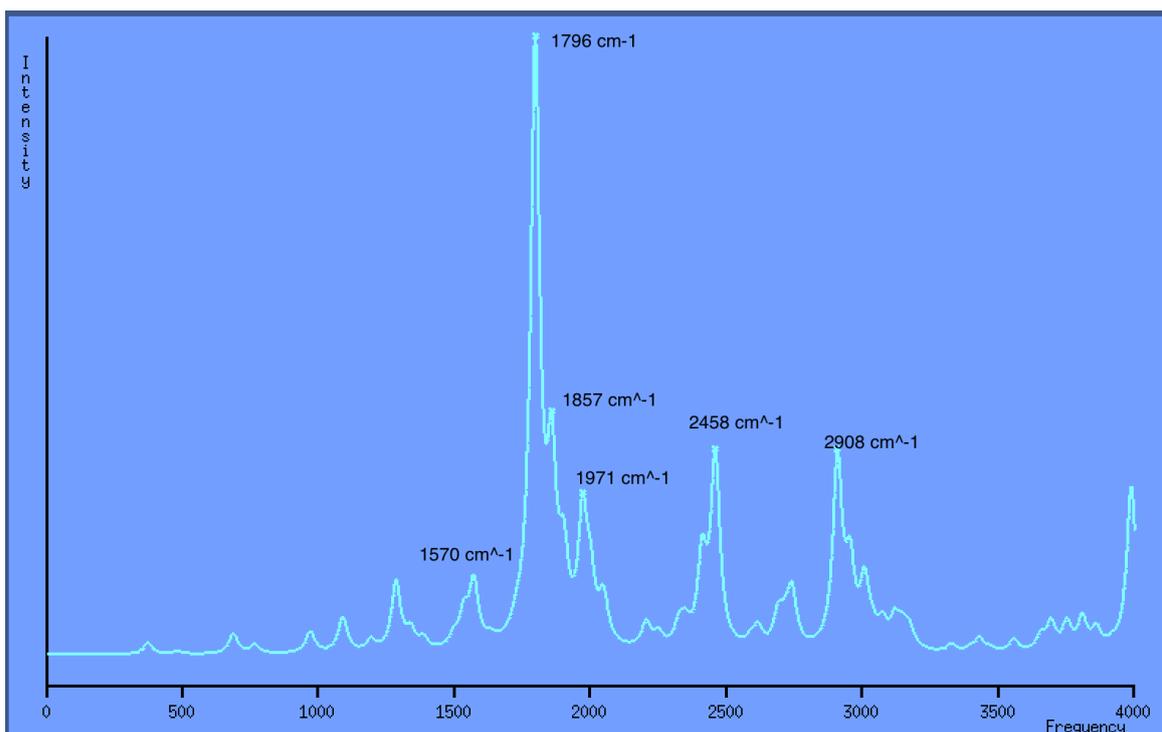


Figure 44: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 310 K. Produced using Molden [94].

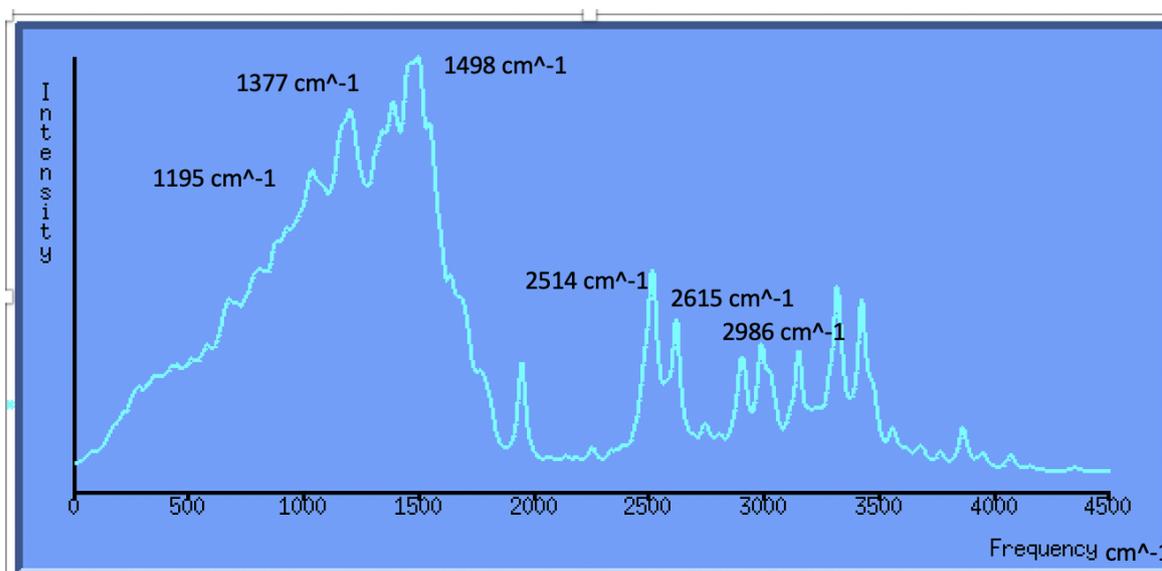


Figure 45: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 673 K. Produced using Molden [94].

In general, the peaks are consistent throughout the figures 41 to 45. The IR spectra at 150 K, figure 42 has been included as while the peaks occur in relatively similar positions to the other spectra the peaks are not as intense or strong. This is a result of the simulation having not fully optimised before the IR calculation being performed. The IR spectrum at 310 K contains a peak at 1976 cm^{-1} , this is a large change compared to the other graphs.

This will be due to the structure not being fully minimised before the IR calculation has been run. As the temperature of SF is increased the IR changes, in both position of the peaks and the structure. The peaks occur at a lower frequency when the temperature is increased; this is due to having an increase in energy and the bonds experiencing more vibration. The peaks become broader and less well defined at higher temperatures, this will be due to SF undergoing a phase change from solid to liquid, and thus the bonds become longer and weaker producing a broader signal. This is reinforced by the structure of the graphs, as the temperature approaches 673 K, the IR contains more 'noise', a higher number of peaks with less definition, which is not seen at lower temperatures such as 10 K. At 10 K the peaks are well defined and have a strong signal.

4.5 Behaviour of Water

There are two parts to this investigation - firstly looking at the positions of the water in the silk cell and secondly seeing the displacement of these molecules. This research has been completed using the geometry optimisation and molecular dynamics methods as described above. Unlike the Ramachandran plots the behaviour of water has only been simulated at 310 K (for molecular dynamics runs). The varying positions allow the average energy of the water molecule in individual sites to be calculated. Which then shows the most favourable position and where the water molecule will most likely sit in a hydrated cell. The molecules energy has been gained by performing a geometry optimisation for each position. The most and least favourable positions have been shown in table 17 and figure 47. The energies in table 17 have been calculated using equation 17. The positions of water molecules have been shown in figure 46, where the black molecules indicate the water molecule position.

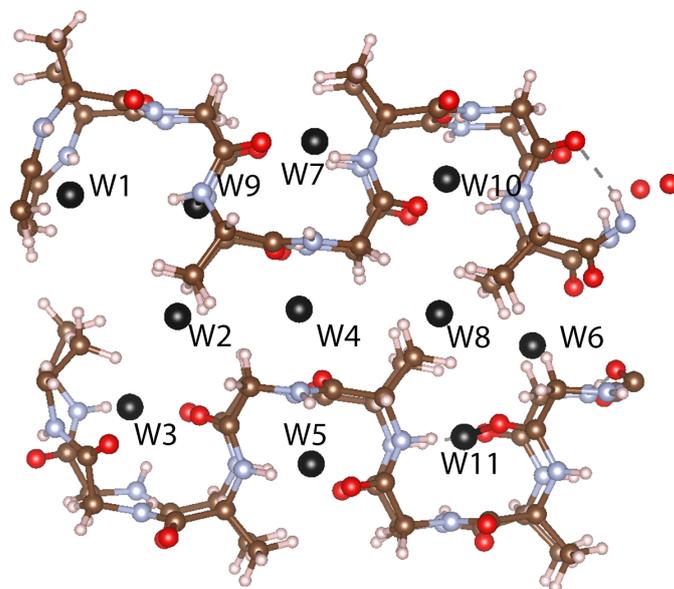


Figure 46: An image showing the various water positions.

$$\begin{aligned} \text{Energy of water position} = & \text{energy of combined system} - (\text{energy of optimised SF} \\ & + \text{energy of optimised water structure}) \end{aligned} \quad (17)$$

Position of water molecules (arbitrary names)	Energy of water position (KJ/mol)
W ₆	-83.918
W ₄	-82.763
W ₈	-35.870
W ₁₁	-3.892
W ₃	10.459
W ₉	11.167
W ₇	15.981
W ₁	19.597
W ₂	57.021
W ₁₀	58.556
W ₅	78.371

Table 17: Table showing the water energy in various positions.

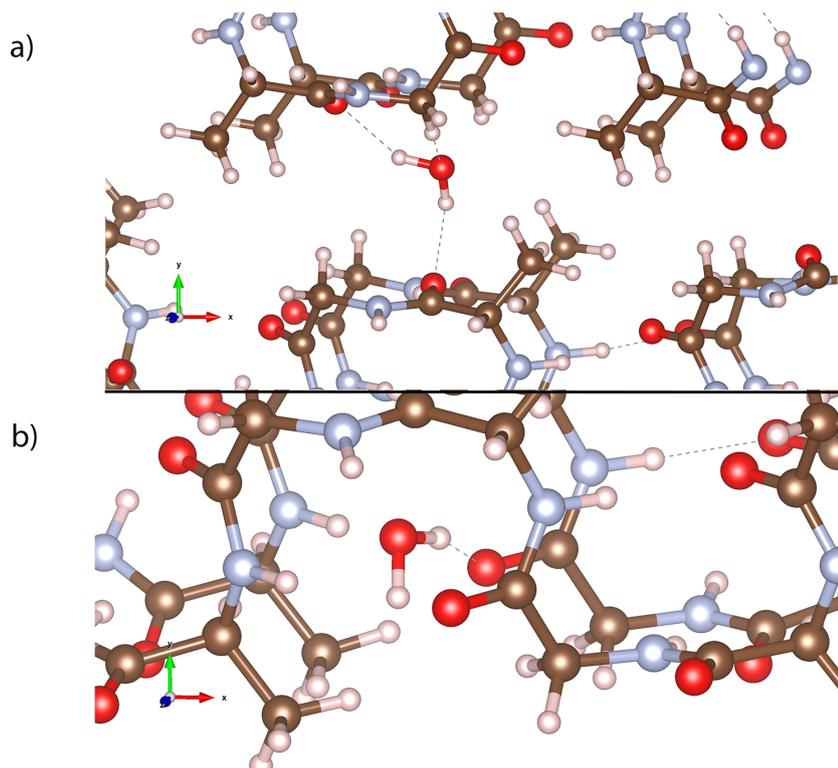


Figure 47: An image to show the most and least favourable water positions. Created by the Author using Vesta [79]. a) Most favourable position for water - W_6 . b) Least favourable position for water - W_5 .

The energy of the position correlates to the number of hydrogen bonds that the molecule can form at any one time. In general, if the water molecule sits between the chains in the y-direction, the energy will be lower as the water can form more hydrogen bonds. The most favourable position can form 3 hydrogen bonds whereas the least favourable can only form 1 hydrogen bond. W_5 is an unfavourable site with a binding energy of +60 KJ/mol this shows it requires energy to be placed in this site and would unlikely appear in this site in experiment.

Furthermore, to the energy of water positions, the energy of the cell as the number of water molecules increases has been investigated. This cell allows for more experimentally accurate results as it has been shown that the experimental SF contains 7.5% water. As the number of molecules increases from zero to ten, it approaches the experimental value. The experimental SF cell has been made with 8% water by weight, and this is being recreated in the hydrated cell [5]. The energy of the increasing molecules is shown in table 18, this energy is the averaged energy per water molecule.

Number of water molecules	Average energy of water molecules (KJ/mol)
1	-83.918
2	-62.884
4	-17.40
6	-15.34
8	-18.26
10	-12.25

Table 18: Table showing the water energy in varying amounts.

In table 18, the energy varies greatly; however, this is a result of the position of the water molecules, as shown in table 17. Due to the cell size, not all of the water molecules can fit in the most favourable positions, and therefore some molecules will always appear in less favourable positions.

The cells with 1 and 8 water molecules had been chosen to run further calculations to investigate the water displacement when the cell is run with molecular dynamics. Two cells were chosen to provide a comparison of the water molecules. Do they behave in the same way when there is a single water molecule versus eight water molecules in one system? For the one molecule system, ran for 1592 femtoseconds with the eight molecule system have been run for 2117 femtoseconds. This simulation produces a trajectory for each of the water molecules in the cell. The trajectory for the singular water is shown in table 19. The trajectories for the eight molecule system are shown in tables 20 and 21.

Water molecules	Displacement (\AA)	Equivalent position at 0 fs	Equivalent position at 1592 fs
1	0.818	W_1	W_1

Table 19: Table showing the water displacement in the one water molecule system.

Water molecules	Displacement (\AA)	Equivalent position at 0 fs	Equivalent position at 2117 fs
1	0.559	W_4	W_2
2	1.394	W_9	W_1
3	1.314	W_{11}	W_8
4	0.062	W_4	W_8
5	0.540	W_5	W_5
6	0.718	W_{10}	W_{10}
7	0.907	W_3	W_3
8	2.295	W_7	W_7

Table 20: Table showing the water displacement in the eight water molecule system.

Water molecules	Displacement (Å)	Displacement in x (Å)	Displacement in y (Å)	Displacement in z (Å)
1	0.559	-0.557	0.027	0.027
2	1.394	0.364	-1.275	0.430
3	1.314	-0.969	0.869	-0.181
4	0.062	-0.027	-0.044	0.034
5	0.540	-0.004	0.023	-0.540
6	0.718	-0.717	0.005	-0.029
7	0.907	-0.148	-0.194	0.874
8	2.295	2.017	0.015	-1.094

Table 21: Table showing the water displacement as a sum of the individual vectors in the eight water molecule system.

While in general water molecules that are in the most favourable positions experience a lower displacement, this is not true for all cases. This may be a result that the tables 19 and 20 only show the displacement at the final time frame instead of the maximum displacement of each water molecule. The displacement for all atoms mainly occurs in the x-direction; it appears the water will travel along the chains but will not migrate through the chains. In general the water molecules diffuse from a less favourable position to a more favourable position. This diffusion may occur in more giant cells. This diffusion in the x, y and z-axis are shown in figures 48, 49 and 50.

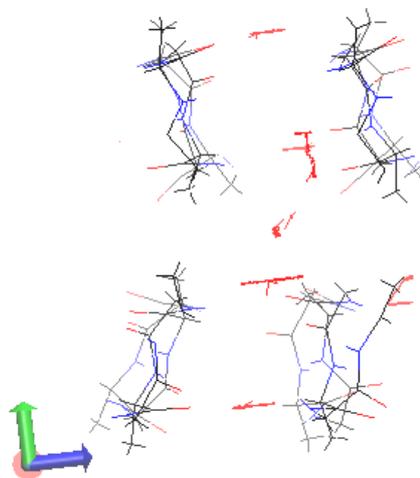


Figure 48: A figure showing the water diffusion in the x-direction. Hydrogen atoms have been omitted.

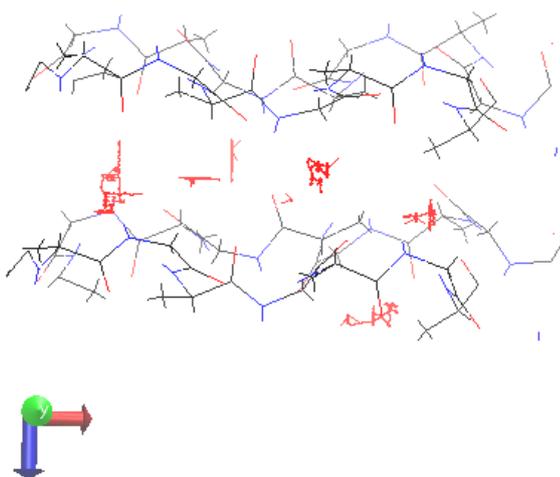


Figure 49: A figure showing the water diffusion in the y-direction. Hydrogen atoms have been omitted.

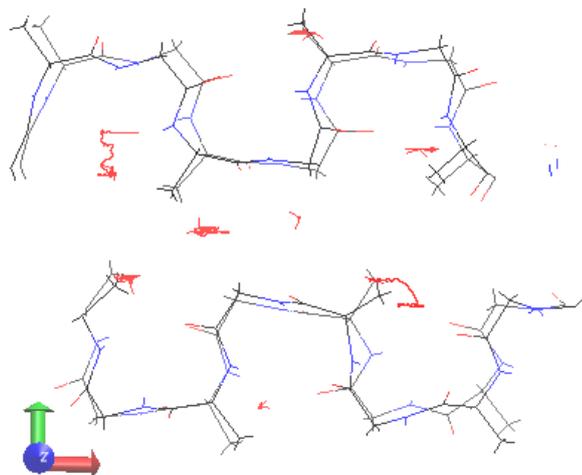


Figure 50: A figure showing the water diffusion in the z-direction. Hydrogen atoms have been omitted.

Classical calculations have been performed to investigate the MSD, diffusion coefficient and activation energy of water in SF. This has been performed on both $(-\text{Ala-Gly-})_{1028}$ and $(-\text{Ala-Gly-})_{128}$. The general trend observed is that the water molecules are displaced more at higher temperatures for both $(-\text{Ala-Gly-})_{128}$ and $(-\text{Ala-Gly-})_{1024}$ SF crystal models. Dissimilarly, the water molecules incorporated in the more extensive $(-\text{Ala-Gly-})_{1024}$ SF crystal follow a smoother displacement gradient over the 50 ps timeframe. Whereas, the smaller $(-\text{Ala-Gly-})_{128}$ SF crystal model, suggest a more erratic displacement over the 50 ps timeframe. This displacement could infer greater stability of the system and more hydrogen bond interactions occurring between the polymer chains and water molecules in the larger model. This then allows the diffusion coefficient to be calculated. Which can then be used to produce an Arrhenius plot, and thus the activation energy can be calculated. The activation energy for the $(-\text{Ala-Gly-})_{1024}$ cell is 9.232 kJ/mol, and for $(-\text{Ala-Gly-})_{128}$ is 9.298 kJ/mol. This is shown in more detail in the attached paper "Evaluating a Bombyx mori silk fibroin model using a range of atomistic simulation tools".

4.6 Behaviour of Magnesium

Magnesium has been simulated in SF cell due to the requirement of having an ion in the battery. Mg^{-2} was chosen due to little choice provided by CP2K for DZVP and BLYP. Similar to above the behaviour of magnesium has been investigated in the $(-\text{Ala-Gly})_{16}$ SF cell. However, there is no classical mechanics comparison for the data and very little experimental data either which makes verifying the data more difficult. Both the energy of Mg in various positions and as the number of magnesium ions increases has been investigated. This has been completed for hydrated and non-hydrated systems to see how magnesium ions interact with the protein chain and water.

This energy is calculated via the geometry optimisation method as described in the methodology, the energies for the various positions are detailed in tables 22 and 23 with the corresponding images for the most and least favourable positions included in figures 52 and 53. The energies have been calculated using equation 18, where the energy of optimised water structure is equal to zero for the non-hydrated state. The Mg positions have been included in figure 51.

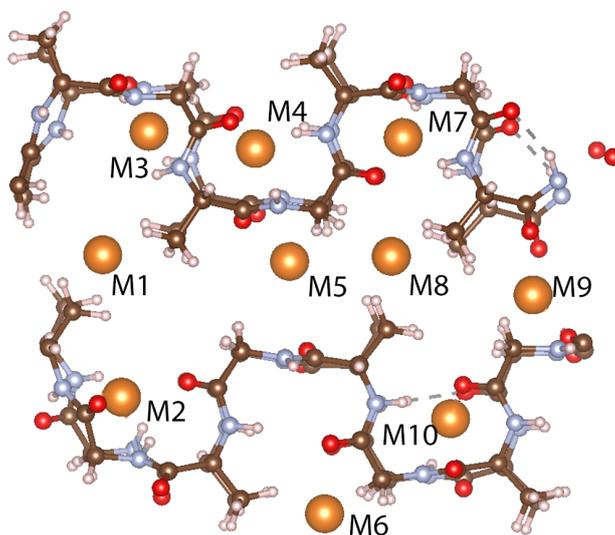


Figure 51: An image showing the various magnesium positions.

$$\begin{aligned} \text{Energy of Mg position} = & \text{energy of combined system} - (\text{energy of optimised SF} \\ & + \text{energy of optimised water structure} + \text{energy of Mg ion}) \end{aligned} \quad (18)$$

Position of Mg ions (arbitrary names)	Energy of Mg position (KJ/mol)
M ₁₀	-380.555
M ₉	257.250
M ₃	265.487
M ₄	513.663
M ₆	567.003
M ₅	609.667
M ₇	759.515
M ₂	844.159
M ₈	962.947
M ₁	1570.839

Table 22: Table showing the magnesium energy in various positions for the non-hydrated cell.

Position of Mg ions (arbitrary names)	Energy of Mg position (KJ/mol)
M ₇	-395.754
M ₄	-290.500
M ₃	119.395
M ₈	147.102
M ₁	215.340
M ₆	220.981
M ₉	550.130
M ₂	712.243
M ₅	2146.668
M ₁₀	2645.452

Table 23: Table showing the magnesium energy in various positions for the hydrated cell.

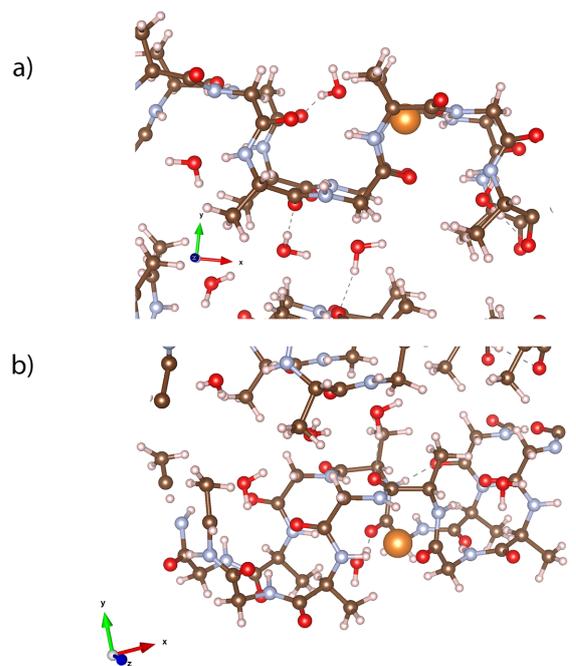


Figure 52: An image to show the most and least favourable magnesium positions for the hydrated system. Created by the Author using Vesta [79]. a) Most favourable position for magnesium - M_7 . b) Least favourable position for magnesium - M_{10} .

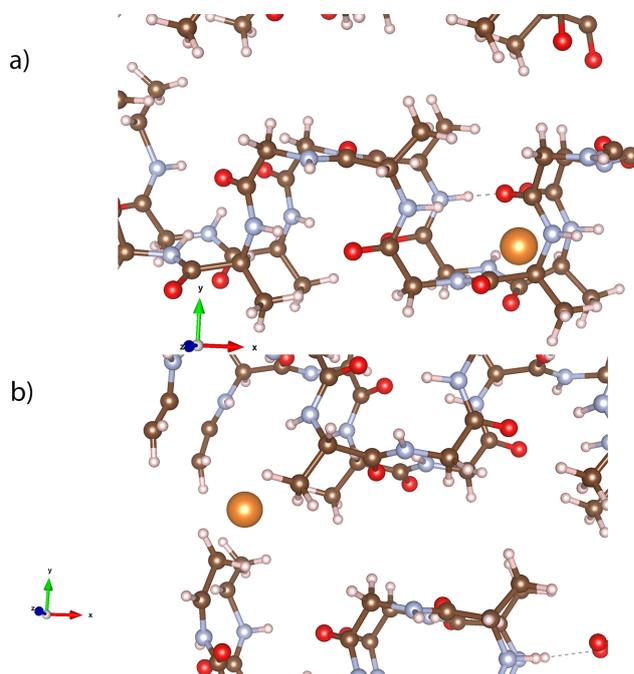


Figure 53: An image to show the most and least favourable magnesium positions for the non-hydrated system. Created by the Author using Vesta [79]. a) Most favourable position for magnesium - M_{10} . b) Least favourable position for magnesium - M_1 .

From figures 52 and 53, it can be seen that the magnesium ion will have lower energy

when it is near an oxygen or nitrogen atom in the chain, with the higher energy magnesium ions generally being in a wider space or closer to hydrogen atoms. This result is opposite to the water energies trends shown above. The reason for this is that the magnesium ions will not form hydrogen bonds with hydrogen atoms but will form an ionic bond when in the presence of oxygen and nitrogen, the magnesium will be attracted to regions of high electron density. As a result, it will have lower energy when the charge of the ion can be shared across multiple atoms such as oxygen and nitrogen. The hydration has a substantial effect on the energy of the magnesium ion. For M_7 the introduction of eight water molecules has moved the magnesium ion by 1.6332 \AA , this is into a more favourable position and thus lower energy.

This investigation of the positions and its effect on the magnesium ions' energy then allows for the calculation of the average energy of magnesium molecules in both hydrated and non-hydrated cells. This has been completed in the same method as previously discussed.

Number of magnesium ions	Average energy of magnesium ions (KJ/mol)
1	215.34
2	484.14
4	12.60
6	254.28
8	-118.47
10	159.48

Table 24: Table showing the magnesium energy in varying amounts for the hydrated system.

Number of magnesium ions	Average energy of magnesium ions (KJ/mol)
1	1570.839
2	-15.62
4	-59.01
6	140.44
8	98.54
10	-286.50

Table 25: Table showing the magnesium energy in varying amounts for the non-hydrated system.

The energy of the magnesium relies heavily on the position of magnesium and water. The most significant energy difference is seen between the singular magnesium ion hydrated and non-hydrated systems, the change is due to the number of ionic bonds and instant dipoles that can be formed between the magnesium ions and the water molecules. As the number of magnesium ions increases, more bonds can be formed which lowers the energy of the system,

the number of water molecules is kept constant and in the same position. This allows for the charge of magnesium to be shared across atoms, which, again, lowers the energy of the system. However, there appears to be a very weak correlation between energies for either system as the number of magnesium ions is increased, shown in figure 54.

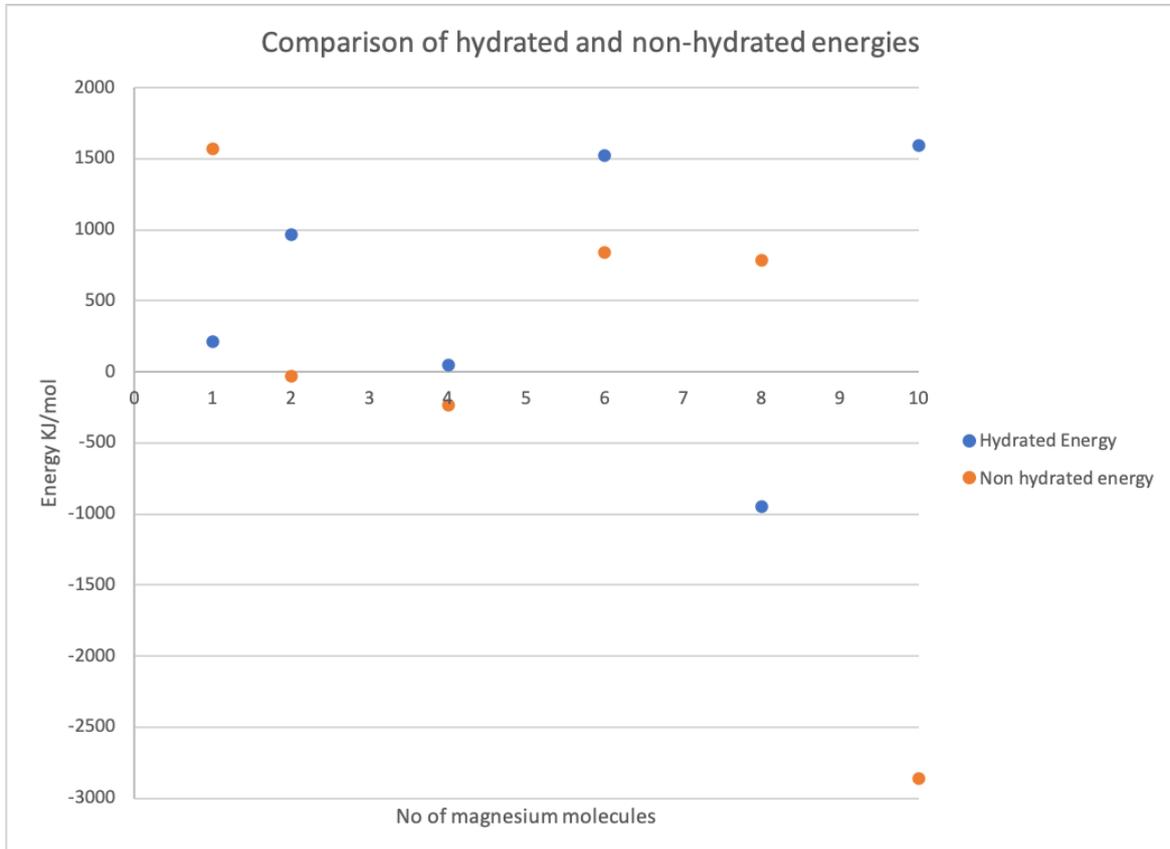


Figure 54: A graph showing the comparison of the energy of the hydrated and non-hydrated system.

There is no classical mechanical results or experimental results to show agreement or disagreement with these results. This lack of data makes it difficult to state the exact relation between energy and number of magnesium molecules. Nevertheless, there is a strong relationship between the position of the magnesium ion and the energy of the ion.

A hydrated system with added magnesium has been run using molecular dynamics for 558.000 fs. Eight magnesium ions were chosen; this allows for a range of positions to be tested to see how the magnesium behaves in a more extended period. Similar to above the water behaves in the same way, the water diffuses along the x-axis with little movement in other directions, shown in figures 55, 56 and 57. The water diffusion has been shown in table 26. As above this is a snapshot of the diffusion at 558 fs and does not show the furthest displacement experienced by the water or magnesium ions. The equivalent positions have not been included in table 26 as the positions and energies are of relevance for the Mg ions.

Water molecules	Displacement (Å)	Displacement in x (Å)	Displacement in y (Å)	Displacement in z (Å)
1	0.273	0.262	0.078	0.001
2	0.528	0.501	0.028	-0.162
3	0.368	-0.013	0.045	-0.365
4	1.040	1.034	0.062	-0.092
5	0.237	0.001	0.218	-0.092
6	1.189	-1.147	0.132	-0.285
7	1.274	0.652	1.054	0.294
8	1.029	0.085	-0.810	0.629

Table 26: Table showing the water displacement in the charged system.

For figures 55, 56 and 57, the red trajectory lines show the diffusion pathways of the water molecules, and the orange shows the magnesium pathways the displacement is minimal.

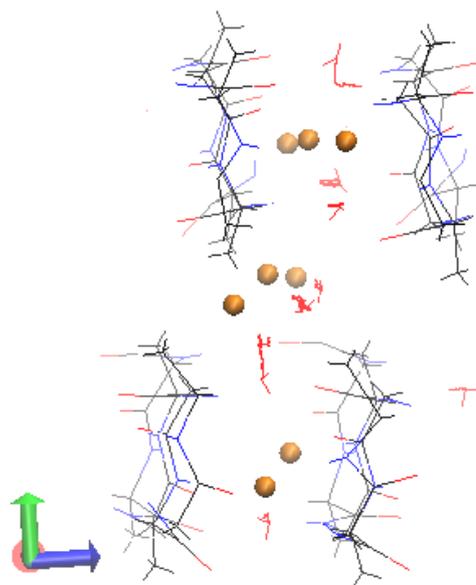


Figure 55: A figure showing the water and Mg diffusion in the x-direction. Hydrogen atoms have been omitted.

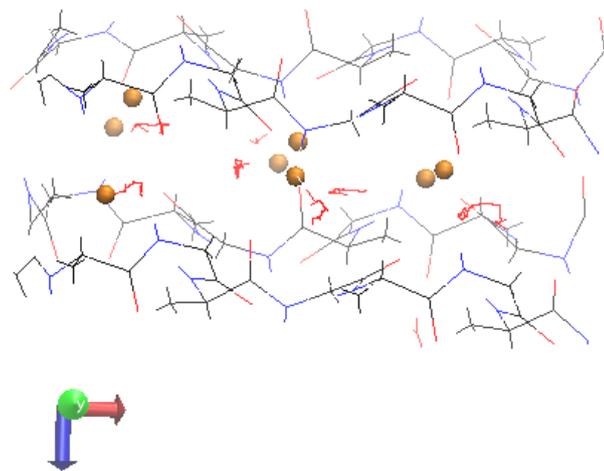


Figure 56: A figure showing the water and Mg diffusion in the y-direction. Hydrogen atoms have been omitted.

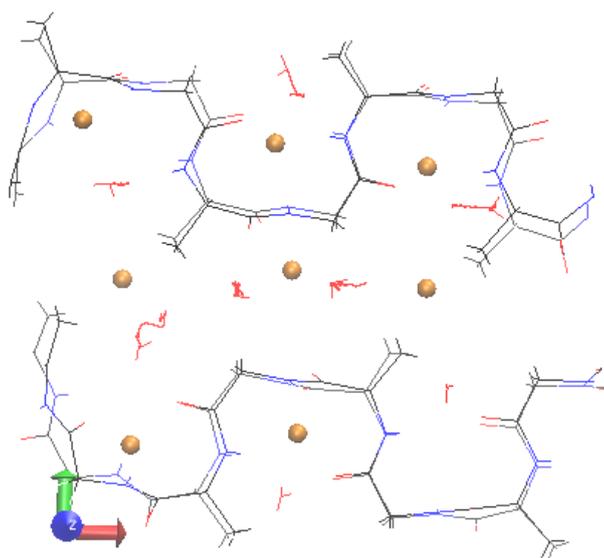


Figure 57: A figure showing the water and Mg diffusion in the z-direction. Hydrogen atoms have been omitted.

As shown in figures 55, 56 and 57, the magnesium ions have little to no displacement. There are several possible reasons for this; the most likely is that the simulation has not been run for a sufficient duration of time, spanning 558 fs whereas the water displacement calculations were run for 2117 fs. This time is due to a restriction on the amount of computational time provided for the more extensive simulations. One reason for the diffusion not being seen is that the magnesium is a large atom and is unlikely to experience large enough forces to move the Mg ions a significant distance in the DFT calculations. The magnesium diffusion has been shown in table 27. Here the equivalent positions have not changed due to the small displacement but are included for completeness.

Magnesium ion	Displacement (\AA)	Equivalent position at 0 fs	Equivalent position at 558 fs
1	0.0454	M ₃	M ₃
2	0.0325	M ₁	M ₁
3	0.0258	M ₂	M ₂
4	0.0596	M ₅	M ₅
5	0.0494	M ₄	M ₄
6	0.0714	M ₉	M ₉
7	0.0178	M ₆	M ₆
8	0.0401	M ₇	M ₇

Table 27: Table showing the magnesium displacement in the charged system.

4.6.1 Electron Density

An electron density calculation has been performed on the charged cell. This calculation allows for better visualisation of where the electrons are likely to appear. This is shown in figure 58 and 59. The area around the magnesium ion is empty of electrons as expected due to it carrying a charge of +2 and thus having an empty outer electron shell. The shape of the electrons around the water molecules can be seen. This shape is amplified with the presence of hydrogen bonds.

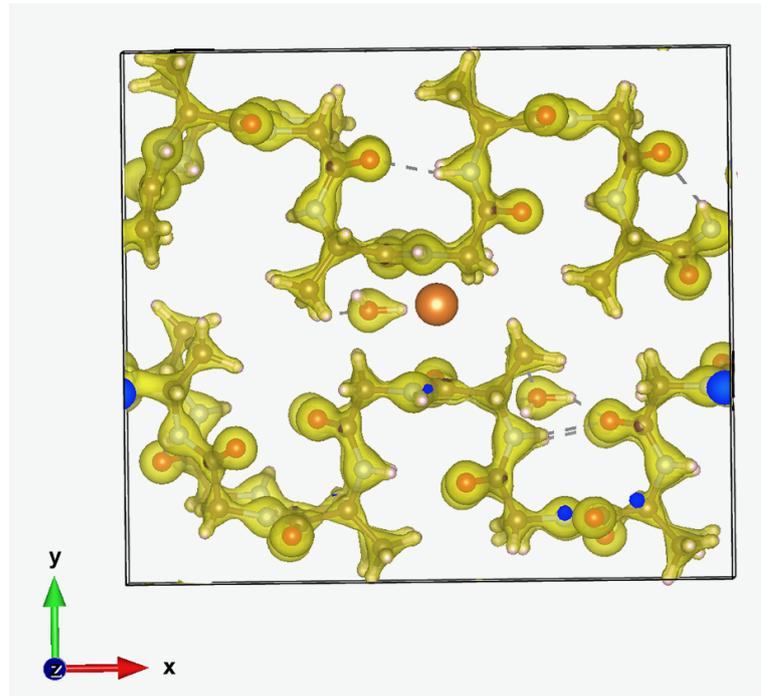


Figure 58: A figure showing the electron density of the charged cell.

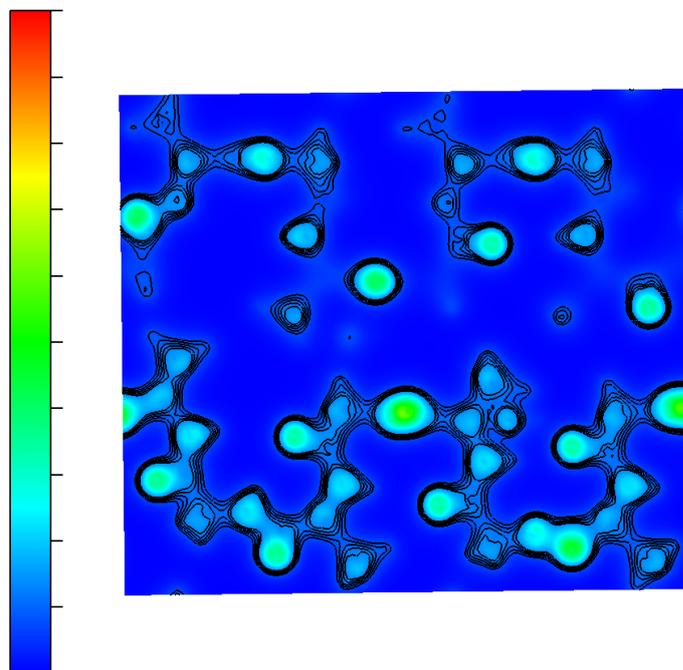


Figure 59: A figure showing the electron density map of the charged cell. Where the key is shown on the left-hand side, the colours are arbitrary, with red showing a greater area of electron density and the blue showing an area of low electron density.

It is worth noting that for figure 59, the electron density is taken at a set distance in the cell and due to the shape of the SF (i.e. not being flat to a plane) the splice may not contain a complete electron picture.

4.7 Behaviour of Choline

The behaviour of choline has been investigated for both individual choline molecules and choline nitrate in a silk cell. Several simulations have been performed, similar to previous methods. The structure of choline and choline nitrate have been shown in figures 60 and 61.

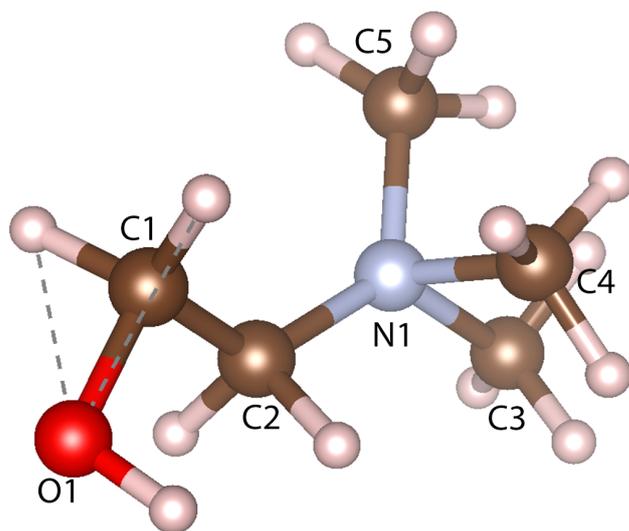


Figure 60: A figure showing the choline structure

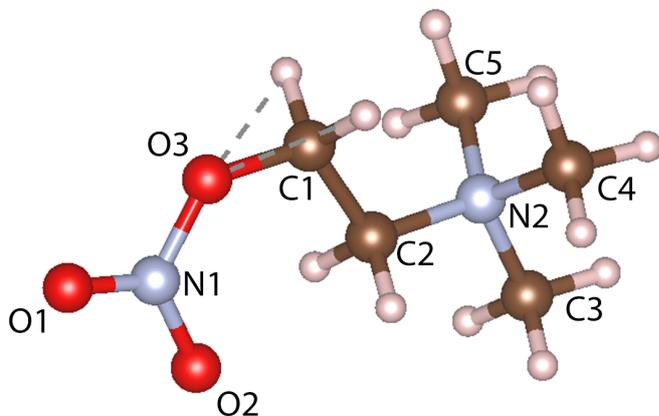


Figure 61: A figure showing the simulated choline nitrate structure; this is not the experimental structure of choline nitrate.

The bond lengths for both choline and choline nitrate have been compared against literature to check if the structures are valid and can be used in the full system. The comparison is in tables 28 and 29. In both tables the hydrogen - x bonds have been omitted. Table 28 has been taken from table 10 in the methodology but is shown to provide an easy comparison of nitrogens effects on the bond lengths. The atom labels are shown in figure 61.

Atoms		Bond length (Å)
Atom 1	Atom 2	
O1	C1	1.3747
C1	C2	1.4140
C2	N1	1.5419
N1	C3	1.4115
N1	C4	1.5617
N1	C5	1.4713

Table 28: Table showing the bond lengths of choline. Hydrogen bonds have been omitted.

There have been very few investigations of choline in silk previously. For choline in SF,

Atoms		Bond length (Å)
Atom 1	Atom 2	
O1	N1	1.193
O2	N1	1.193
N1	O3	1.328
O3	C1	1.396
C1	C2	1.552
C2	N2	1.498
N2	C3	1.484
N2	C4	1.483
N2	C5	1.483

Table 29: Table showing the bond lengths of choline nitrate. Hydrogen bonds have been omitted.

the structure has been assumed to form a crystal-like structure, using an ordered repeating structure. This was tested in both the x and y direction, to see which structure is likely to appear experimentally. As shown previously, choline is in the silk system at a ratio of 1:3 with regards to molecular weight. This research has led to 27 choline molecules in SF. The two structures in the simulation have layers of choline nitrate in either the xy plane or the zx plane. The two structures are shown in figures 62 and 63.

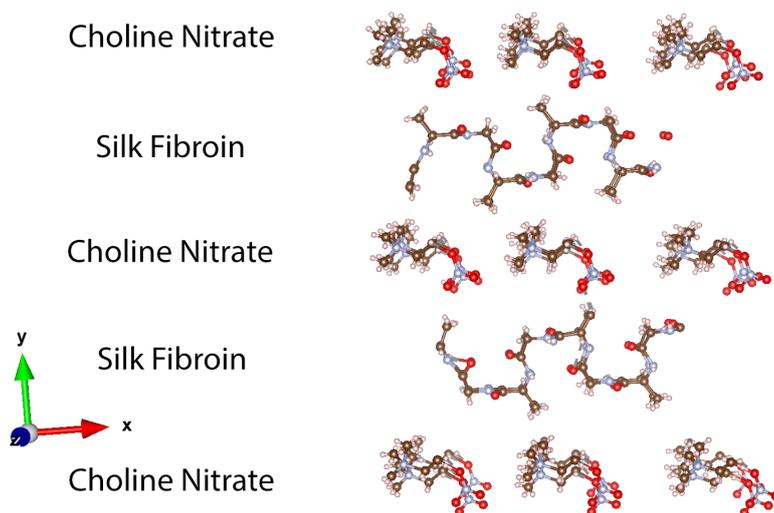


Figure 62: A figure showing the choline nitrate in the xy plane.

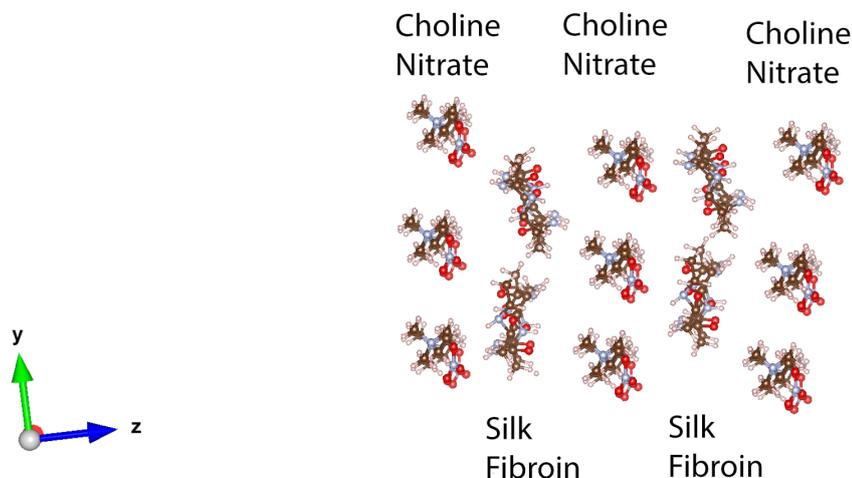


Figure 63: A figure showing the choline nitrate in the zx plane.

Both structures were run under MD, due to the computational cost the simulations were run for a short period: 6.8 fs for the choline nitrate in the xy plane and 7.45 fs for the choline nitrate in the zx plane. No movement of the choline nitrate occurred in either simulation. However, due to the short time frame, no diffusion has been expected. It would not be realistic to compare the energies of the two structures due to the change in structure.

Literature suggests that choline will diffuse through the cell similar to water and magnesium, this has not been shown in silk, but in secondary material [66]. This diffusion has not been seen in this project; this may have some contributing factors. One of the main issues will be that the simulation has not been run for a long enough period to see any such movement, due to the restrictions on computational time given to this project. However, in this simulation, choline will sit between chains without causing any damage to the structure. Another issue of these simulations is that the choline nitrate has been assumed in an unlikely form. The simulations have used a single molecule of choline nitrate, but this is more likely to take on the form of two separate species (cationic choline and an anionic/zwitterionic molecule (anionic/zwitterionic depends on the pH)). This simulation has not simulated pH values, and thus the exact form of the anionic/zwitterionic can not be extracted from this project. However, this allows for the simulations to show how the cell will act when it has

been saturated with choline, using the experimentally accurate weights. Another reason for not seeing the diffusion would be the size of the cell, as a small cell will restrict movement and a larger cell will reduce the forces acting on the molecule.

Even though the choline nitrate structure is incorrect, the overall cell is stable in this simulation. This stability will allow for both the silk structure and the choline structure to be used in further calculations. The chosen structure is the choline nitrate seeded in the zx plane. There is little difference between the two; however, this structure has run for more steps in the same time and so will be cheaper when more molecules are added to the structure. [107] has described a range of different choline ionic liquids, a mixture of cations and anions and has demonstrated that the two groups are mobile in PPy-gelatin. This reinforces that the structure used in this project is incorrect, using the correct anion-cation structure may show a diffusion as the molecules will be much smaller and liable to more significant molecule interaction, this is due to the molecule experiencing larger forces encouraging the diffusion.

4.8 Behaviour of the complete cell

A complete cell has been simulated with experimentally accurate molecular weights - 7% of water, 1:3 choline and sufficient Mg ions in the cell. While shown above the choline is not in the correct structure; it provides a method for gaining more detailed information about the entire cell. The calculation has been run for 31.45 fs; this is a fraction of the time the water diffusion simulations were run for; however, this simulation has a much higher computational cost. From data generated, both the water and magnesium have experienced minimal diffusion, again due to the restricted time frame, but this would increase the longer the simulation. Where the water has travelled less than 0.1 \AA , and the magnesium has moved less than 0.2 \AA . The displacements are shown in tables 30 and 31. The previously found equivalent positions do not apply in this situation due to the change in structure and addition of choline.

Water molecules	Displacement (Å)	Displacement in x (Å)	Displacement in y (Å)	Displacement in z (Å)
1	0.0802	-0.037	-0.068	0.020
2	0.135	0.108	-0.055	-0.060
3	0.0671	0.040	-0.050	-0.021
4	0.0828	-0.056	-0.022	0.057
5	0.0921	0.045	0.016	-0.079
6	0.0527	0.041	-0.011	-0.032
7	0.0589	0.044	-0.011	-0.037
8	0.0722	-0.005	-0.037	0.062

Table 30: Table showing the water displacement in the full system.

Magnesium molecules	Displacement (Å)	Displacement in x (Å)	Displacement in y (Å)	Displacement in z (Å)
1	0.103	0.0875	-0.046	-0.031
2	0.110	-0.098	0.001	0.050
3	0.179	-0.045	-0.007	-0.173
4	0.0210	-0.020	-0.004	-0.005
5	0.0543	0.035	0.028	0.031
6	0.0725	0.011	0.050	-0.051
7	0.102	-0.083	0.060	0.0001
8	0.0592	0.054	0.001	-0.023

Table 31: Table showing the magnesium displacement in the full system.

The diffusion of magnesium and water is consistent with what has been seen previously and agrees with the simulations via classical mechanics. The diffusions generally occur in the x-direction; however, a more significant displacement has been seen on the magnesium in the full cell compared to the magnesium in the hydrated cell. This is due to the increased charge placed on the system with the introduction of choline nitrate. This generates more force on the charged ions and generates a larger displacement. This charge appears to have had little effect on the choline nitrate due to their increased size. Choline nitrate has not diffused in either simulation containing the molecule.

On further investigation into the simulation, it appears that both the choline and the silk structure have broken down in the 31.45 fs simulation. This is shown in figures 64 and 65. The breakdown of SF is a result of the exclusion radius for seeding magnesium being too small; it has been kept the same as for the water molecules- 1.4 angstroms. This simulation has shown that this needs to be increased as at this distance, the magnesium binds into the structure and breaks the silk chain. The silk chain has broken in multiple places, as shown

in figure 65.

The choline nitrate has also broken down; this further supports the idea above that the choline structure used is not the experimental structure. The choline nitrate has been simulated as a single molecule instead of two molecules which form the ionic liquid. The choice to model the ionic liquid as a crystalline structure may have also resulted in the structural deformation.

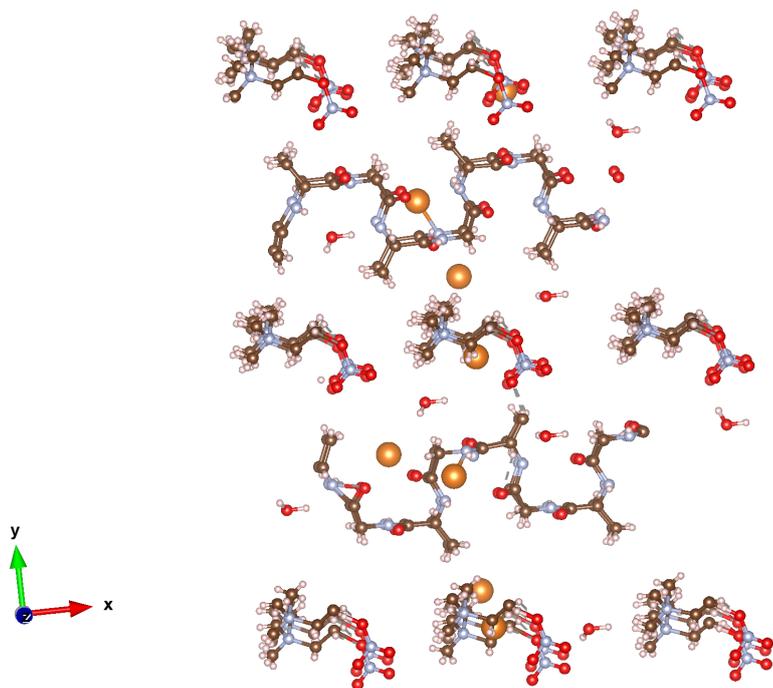


Figure 64: A figure showing the first frame of the full simulation.

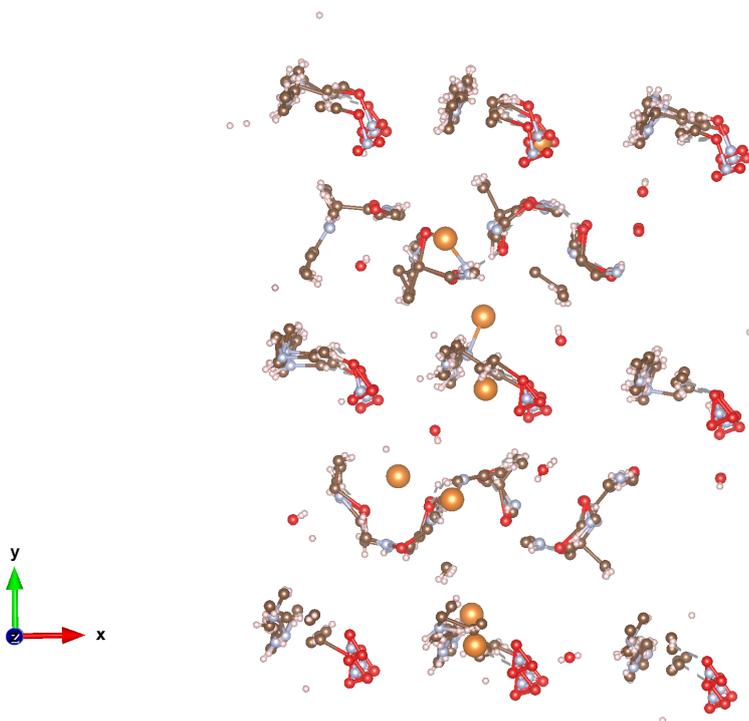


Figure 65: A figure showing the last frame of the full simulation.

To improve this simulation, a larger exclusion radius for the magnesium ion will be needed, and a further investigation into the choline nitrate structure will be required.

5 Conclusions

This project generates several exciting conclusions. The conclusions can be drawn from the computational simulation of SF and about the SF material itself. A range of calculations have been performed throughout this project; these aim to answer questions set out in the introduction. Further to this, other questions which have been investigated include: How does SF act when excited? How does the SF material behave under varying temperatures?

5.1 Is the generated structure valid?

Through the simulations, the generated simplified structure is an accurate estimation of the experimental structure of SF. The simplified structure accurately recreates experimental Ramachandran plots, IR spectra and thermal expansion behaviour, giving a high level of confidence in this structure. This is an important conclusion as it shows that this project has been completed as expected and has validated the model for use in more complicated simulations paving the way to simulate a battery in the body.

5.2 Is the computational time enough for all calculations to be successfully run?

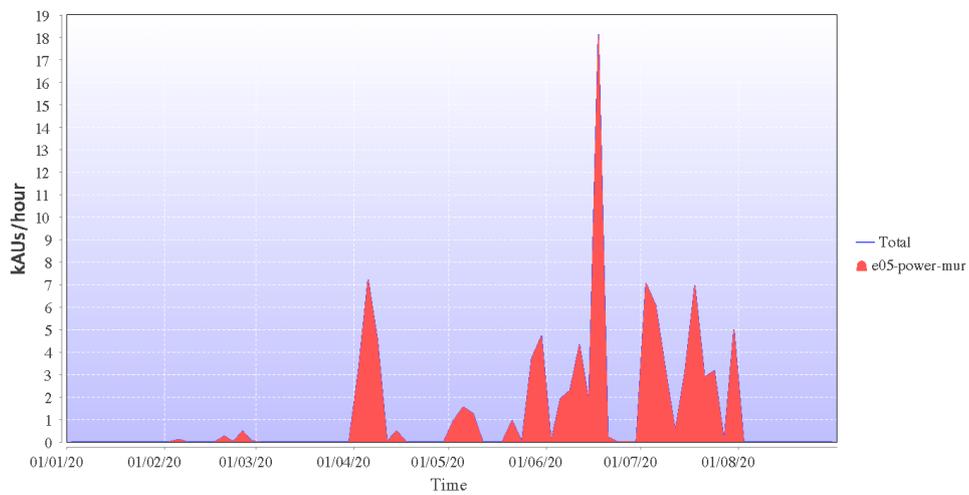
The computational time requested for this project from ARCHER was not sufficient for the calculations performed. 2500 kAUs were requested, and 7000 kAUs were used in total. This extra time was provided from kind contributions from students in the research group, allowing more complicated calculations as needed. The total ARCHER time used is below in figure 66. The HEC was sufficient for smaller calculations with no time limit so allowed for calculations to run for extended periods; this was useful for IR calculations where the run time exceeded four weeks.

ARCHER User Usage Report: Mr Benjamin Deacon , (bdeacon)

Report produced on 2020-08-18 .

A total of 211 jobs were submitted on ARCHER by Mr Benjamin Deacon account bdeacon during the period Jan - Aug 2020 .

Usage in kAUs :



Resource	Budget	Raw Usage / kAUs	Charged Usage / kAUs	Jobs
XC	e05-power-mur	7,008.701	7,008.701	211

Usage broken down by job size.

Cores	Raw Usage / kAUs	Charged Usage / kAUs	Jobs
17-32	0	0	1
65-128	4,016.423	4,016.423	148
129-256	986.45	986.45	48
257-512	1,313.78	1,313.78	9
513-1024	692.048	692.048	4

Figure 66: An overview of the number of jobs run on ARCHER.

5.3 How do various species behave in SF?

Water readily diffuses through the SF structure generally in the x-direction, along the silk chains, the water movement is anisotropic. The diffusion in the x-axis, however, does not mean no diffusion occurs in the z or y-direction, but the primary diffusion occurs in the x-direction for the water molecules. The water diffusion is not affected by the addition of either magnesium or choline, and will still diffuse along the chains. The water can form bridges between the chains via hydrogen bonds; these positions are very favourable. When water is forming these bridges, the diffusivity is decreased dramatically for that individual molecule. Again this decreased movement is seen across all systems, regardless if magnesium is present in the system. The behaviour of the water agrees with both experimental and classical mechanical results. This agreement is a positive result and shows the accuracy of DFT.

The magnesium ions do not appear to diffuse in the SF structure, this will be the result of a combination of the Mg ion being more significant in size than the water molecules and that the simulations were not able to run for an extended period. The Mg ion does not change behaviour in different positions like the water molecules, i.e. no diffusion has been observed regardless of the position of the Mg. Mg does not affect the behaviour of water and does not appear to make the water positions any more or less favourable.

Choline, similar to Mg ions, does not move in SF, due to the size of the molecule and structure. It is expected that choline with a crystalline structure will not move significantly at a low temperature. Looking at the choline simulations it is evident that the choline nitrate is not the correct structure, e.g. it should not be in a crystal form, and it will not be a single molecule but two different molecules to form the ionic liquid. This assumption is evidenced by the structure of both the SF and choline nitrate structure deteriorating in a short time frame. The results of the full cell show that further work is needed to investigate the structure of choline nitrate and how it will behave when near the silk structure.

5.4 How does the structure of SF behave when undergoing vibrational excitement?

The IR of the silk has confirmed the structure of the SF. Where the alanine and glycine spectra peaks can be linked to the peaks in the SF IR. The secondary structure can be elucidated from the IR spectrum, which, as shown in the results, strongly agrees with the secondary structure found via producing Ramachandran plots. An important note is that the IR spectra can be reproduced from theory alone, allowing IR spectra to be generated

for a range of SF structures with high confidence that they will accurately resemble the experimental spectrum.

5.5 How does the SF material behave under varying temperatures?

The behaviour of SF under varying temperature has been investigated in three different methods - Ramachandran plots, IR performed at increasing temperatures and investigation of thermal expansion and thermal expansion coefficients. The Ramachandran plots show the material begins to lose its' structure at the higher temperature; this is expected as it will start to undergo a phase change. The results of the IR spectrum reinforce this at higher temperatures where the peaks become broader, and there is more noise in the spectrum. Demonstrating the secondary structure is breaking down, and the bonds are less rigid. This shows good agreement with the Ramachandran plots at higher temperatures. The thermal expansion shows that the material behaves isotropically when hydrated, I.e. all parameters behave uniformly as the temperature is increased from 0 K to 673 K. Thermal expansion shows the material when hydrated contracts, this agrees with experiment. An experiment has shown good agreement with the data in the range 273 K to 473 K. The thermal expansion coefficients have shown the material will expand at a similar rate above 273 K. Below this the behaviour will change as the SF will undergo a phase change and begin to turn more crystalline, resulting in drastically different coefficients. The thermal expansion of the cells shows a significant change between the hydrated and non- hydrated states. Where the volume of a hydrated silk cell decreases as the temperature increases in contrast to the non-hydrated cell where the volume increases.

5.6 How does DFT data compare to classical mechanics and experimental data?

All water simulations performed have replicated the results seen classically providing a strong basis to confirm that both the SF model is correct and accurate and that both DFT and classical mechanics are valid methods for elucidating this material. The thermal expansion or contraction as it appears has agreed with experiment; no classical data has been gathered for the thermal expansions. The Ramachandran plots have demonstrated agreement with both the classical results and experimental results. The DFT data has been shown to align more accurately with the experimental results than the classical results do. This is apparent as DFT has included data points in the top left β sheet region, which is missing in the classical result. The experimental IR can be accurately reproduced from the theory of DFT. The IR

spectra produced have shown good agreement of the position of the prominent peak at 1650 cm^{-1} in experimental results, this peak generated by the DFT moves with the change in temperature, but the peak is generally recreated in a similar position as shown in the results section.

The accuracy of the DFT method used, BLYP and DZVP, has been shown consistently in the calculations performed. With the DFT repeatedly outperforming the classical mechanical simulations.

6 Future Work

Further to work contained in this thesis, there are several ideas to continue this work going forwards. There are a range of barriers to the future of the project.

For future projects, the computational resource will be a limiting factor. More complex calculations and a higher number of calculations will be required. To be able to run this more computational time will be required on ARCHER, where the HEC will not be sufficient to run the chosen calculations.

The availability of basis sets and functionals have restricted which atoms can be simulated in specific methods - BLYP with DZVP has been chosen due to the broad range of atoms and low computational cost. However, a second basis set would have to be chosen if, for example, Ca^{2+} was to be included in the system. Another suggestion would be to choose a more extensive basis set such as B3LYP which is a hybrid exchange-correlation functional and may provide more accurate results although BLYP and DZVP have successfully replicated experimental results and as such is sufficient for the project so far. Changing the basis set and functional to B3LYP and TZVP would also increase the computational cost by a significant amount. Simulations via B3LYP is unlikely to be feasible for a system of this size due to the cost and complexity of the calculation.

Scientific knowledge has a substantial effect on future projects. This is shown by having no available structure for a simplified silk molecule and had to be created specifically for this project. There have been very few simulations on SF previous to this project, and no simulations on the behaviour of choline in silk and have thus meant little to no benchmark data with experimental results in certain areas. The complexity of the task is a significant restriction as there has to be a tradeoff between what could be accomplished in 12 months compared to how long the task would take.

The future of the project will be concerned with simulating a complete SF biological battery. This will consist of five aims:

- Rerun simulations with a larger simulation cell
- Simulate a more complex SF structure
- Simulate the behaviour of different ions
- Simulate the behaviour of choline

- Simulate a range of molecules interacting with SF battery

Rerunning this project with a larger cell would provide a better comparison to classical mechanics. This project has focused on a $1 \times 1 \times 1$ cell in contrast to the $2 \times 2 \times 2$ and $4 \times 4 \times 4$ cells implemented by classical mechanics. Using a larger cell will provide a more accurate comparison, especially for producing Ramachandran plots as the number of data points will be increased substantially. A larger cell would allow for a direct comparison of the energy of the two simulation methods. This would show how accurate the force field, from classical mechanics is, in comparison to the all electron approach used by DFT.

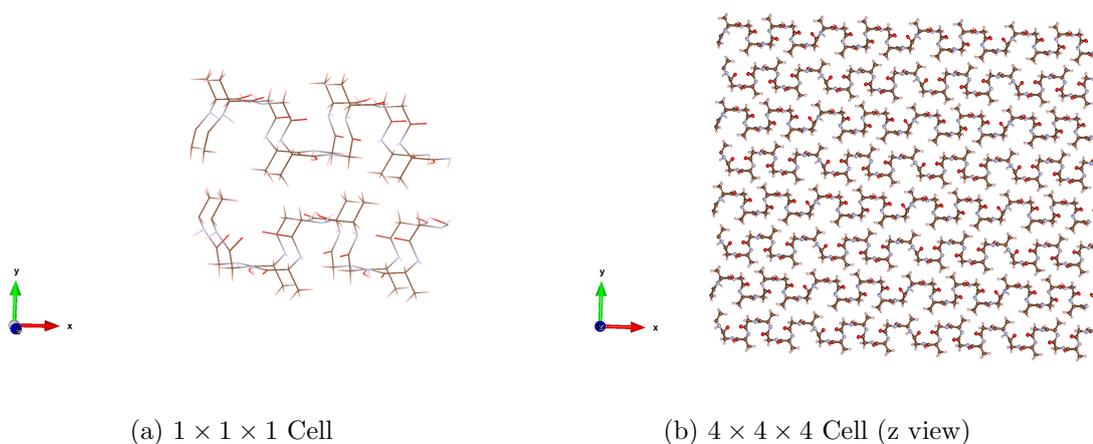


Figure 67: An image to show the size difference in silk cells. Created by the Author using Vesta [79].

To simulate a more complex SF structure. This project has focused on a repeating structure of $(-Ala-Gly-)_n$ whereas natural SF has a more complex structure of $(-Ala-Gly-Ser-Gly-Ala-Gly-)_n$. The complexity of this task is due to the lack of a preexisting template file for the complex structure; this means to simulate this a template would have to be created, optimised and compared to experimental data to validate the model. This may take an extended period depending on the first initial guess. Further to this, other side chains may be added into the structure; however, this would more accurately show synthetic silk but less accurately represent SF from *B. Mori* silkworms. Furthermore, the increase in complexity of the structure would increase the computational cost.

This project has simulated Mg^{2+} in SF, but batteries can contain ions such as Ca^{2+} and Li^{1+} , as shown in the literature review. Simulations of multiple ions would allow further calculations into the diffusivity of different ions in the silk structure, and a better understanding of the behaviour of SF with different molecules. From this, the diffusion coefficients can be

calculated and compared to provide a complete view of how the change in ion changes the diffusion coefficient and the Arrhenius plot for each ion. The diffusion coefficient and Arrhenius plot has been calculated for classical mechanics but not for the DFT approach. Further diffusion calculations were not performed due to computational limitations. A limiting factor of this is that the CP2K basis set and functionals do not contain data for Ca^{2+} and Li^{1+} at the BLYP level and has thus been excluded from this project. In future a different basis set and functional can be chosen to allow for these ions to be simulated.

As shown in the results and discussion, the choline molecules have been simulated. After a comparison to literature and analysis of the results, it was concluded that the choline nitrate structure used was incorrect. Future projects should aim to successfully simulate choline nitrate as two singular molecules to accurately represent the ionic liquid. This correct form may lead to diffusion of the choline through the cell as literature has shown. The choline in the correct form may also change the behaviour of the full cell as shown in the results.

Further development of this project will provide a more complete idea of how the battery will behave and decay in the body. The mechanism of the decay of silk is not well known and further work will show how the silk breaks down over time. The rate of decay will be determined by the thickness of the SF protective layer. This rate and thickness can be predicted through the DFT simulations. This knowledge is vital for controlling how long the battery will be present in the body and thus the battery can be tailored for individual operations.

The final results of the experimental batteries are to be used in the body and as such, there will be more molecules present than just the simulated SF, choline, Mg and water. To more precisely simulate the working environment a more extensive range of molecules would have to be included in the cell, this would include but is not limited to lipids, other ions such as Ca^{2+} , ATP (adenosine triphosphate) and enzymes produced by the body. Due to the size of some of these molecules, it may be unfeasible to simulate all molecules via DFT but would be more realistic to use classical mechanics to simulate the full cell, while using DFT to investigate single molecules and the interactions between molecules. This research has been excluded from this project due to time constraints and computational time.

Supporting Information

Ab Initio Simulations of the Degradation of Biodegradable Batteries

National HPC resource:	Archer
Consortium theme:	Energy Generation, Storage and Transport
Project leader:	Dr Samuel Murphy
Investigator(s):	Benjamin Deacon, Dr Samuel Murphy
Principal user(s):	Benjamin Deacon
Total AU allocation request:	2,488.32 kAUs
Estimate of job type:	Standard
Storage space required (Gb):	home 30Gb, work 300Gb
Code(s) to be used:	CP2K
Other national HEC funding:	None

Description of project

Transient implantable medical bionics (TIMBs), such as, biodegradable batteries that disappear after their operation, are gaining attention because they potentially facilitate the deployment of novel instructive biomaterials for regenerative medicine. Implantable biodegradable and biocompatible batteries may be capable of satisfying the power requirements of some biomedical devices before harmlessly degrading.

One material of particular interest for the construction of biodegradable batteries is Bombyx Mori silk. At Lancaster we are developing a biodegradable battery that will utilise silk both in the electrolyte and to encase the battery. Using the silk offers the battery a degree of protection that enables the device to operate for a number of days before it harmlessly degrades. Key to tuning the lifetime of the battery is understanding how the structure of the silk changes under different operating conditions and how this changes the diffusivity of the cations (i.e. Mg^{2+}) and other species such as choline nitrate used as the ionic liquid in the electrolyte.

Therefore, this project will examine structural transformations and specie mobility in a simplified silk system constructed from chains of alternating glycine and alanine proteins using *ab initio* molecular dynamics.

Justification of resources requested

The simulations will use the DFT+D3 approach and two different exchange correlation functionals, PBE and BLYP in the cp2k code. The underlying silk backbone contains 272 atoms, and this is increased depending on the weight percentage of water required (typically 5% by weight) and the concentration of ions (i.e. Mg^{2+} and Ca^{2+}) and choline nitrate, chloride and phosphate). For each functional we will explore systems comprised of the silk, water, Mg^{2+} or Ca^{2+} with each of choline nitrate, chloride and phosphate. The simulations will be seeded randomly with water, cations and choline molecules, therefore to provide some

statistics we will consider three different initial configurations leading to a total of nine simulations for each cation and functional. Each different structure will be MD simulations will be run for 5000 steps and these simulations will require 48 hours on 4 nodes.

A summary of the planned simulations has been presented below:

Type of Simulation	Number of Simulations	Time	kAUs per job	Total kAUs
Fully hydrated Silk + Mg ion	18	48:00:00	69.120	1,244.16
Fully hydrated Silk + Ca ion	18	48:00:00	69.120	1,244.16
Total				2,488.32

Justification the choice of software

CP2K has been used as it is open source software that run very efficiently on Archer and supports the functionals required for this study.

```
&GLOBAL
  PROJECT ORGANIC_MOL
  RUN_TYPE GEO_OPT
  PRINT_LEVEL MEDIUM
&END GLOBAL
&FORCE_EVAL
  METHOD QS
  &SUBSYS
    &CELL
      ABC 4.30833 6.85699 2.44154
    &END CELL
    &TOPOLOGY
      COORD_FILE_NAME alanine.xyz
      COORDINATE xyz
    &END TOPOLOGY
    &KIND H
      BASIS_SET DZVP-GTH-PBE
      POTENTIAL GTH-PBE-q1
    &END KIND
    &KIND O
      BASIS_SET DZVP-GTH-PBE
      POTENTIAL GTH-PBE-q6
    &END KIND
    &KIND C
      BASIS_SET DZVP-GTH-PBE
      POTENTIAL GTH-PBE-q4
    &END KIND
    &KIND N
      BASIS_SET DZVP-GTH-PBE
      POTENTIAL GTH-PBE-q5
    &END KIND
  &END SUBSYS
&DFT
  BASIS_SET_FILE_NAME ./BASIS_SET
  POTENTIAL_FILE_NAME ./POTENTIAL
  &QS
    EPS_DEFAULT 1.0E-10
  &END QS
  &MGRID
    CUTOFF 300
    NGRIDS 4
    REL_CUTOFF 60
  &END MGRID
  &SCF
    SCF_GUESS ATOMIC
    EPS_SCF 1.0E-07
    MAX_SCF 300
    &DIAGONALIZATION T
      ALGORITHM STANDARD
    &END DIAGONALIZATION
    &MIXING T
      ALPHA 0.4
      METHOD KERKER_MIXING
    &END MIXING
```

```
&PRINT
  &RESTART OFF
  &END RESTART
&END PRINT
&END SCF
&XC
  &XC_FUNCTIONAL PBE
  &END XC_FUNCTIONAL
  &vdW_POTENTIAL
    DISPERSION_FUNCTIONAL PAIR_POTENTIAL
    &PAIR_POTENTIAL
      TYPE DFTD3
      CALCULATE_C9_TERM .TRUE.
      PARAMETER_FILE_NAME ./dftd3.dat
      REFERENCE_FUNCTIONAL PBE
      D3_SCALING 0.00000000E+000 0.00000000E+000
0.00000000E+000
      R_CUTOFF 2
    &END PAIR_POTENTIAL
  &END vdW_POTENTIAL
&END XC
&END DFT
&PRINT
  &FORCES ON
  &END FORCES
&END PRINT
&END FORCE_EVAL
&MOTION
  &GEO_OPT
    TYPE MINIMIZATION
    MAX_DR 1.0E-03
    MAX_FORCE 1.0E-03
    RMS_DR 1.0E-03
    RMS_FORCE 1.0E-03
    MAX_ITER 200
    OPTIMIZER CG
  &CG
    MAX_STEEP_STEPS 0
    RESTART_LIMIT 9.0E-01
  &END CG
&END GEO_OPT
&END MOTION
```

```

&FORCE_EVAL
METHOD Quickstep
&DFT
BASIS_SET_FILE_NAME ./BASIS_MOLOPT
POTENTIAL_FILE_NAME ./GTH_POTENTIALS
&PRINT
&MOMENTS
PERIODIC FALSE
&END
&END
&QS
EPS_DEFAULT 1.0E-12
EXTRAPOLATION ASPC
EXTRAPOLATION_ORDER 3
&END QS
&POISSON
PERIODIC NONE
POISSON_SOLVER MT
&END POISSON
&MGRID
CUTOFF 350
NGRIDS 5
&END
&SCF
MAX_SCF 100
SCF_GUESS RESTART
EPS_SCF 2.0E-7
&DIAGONALIZATION
ALGORITHM STANDARD
&END
&OUTER_SCF
MAX_SCF 15
EPS_SCF 2.0E-7
&END
! number of unoccupied states required for XAS
ADDED_MOS 40
&PRINT
&RESTART
&EACH
QS_SCF 0
GEO_OPT 1
&END
ADD_LAST NUMERIC
FILENAME RESTART
&END
&RESTART_HISTORY OFF
&END
&END
&END SCF
&XC
&XC_FUNCTIONAL PBE
&END XC_FUNCTIONAL
&END XC
&END DFT

```

```

&SUBSYS
  &CELL
    ABC [angstrom] 17.8 15.7558 11.4904
  &END
  &TOPOLOGY
    COORD_FILE_NAME ./silk.xyz
    COORDINATE xyz
  &END
  &KIND 0
    BASIS_SET TZV2P-MOLOPT-GTH
    POTENTIAL GTH-PBE-q6
  &END KIND
  &KIND C
    BASIS_SET TZV2P-MOLOPT-GTH
    POTENTIAL GTH-PBE-q4
  &END KIND
  &KIND H
    BASIS_SET TZV2P-MOLOPT-GTH
    POTENTIAL GTH-PBE-q1
  &END KIND
  &KIND N
    BASIS_SET TZV2P-MOLOPT-GTH
    POTENTIAL GTH-PBE-q5
  &END KIND
&END SUBSYS
&END FORCE_EVAL
&GLOBAL
  PRINT_LEVEL LOW
  PROJECT IR
  RUN_TYPE VIBRATIONAL_ANALYSIS
&END GLOBAL
&MOTION
  &GEO_OPT
  ! to calculate vibrational spectrum tight convergence is required
  because frequencies are very sensitive to force constant
  MAX_FORCE 0.00005
  MAX_ITER 1600
  OPTIMIZER BFGS
  &BFGS
  &END
  &END
&END
! setup parameters to perform a Normal Modes analysis
&VIBRATIONAL_ANALYSIS
! Calculation of the IR-Intensities.
  INTENSITIES
! Specify the number of processors to be used per replica
environment (for parallel runs)
  NPROC_REP 32
! Specify the increment to be used to construct the HESSIAN with
finite difference method
  DX 0.001
  &PRINT
  &PROGRAM_RUN_INFO ON

```

&END
&END
&END

```
&GLOBAL
  PROJECT ORGANIC_MOL
  RUN_TYPE MD
  PRINT_LEVEL MEDIUM
&END GLOBAL
&FORCE_EVAL
  METHOD QS
  &SUBSYS
    &CELL
      ABC 17.8 15.7558 11.4904
    &END CELL
    &TOPOLOGY
      COORD_FILE_NAME silk.xyz
      COORDINATE xyz
    &END TOPOLOGY
    &KIND H
      BASIS_SET DZVP-GTH-BLYP
      POTENTIAL GTH-BLYP-q1
    &END KIND
    &KIND O
      BASIS_SET DZVP-GTH-BLYP
      POTENTIAL GTH-BLYP-q6
    &END KIND
    &KIND C
      BASIS_SET DZVP-GTH-BLYP
      POTENTIAL GTH-BLYP-q4
    &END KIND
    &KIND N
      BASIS_SET DZVP-GTH-BLYP
      POTENTIAL GTH-BLYP-q5
    &END KIND
  &END SUBSYS
&DFT
  BASIS_SET_FILE_NAME ./BASIS_SET
  POTENTIAL_FILE_NAME ./POTENTIAL
&QS
  EPS_DEFAULT 1.0E-7
&END QS
&MGRID
  CUTOFF 200
  NGRIDS 4
  REL_CUTOFF 30
&END MGRID
&SCF
  SCF_GUESS ATOMIC
  EPS_SCF 1.0E-05
  MAX_SCF 50
  &DIAGONALIZATION T
    ALGORITHM STANDARD
  &END DIAGONALIZATION
  &MIXING T
    ALPHA 0.4
    METHOD KERKER_MIXING
  &END MIXING
```

```

&PRINT
  &RESTART OFF
  &END RESTART
&END PRINT
&END SCF
&XC
  &XC_FUNCTIONAL BLYP
  &END XC_FUNCTIONAL
  &vdW_POTENTIAL
    DISPERSION_FUNCTIONAL PAIR_POTENTIAL
    &PAIR_POTENTIAL
      TYPE DFTD3
      CALCULATE_C9_TERM .TRUE.
      PARAMETER_FILE_NAME ./dftd3.dat
      REFERENCE_FUNCTIONAL BLYP
      D3_SCALING 0.00000000E+000 0.00000000E+000
0.00000000E+000
      R_CUTOFF 2
    &END PAIR_POTENTIAL
  &END vdW_POTENTIAL
&END XC
&END DFT
&END FORCE_EVAL
&MOTION
  &MD
    ENSEMBLE          NVT
    STEPS              20000
    TIMESTEP          0.05
    TEMPERATURE        310
    ANGVEL_ZERO                T
    ANGVEL_TOL         0.001
    COMVEL_TOL         0.001
    TEMP_TOL           200
  &THERMOSTAT
    TYPE NOSE
  &END THERMOSTAT
  &LANGEVIN
    GAMMA 0.001
  &END LANGEVIN
&END MD
&PRINT
  &TRAJECTORY SILENT
    FILENAME =Silk.xyz
  &EACH
    MD 1
  &END EACH
&END TRAJECTORY
  &VELOCITIES SILENT
    FILENAME =Silk.vel
  &EACH
    MD 1
  &END EACH
&END VELOCITIES
&FORCES SILENT

```

```
FILENAME =Silk.force
&EACH
  MD 1
&END EACH
&END FORCES
&END PRINT
&END MOTION
```

```
&GLOBAL
  PROJECT ORGANIC_MOL
  RUN_TYPE ENERGY_FORCE
  PRINT_LEVEL MEDIUM
&END GLOBAL
&FORCE_EVAL
  METHOD QS
  &SUBSYS
    &CELL
      ABC 1 1 1
    &END CELL
    &TOPOLOGY
      COORD_FILE_NAME carbon.xyz
      COORDINATE xyz
    &END TOPOLOGY
    &KIND H
      BASIS_SET DZVP-GTH-PADE
      POTENTIAL GTH-PADE-q1
    &END KIND
    &KIND O
      BASIS_SET DZVP-GTH-PADE
      POTENTIAL GTH-PADE-q6
    &END KIND
    &KIND C
      BASIS_SET DZVP-GTH-PADE
      POTENTIAL GTH-PADE-q4
    &END KIND
    &KIND N
      BASIS_SET DZVP-GTH-PADE
      POTENTIAL GTH-PADE-q5
    &END KIND
    &KIND Mg
      BASIS_SET DZVP-GTH-PADE
      POTENTIAL GTH-PADE-q10
    &END KIND
  &END SUBSYS
&DFT
  BASIS_SET_FILE_NAME ./BASIS_SET
  POTENTIAL_FILE_NAME ./POTENTIAL
  &QS
    EPS_DEFAULT 1.0E-10
  &END QS
  &MGRID
    CUTOFF 300
    NGRIDS 4
    REL_CUTOFF 60
  &END MGRID
  &SCF
    SCF_GUESS ATOMIC
    EPS_SCF 1.0E-07
    MAX_SCF 300
    &DIAGONALIZATION T
      ALGORITHM STANDARD
    &END DIAGONALIZATION
```

```
&MIXING T
  ALPHA 0.4
  METHOD KERKER_MIXING
&END MIXING
&PRINT
  &RESTART OFF
  &END RESTART
&END PRINT
&END SCF
&XC
  &XC_FUNCTIONAL PADE
  &END XC_FUNCTIONAL
  &vdW_POTENTIAL
    DISPERSION_FUNCTIONAL PAIR_POTENTIAL
    &PAIR_POTENTIAL
      TYPE DFTD3
      CALCULATE_C9_TERM .TRUE.
      PARAMETER_FILE_NAME ./dftd3.dat
      REFERENCE_FUNCTIONAL PADE
      D3_SCALING 0.00000000E+000 0.00000000E+000
0.00000000E+000
      R_CUTOFF 2
    &END PAIR_POTENTIAL
  &END vdW_POTENTIAL
&END XC
&END DFT
&PRINT
  &FORCES ON
  &END FORCES
&END PRINT
&END FORCE_EVAL
```

```

#!/usr/bin/perl

use strict;
use warnings;

#Simple supercell script
my $seedname = $ARGV[0];
open INFILE, "$seedname" or die "Can't open $seedname\n";
my @input = <INFILE>;
close INFILE;

#Number of residues in the unitcell
my $num_residue_unit = 32;

#Supercell lattice parameters
my $x_latt = 17.8;
my $y_latt = 15.7558;
my $z_latt = 11.4904;

#Supercell repetitions
my $x_reps = 1;
my $y_reps = 1;
my $z_reps = 1;

#Supercell size
my $final_x = $x_reps*$x_latt;
my $final_y = $y_reps*$y_latt;
my $final_z = $z_reps*$z_latt;

my $change_y = 0.4;
my $new_y_latt = 15.7558+$change_y*15.7558;

#Open output files (xyz for dl_poly and template file for fl_analyzer)
open OUTFILE1, ">y_increase_$change_y.xyz" or die "Can't open
  supercell-$x_reps-$y_reps-$z_reps.xyz\n";
open OUTFILE2, ">template-$x_reps-$y_reps-$z_reps" or die "Can't open
  template-$x_reps-$y_reps-$z_reps\n";

print OUTFILE1 "272\n";
print OUTFILE1 "supercell boxsize 17.8 18 11.4904\n";

print OUTFILE2 "pdb created using supercell.pl\n";
print OUTFILE2 "CRYST1 17.8 15.7558 11.4904 90.0 90.0 90.0\n";

#Initiate counter
my $counter = 0;
my $atom_counter = 0;

#Loop over all repetitions in x, y and z
for (my $s=1;$s<=$x_reps;$s++)
{
  for (my $t=1;$t<=$y_reps;$t++)
  {
    for (my $u=1;$u<=$z_reps;$u++)
    {

```

```

#Loop over all atoms in unitcell
for (my $atom=0;$atom<272;$atom++)
{
    #For each atom split calculate position in new supercell
    my @splitline = split(/\s+/, $input[$atom+2]);
    my $atom = $splitline[0];
    my $atom_no = $splitline[1];
    my $atom_name = $splitline[2];
    my $residue_name = $splitline[3];
    my $residue_type = $splitline[4];
    my $residue_no = $splitline[5];
    my $initial_x = $splitline[6];
    my $initial_y = $splitline[7];
    my $initial_z = $splitline[8];
    my $beta = $splitline[9];
    my $weight = $splitline[10];
    my $atom_type = $splitline[11];

    #Convert to fractional
    $initial_x = $initial_x/$x_latt;
    $initial_y = $initial_y/$y_latt;
    $initial_z = $initial_z/$z_latt;

    my $current_residue_count =
        ($counter*$num_residue_unit)+$residue_no;

    #Convert to new fraction co-ordinates
    my $final_atom_x = $initial_x/$x_reps;
    my $final_atom_y;
    if ($residue_type eq 'A' or $residue_type eq 'C')
    {
        $final_atom_y = $initial_y/$y_reps + $change_y;
    }
    else
    {
        $final_atom_y = $initial_y/$y_reps;
    }

    my $final_atom_z = $initial_z/$z_reps;

    if ($final_atom_x < 0)
    {
        $final_atom_x = $final_atom_x+1;
    }
    elsif ($final_atom_x >= 1.0)
    {
        $final_atom_x = $final_atom_x-1;
    }
    if ($final_atom_y < 0)
    {
        $final_atom_y = $final_atom_y+1;
    }
    elsif ($final_atom_y >= 1.8)
    {

```

```

        $final_atom_y = $final_atom_y-1;
    }
    if ($final_atom_z < 0)
    {
        $final_atom_z = $final_atom_z+1;
    }
    elsif ($final_atom_z >= 1.0)
    {
        $final_atom_z = $final_atom_z-1;
    }

    #Calculate final cartesian co-ordinates
    my $cart_x = $final_x*$final_atom_x;
    my $cart_y = $final_y*$final_atom_y;
    my $cart_z = $final_z*$final_atom_z;

    $atom_counter++;

    print OUTFILE1 "$atom_type $cart_x $cart_y $cart_z
    $residue_type\n";
    print OUTFILE2 sprintf("ATOM %4.0f   %s   %s %5.0f
    %7.3f %7.3f %7.3f 1.00 0.00           %s\n",
    $atom_counter, $atom_name, $residue_type,
    $current_residue_count, $cart_x, $cart_y, $cart_z,
    $atom_name);

    }
    $counter++;
}
}
}

```

```

#!/usr/bin/perl

use strict;
use warnings;

#Simple supercell script
my $seedname = $ARGV[0];
open INFILE, "$seedname" or die "Can't open $seedname\n";
my @input = <INFILE>;
close INFILE;

#Number of residues in the unitcell
my $num_residue_unit = 32;

#Supercell lattice parameters
my $x_latt = 17.8;
my $y_latt = 15.7558;
my $z_latt = 11.4904;

#Supercell repetitions
my $x_reps = 1;
my $y_reps = 1;
my $z_reps = 1;

#Supercell size
my $final_x = $x_reps*$x_latt;
my $final_y = $y_reps*$y_latt;
my $final_z = $z_reps*$z_latt;

my $change_z = 0.7;

#Open output files (xyz for dl_poly and template file for fl_analyzer)
open OUTFILE1, ">z_increase_$change_z.xyz" or die "Can't open
  supercell-$x_reps-$y_reps-$z_reps.xyz\n";
open OUTFILE2, ">template-$x_reps-$y_reps-$z_reps" or die "Can't open
  template-$x_reps-$y_reps-$z_reps\n";

print OUTFILE1 "272\n";
print OUTFILE1 "supercell boxsize 17.8 15.7558 30\n";

print OUTFILE2 "pdb created using supercell.pl\n";
print OUTFILE2 "CRYST1 17.8 15.7558 11.4904 90.0 90.0 90.0\n";

#Initiate counter
my $counter = 0;
my $atom_counter = 0;

#Loop over all repetitions in x, y and z
for (my $s=1;$s<=$x_reps;$s++)
{
  for (my $t=1;$t<=$y_reps;$t++)
  {
    for (my $u=1;$u<=$z_reps;$u++)
    {

```

```

#Loop over all atoms in unitcell
for (my $atom=0;$atom<272;$atom++)
{
    #For each atom split calculate position in new supercell
    my @splitline = split(/\s+/, $input[$atom+2]);
    my $atom = $splitline[0];
    my $atom_no = $splitline[1];
    my $atom_name = $splitline[2];
    my $residue_name = $splitline[3];
    my $residue_type = $splitline[4];
    my $residue_no = $splitline[5];
    my $initial_x = $splitline[6];
    my $initial_y = $splitline[7];
    my $initial_z = $splitline[8];
    my $beta = $splitline[9];
    my $weight = $splitline[10];
    my $atom_type = $splitline[11];

    #Convert to fractional
    $initial_x = $initial_x/$x_latt;
    $initial_y = $initial_y/$y_latt;
    $initial_z = $initial_z/$z_latt;

    my $current_residue_count =
        ($counter*$num_residue_unit)+$residue_no;

    #Convert to new fraction co-ordinates
    my $final_atom_x = $initial_x/$x_reps;
    my $final_atom_z;
    my $final_atom_y = $initial_y/$y_reps + 0;

    if ($residue_type eq 'A' or $residue_type eq 'B')
    {
        $final_atom_z = $initial_z/$z_reps + $change_z;
    }
    else
    {
        $final_atom_z = $initial_z/$z_reps;
    }

    if ($final_atom_x < 0)
    {
        $final_atom_x = $final_atom_x+1;
    }
    elseif ($final_atom_x >= 1.0)
    {
        $final_atom_x = $final_atom_x-1;
    }
    if ($final_atom_y < 0)
    {
        $final_atom_y = $final_atom_y+1;
    }
    elseif ($final_atom_y >= 1.0)
    {

```

```

        $final_atom_y = $final_atom_y-1;
    }
    if ($final_atom_z < 0)
    {
        $final_atom_z = $final_atom_z+1;
    }
    elsif ($final_atom_z >= 1.8)
    {
        $final_atom_z = $final_atom_z-1;
    }

    #Calculate final cartesian co-ordinates
    my $cart_x = $final_x*$final_atom_x;
    my $cart_y = $final_y*$final_atom_y;
    my $cart_z = $final_z*$final_atom_z;

    $atom_counter++;

    print OUTFILE1 "$atom_type $cart_x $cart_y $cart_z
    $residue_type\n";
    print OUTFILE2 sprintf("ATOM %4.0f   %s   %s %5.0f
    %7.3f %7.3f %7.3f 1.00 0.00           %s\n",
    $atom_counter, $atom_name, $residue_type,
    $current_residue_count, $cart_x, $cart_y, $cart_z,
    $atom_name);

    }
    $counter++;
}
}
}

```

```
1 #!/usr/bin/perl
2
3 use strict;
4 use warnings;
5
6 #Simple supercell script
7 my $seedname = $ARGV[0];
8 open INFILE, "$seedname" or die "Can't open $seedname\n";
9 my @input = <INFILE>;
10 close INFILE;
11
12 #Number of residues in the unitcell
13 my $num_residue_unit = 32;
14
15 #Supercell lattice parameters
16 my $x_latt = 17.8;
17 my $y_latt = 15.7558;
18 my $z_latt = 11.4904;
19
20 #Supercell repetitions
21 my $x_reps = 1;
22 my $y_reps = 1;
23 my $z_reps = 1;
24
25
26 #Open output files (xyz for dl_poly and template file for fl_analyzer)
27 open OUTFILE1, ">supercell-$x_reps-$y_reps-$z_reps.xyz" or die "Can't
... open supercell-$x_reps-$y_reps-$z_reps.xyz\n";
28 open OUTFILE2, ">template-$x_reps-$y_reps-$z_reps" or die "Can't open
... template-$x_reps-$y_reps-$z_reps\n";
29
30 print OUTFILE1 "272\n";
31 print OUTFILE1 "supercell boxsize 17.8 15.7558 11.4904\n";
32
33 print OUTFILE2 "pdb created using supercell.pl\n";
34 print OUTFILE2 "CRYST1 17.8 15.7558 11.4904 90.0 90.0 90.0\n";
35
36 #Initiate counter
37 my $counter = 0;
38 my $atom_counter = 0;
39
40 #Loop over all repetitions in x, y and z
41 for (my $s=1;$s<=$x_reps;$s++)
42 {
43     for (my $t=1;$t<=$y_reps;$t++)
44     {
45         for (my $u=1;$u<=$z_reps;$u++)
46         {
47             #Loop over all atoms in unitcell
```

```
48     for (my $atom=0;$atom<276;$atom++)
49     {
50         #For each atom split calculate position in new supercell
51         my @splitline = split(/\s+/, $input[$atom+2]);
52         my $atom = $splitline[0];
53         my $atom_no = $splitline[1];
54         my $atom_name = $splitline[2];
55         my $residue_name = $splitline[3];
56         my $residue_type = $splitline[4];
57         my $residue_no = $splitline[5];
58         my $cart_x = $splitline[6];
59         my $cart_y = $splitline[7];
60         my $cart_z = $splitline[8];
61         my $beta = $splitline[9];
62         my $weight = $splitline[10];
63         my $atom_type = $splitline[11];
64
65         $atom_counter++;
66
67         print OUTFILE1 "$atom_type $cart_x $cart_y $cart_z
... $residue_type $residue_no $atom_name\n";
68         print OUTFILE2 sprintf("ATOM   %4.0f  %3s %1s %1s
... %2.0f      %7.3f %7.3f %7.3f  1.00 0.00          %s\n", $atom_no,
... $atom_name, $residue_name, $residue_type, $residue_no, $cart_x, $cart_y,
... $cart_z, $atom_type);
69
70     }
71     $counter++;
72 }
73 }
74 }
75
76
```

The Royal Society (U.K.) - License Terms and Conditions

This is a License Agreement between Benjamin Deacon ("You") and The Royal Society (U.K.) ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by The Royal Society (U.K.), and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	21-Sep-2020	Type of Use	Republish in a thesis/dissertation
Order license ID	1064524-1	Publisher	W. Bowyer and J. Nichols for Lockyer Davis, printer to the Royal Society
ISSN	0261-0523	Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Philosophical transactions of the Royal Society of London	Country	United Kingdom of Great Britain and Northern Ireland
Author/Editor	Royal Society (Great Britain), GREW, Nehemiah, Halley, Edmund, MORTIMER, Cromwell, MUSGRAVE, William, OLDENBURG, Henry, PLOT, Robert., SLOANE, Hans, Waller, Richard, Evelyn, John	Rightsholder	The Royal Society (U.K.)
Date	01/01/1776	Publication Type	Journal
Language	English	URL	https://royalsociety.org/journals/

REQUEST DETAILS

Portion Type	Image/photo/illustration	Distribution	Worldwide
Number of images / photos / illustrations	1	Translation	Original language of publication
Format (select all that apply)	Electronic	Copies for the disabled?	No
Who will republish the content?	Academic institution	Minor editing privileges?	No
Duration of Use	Current edition and up to 5 years	Incidental promotional use?	No
Lifetime Unit Quantity	Up to 499	Currency	GBP
Rights Requested	Main product		

NEW WORK DETAILS

Title	Ab initio simulations of the degradation of biodegradable batteries	Institution name	Lancaster University
Instructor name	Samuel Murphy	Expected presentation date	2020-09-30

ADDITIONAL DETAILS

Order reference number	N/A	The requesting person / organization to appear on the license	Benjamin Deacon
------------------------	-----	---	-----------------

REUSE CONTENT DETAILS

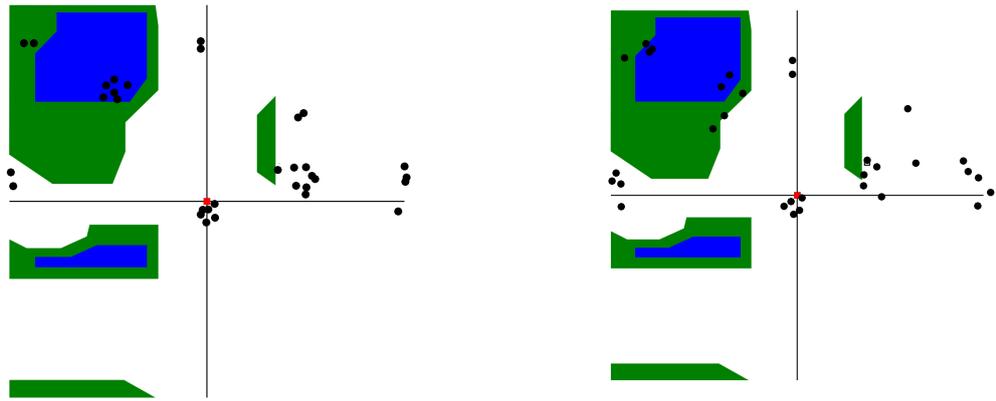
Title, description or numeric reference of the portion(s)	figure 4	Title of the article/chapter the portion is from	N/A
Editor of portion(s)	N/A	Author of portion(s)	Royal Society (Great Britain): GREW, Nehemiah.; Halley, Edmund; MORTIMER, Cromwell.; MUSGRAVE, William; OLDENBURG, Henry.; PLOT, Robert.; SLOANE, Hans; Waller, Richard; Evelyn, John
Volume of serial or monograph	N/A	Issue, if republishing an article from a serial	N/A
Page or page range of portion	430-431	Publication date of portion	1776-01-01

PUBLISHER TERMS AND CONDITIONS

Out of Copyright - any journal requested with a Publication date older than 70 years.

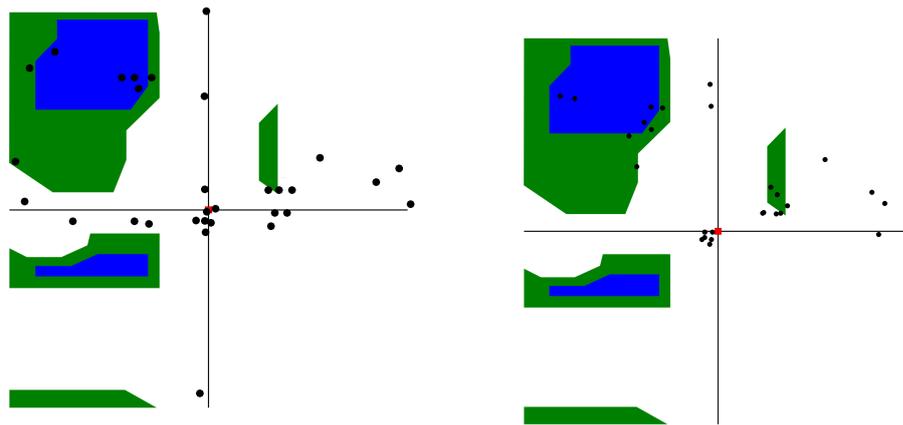
CCC Reproduction Terms and Conditions

- Description of Service; Defective Terms. This Reproduction License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Works"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republisher", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.
- The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.
- Scope of License; Limitations and Obligations.
 - All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.
 - General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 2918 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.
 - Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Order Confirmation or invoice and/or in these terms and conditions. Upon completion of the licensed use, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locations) any further copies of the Work (except for copies printed on paper in accordance with this license and still in User's stock at the end of such period).
 - In the event that the material for which a republication license is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) which are identified in such material as having been used by permission, User is responsible for identifying, and seeking separate licenses (under this Service or otherwise) for, any of such third party materials; without a separate license, such third party materials may not be used.
 - Use of proper copyright notice for a Work is required as a condition of any license granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Republished with permission of (Rightsholder's name), from (Work's title, author, volume, edition number and year of copyright); permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either immediately adjacent to the Work as used (for example, as part of a by-line or footnote but not as a separate electronic link) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.
 - User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is defamatory, violates the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.
- Indemnity. User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy or other tangible or intangible property.
- Limitation of Liability. UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for this license. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors and assigns.
- Limited Warranties. THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS". CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.
- Effect of Breach. Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the license set forth in the Order Confirmation and/or these terms and conditions, shall be a material breach of the license created by the Order Confirmation and these terms and conditions. Any breach not cured within 30 days of written notice thereof shall result in immediate termination of such license without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.
- Miscellaneous.
 - User acknowledges that CCC may, from time to time, make changes or additions to the Service or to these terms and conditions, and CCC reserves the right to send notice to the User by electronic mail or otherwise for the purposes of notifying User of such changes or additions; provided that any such changes or additions shall not apply to permissions already secured and paid for.
 - User of User-related information collected through the Service is governed by CCC's privacy policy, available online here: <https://marketplace.copyright.com/rs-uk/web/mp/privacy-policy>
 - The licensing transaction described in the Order Confirmation is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the license created by the Order Confirmation and these terms and conditions or any rights granted hereunder; provided, however, that User may assign such license in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in the new material which includes the Work(s) licensed under this Service.
 - No amendment or waiver of any terms is binding unless set forth in writing and signed by the parties. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the licensing transaction described in the Order Confirmation, which terms are in any way inconsistent with any terms set forth in the Order Confirmation and/or in these terms and conditions or CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.
 - The licensing transaction described in the Order Confirmation document shall be governed by and construed under the law of the State of New York, USA without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such licensing transaction shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court. If you have any comments or questions about the Service or Copyright Clearance Center, please contact us at 978-750-8400 or send an e-mail to support@copyright.com.



(a)

(b)



(c)

(d)

Figure 68: An image to show the Ramachandran plots at 50K. Created by the Author using VMD [101]. a) Hydrated cell run at 50K with NPT. b) Non-Hydrated cell run at 50K with NPT. c) Hydrated cell run at 50K with NVT. d) Non-Hydrated cell run at 50K with NVT.

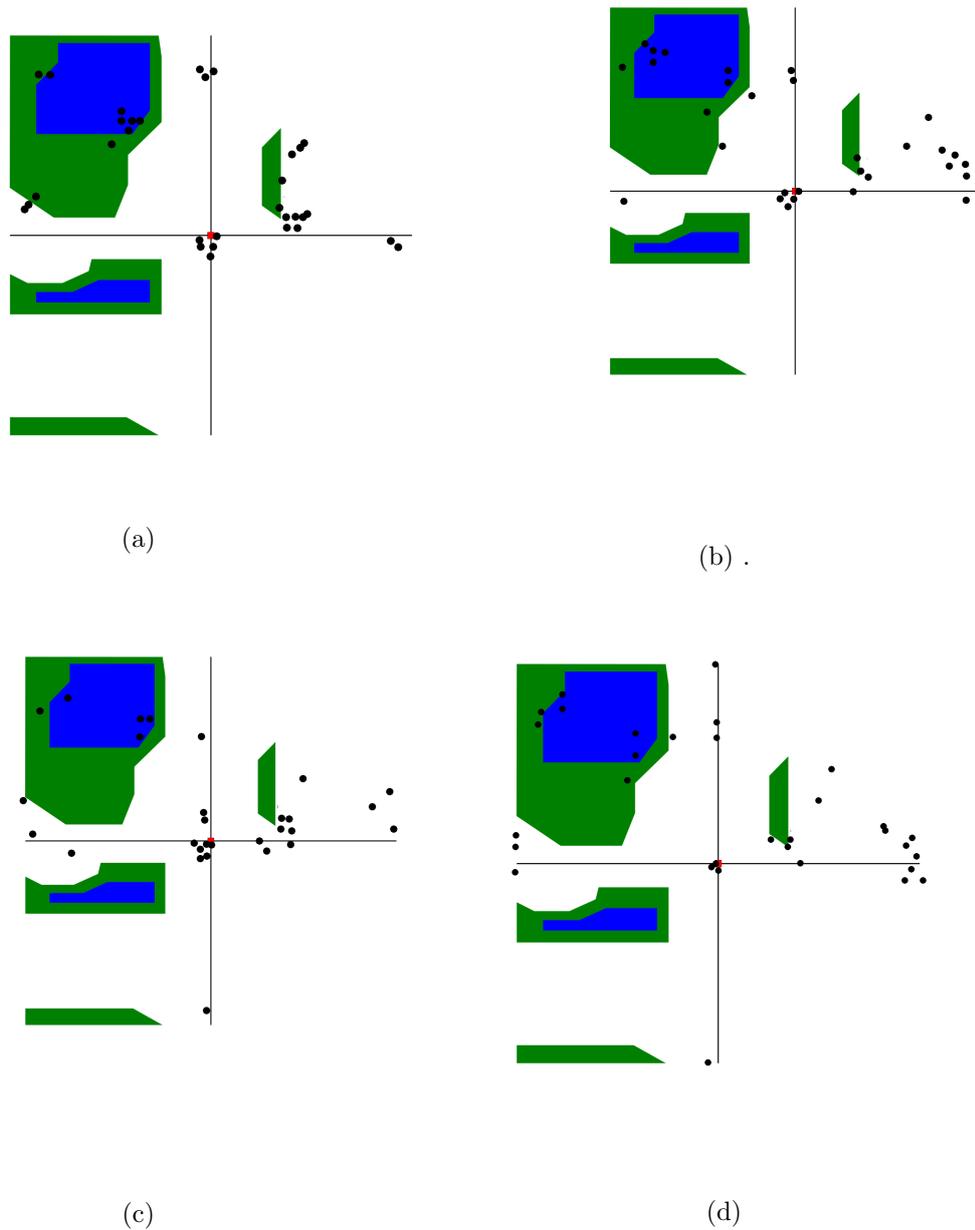
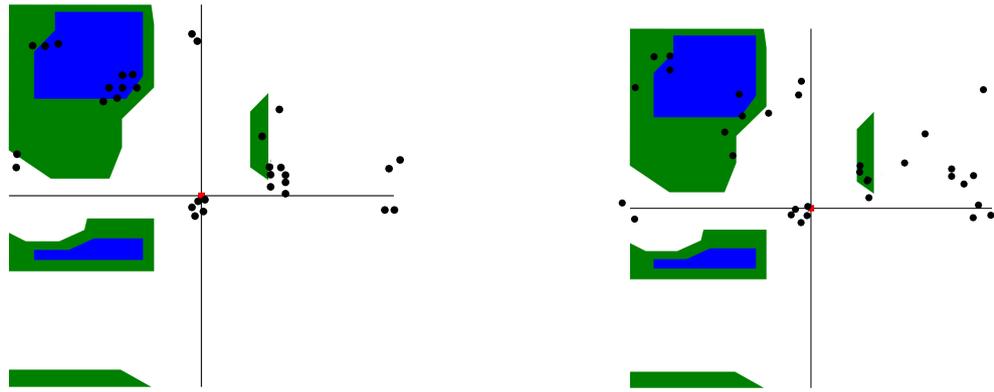
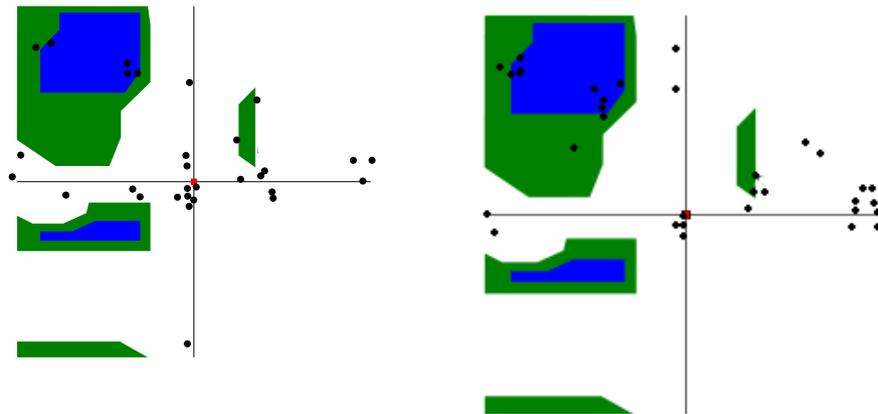


Figure 69: An image to show the Ramachandran plots at 150K. Created by the Author using VMD [101]. a) Hydrated cell run at 150K with NPT. b) Non-Hydrated cell run at 150K with NPT. c) Hydrated cell run at 150K with NVT. d) Non-Hydrated cell run at 150K with NVT.



(a)

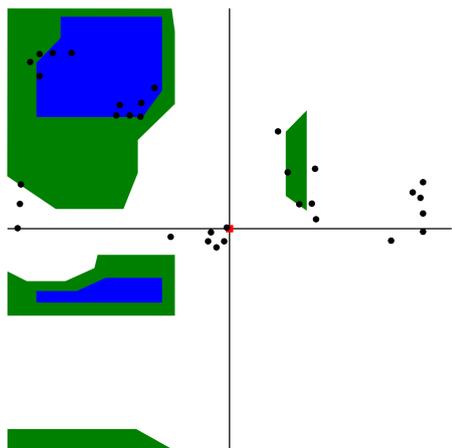
(b)



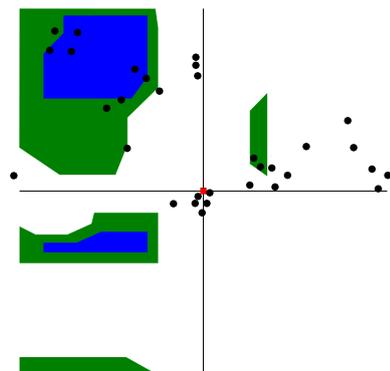
(c)

(d)

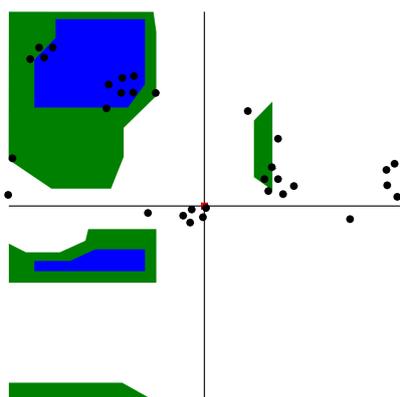
Figure 70: An image to show the Ramachandran plots at 273K. Created by the Author using VMD [101]. a) Hydrated cell run at 673K with NPT. b) Non-Hydrated cell run at 673K with NPT. c) Hydrated cell run at 673K with NVT. d) Non-Hydrated cell run at 673K with NVT.



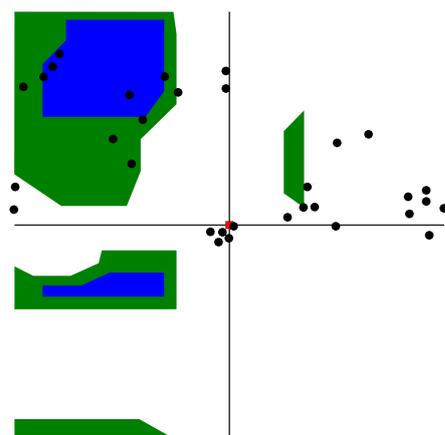
(a)



(b)



(c)



(d)

Figure 71: An image to show the Ramachandran plots at 373K. Created by the Author using VMD [101]. a) Hydrated cell run at 373K with NPT. b) Non-Hydrated cell run at 373K with NPT. c) Hydrated cell run at 373K with NVT. d) Non-Hydrated cell run at 373K with NVT.

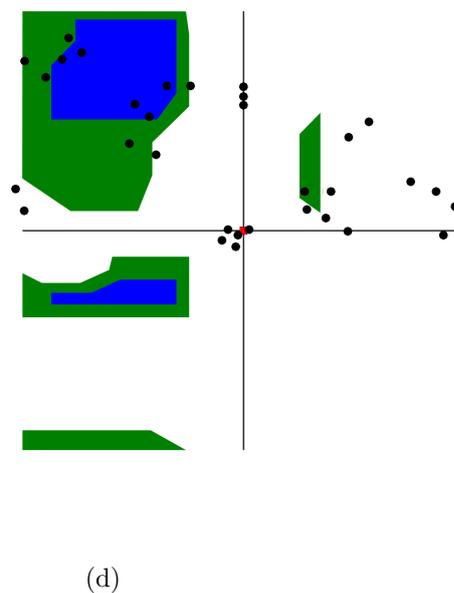
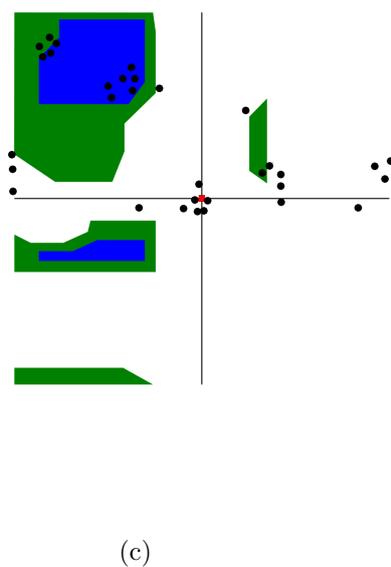
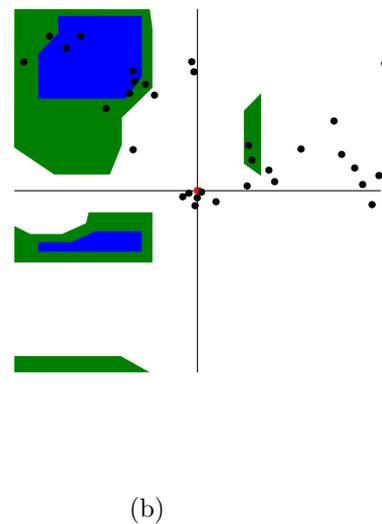
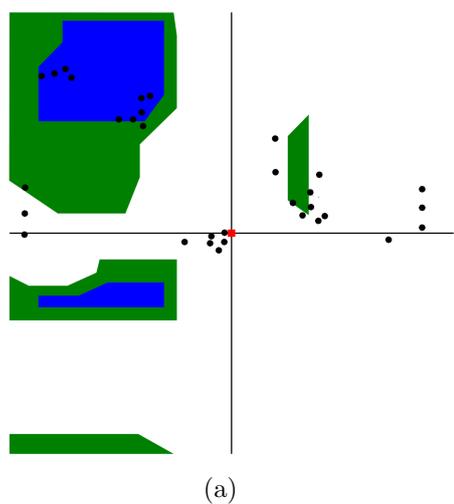
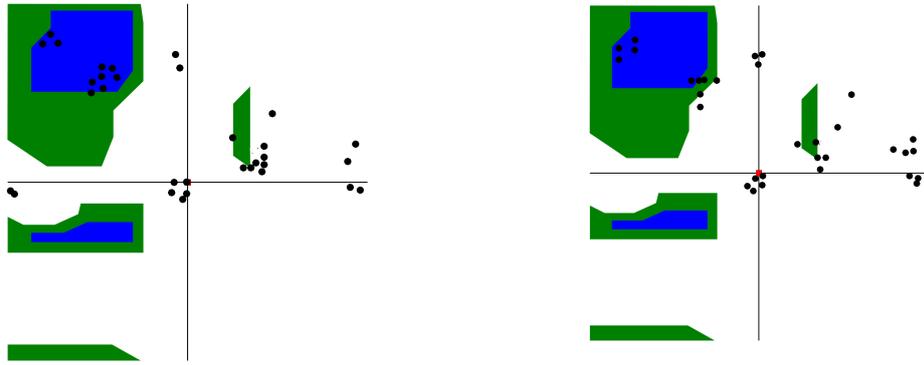
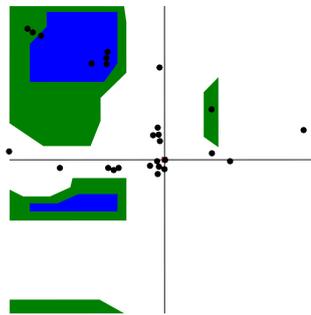


Figure 72: An image to show the Ramachandran plots at 473K. Created by the Author using VMD [101]. a) Hydrated cell run at 473K with NPT. b) Non-Hydrated cell run at 473K with NPT. c) Hydrated cell run at 473K with NVT. d) Non-Hydrated cell run at 473K with NVT.

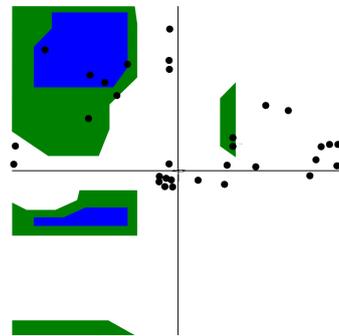


(a)

(b)



(c)



(d)

Figure 73: An image to show the Ramachandran plots at 500K. Created by the Author using VMD [101]. a) Hydrated cell run at 500K with NPT. b) Non-Hydrated cell run at 500K with NPT. c) Hydrated cell run at 500K with NVT. d) Non-Hydrated cell run at 500K with NVT.

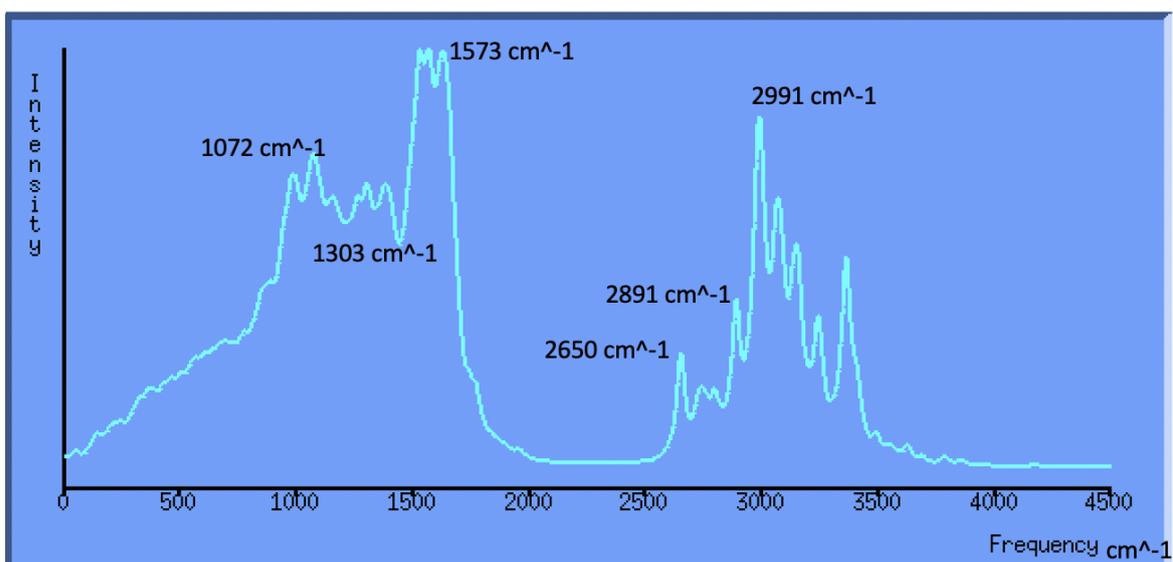


Figure 74: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 50 K. Produced using Molden [94].

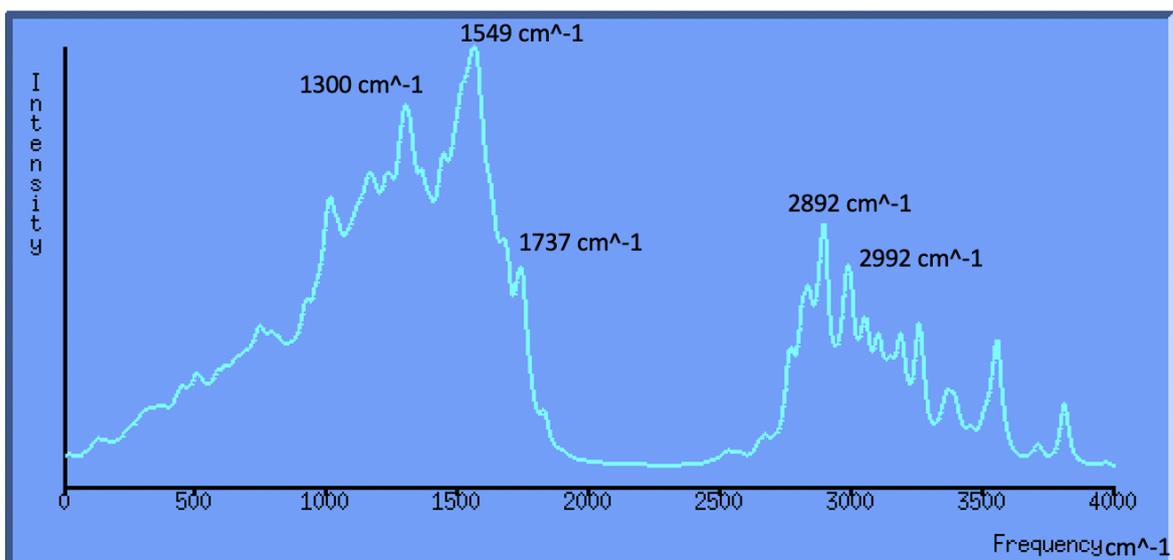


Figure 75: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 273 K. Produced using Molden [94].

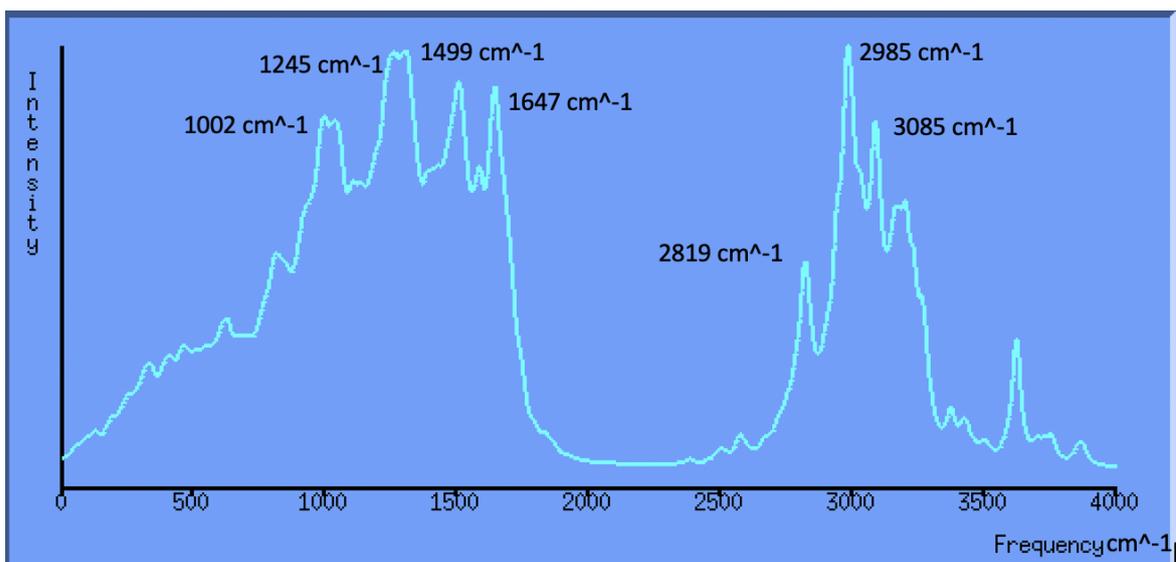


Figure 76: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 373 K. Produced using Molden [94].

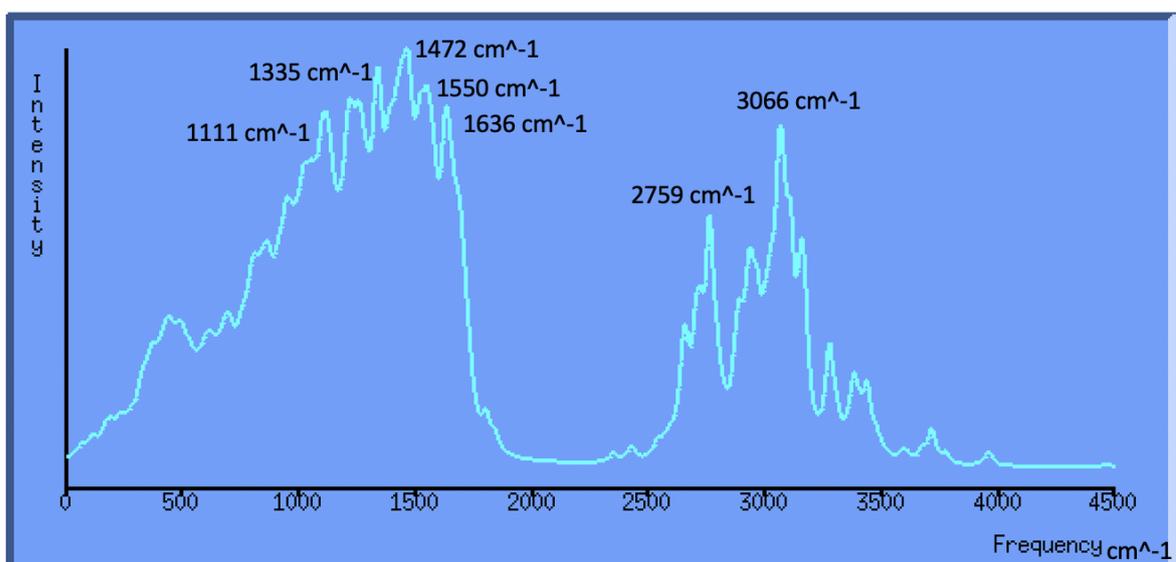


Figure 77: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 473 K. Produced using Molden [94].

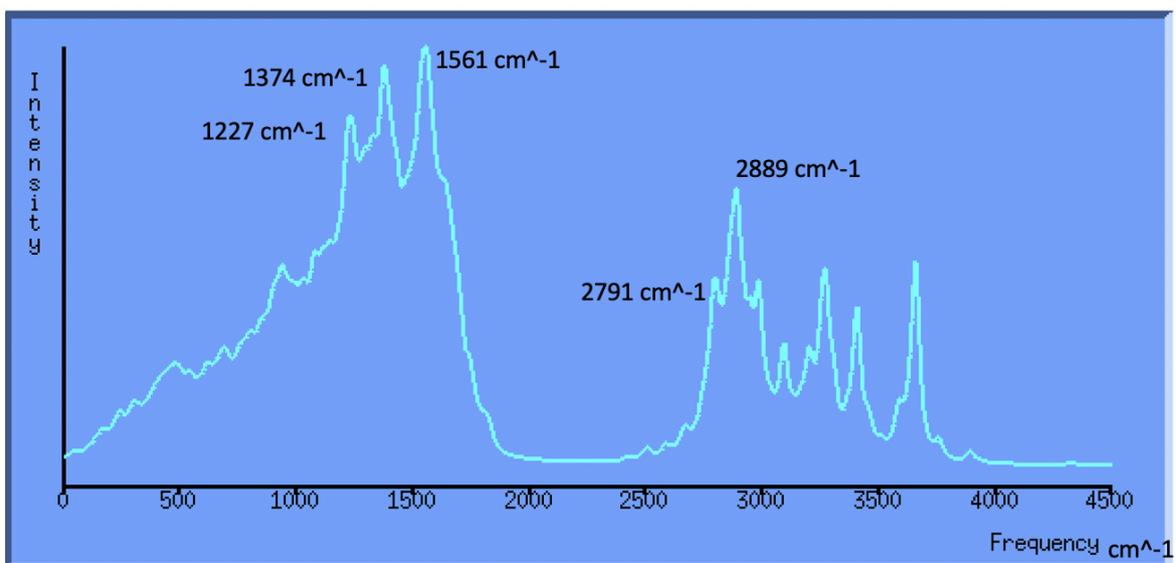


Figure 78: A figure showing the generated IR spectra of simplified SF, via DFT using MD at 500 K. Produced using Molden [94].

References

- [1] Tsutomu Yamane, Kôsukey Umemura, Yasumoto Nakazawa, and Tetsuo Asakura. Molecular dynamics simulation of conformational change of poly(ala-gly) from silk i to silk ii in relation to fiber formation mechanism of bombyx mori silk fibroin. *Macromolecules*, 36(18):6766–6772, Sep 2003.
- [2] Gregory H Altman, Frank Diaz, Caroline Jakuba, Tara Calabro, Rebecca L Horan, Jingsong Chen, Helen Lu, John Richmond, and David L Kaplan. Silk-based biomaterials. *Biomaterials*, 24(3):401–416, Feb 2003.
- [3] Greta Gronau, Sreevidhya T. Krishnaji, Michelle E. Kinahan, Tristan Giesa, Joyce Y. Wong, David L. Kaplan, and Markus J. Buehler. A review of combined experimental and computational procedures for assessing biopolymer structure–process–property relationships. *Biomaterials*, 33(33):8240–8255, Nov 2012.
- [4] Chris Holland, Keiji Numata, Jelena Rnjak-Kovacina, and F. Philipp Seib. The biomedical use of silk: Past, present, future. *Advanced Healthcare Materials*, 8(1):1800465, Sep 2018.
- [5] Xiaoteng Jia, Caiyun Wang, Vijayaraghavan Ranganathan, Bradley Napier, Changchun Yu, Yunfeng Chao, Maria Forsyth, Fiorenzo G. Omenetto, Douglas R. MacFarlane, and Gordon G. Wallace. A biodegradable thin-film magnesium primary battery using silk fibroin–ionic liquid polymer electrolyte. *ACS Energy Letters*, 2(4):831–836, Mar 2017.
- [6] Yeun-Ho Joung. Development of implantable medical devices: From an engineering perspective. *International Neurology Journal*, 17(3):98, 2013.
- [7] SuPing Lyu and Darrel Untereker. Degradability of polymers for implantable biomedical devices. *International Journal of Molecular Sciences*, 10(9):4033–4065, Sep 2009.
- [8] David C. Bock, Amy C. Marschilok, Kenneth J. Takeuchi, and Esther S. Takeuchi. Batteries used to power implantable biomedical devices. *Electrochimica Acta*, 84:155–164, Dec 2012.
- [9] Mohd Noor Islam and Mehmet R. Yuce. Review of medical implant communication system (mics) band and network. *ICT Express*, 2(4):188 – 194, 2016. Special Issue on Emerging Technologies for Medical Diagnostics.

- [10] Marko Bakula, Patrick Pelgrims, and Robert Puers. A wireless powering and communication system for implantable devices based on a royer oscillator with radio and near-field communication links. *Procedia Engineering*, 120:306 – 309, 2015. Eurosenors 2015.
- [11] Carlos Algora and Rafael Peña. Recharging the battery of implantable biomedical devices by light. *Artificial Organs*, 33(10):855–860, Oct 2009.
- [12] Dinesh Bhatia, Sweeti Bairagi, Sanat Goel, and Manoj Jangra. Pacemakers charging using body energy. *Journal of Pharmacy And Bioallied Sciences*, 2(1):51, 2010.
- [13] Kateryna Bazaka and Mohan Jacob. Implantable devices: Issues and challenges. *Electronics*, 2(4):1–34, Dec 2012.
- [14] Robert L. Hammond, Kenny Hanna, Catherine Morgan, Peter Perakis, Nader Najafi, Graham W. Long, and Charles J. Shanley. A wireless and battery-less miniature intracardiac pressure sensor. *ASAIO Journal*, 58(1):83–87, 2012.
- [15] Qingyun Ma, Mohammad Rafiqul Haider, and Yehia Massoud. A low-loss rectifier unit for inductive-powering of biomedical implants. *2011 IEEE/IFIP 19th International Conference on VLSI and System-on-Chip*, Oct 2011.
- [16] Haikal Ismail and Marlia M. Hanafiah. A review of sustainable e-waste generation and management: Present and future perspectives. *Journal of Environmental Management*, 264:110495, 2020.
- [17] S.A.M. Ali, S.-P. Zhong, P.J. Doherty, and D.F. Williams. Mechanisms of polymer degradation in implantable devices: I. poly(caprolactone). *Biomaterials*, 14(9):648 – 656, 1993.
- [18] Alessandro Volta. On the electricity excited by the mere contact of conducting substances of different kinds. *Philosophical Transactions of the Royal Society of London*, 90:403–431, 1800.
- [19] Detchko Pavlov. Chapter 1 - invention and development of the lead–acid battery. In Detchko Pavlov, editor, *Lead-Acid Batteries: Science and Technology (Second Edition)*, pages 3 – 32. Elsevier, Amsterdam, second edition edition, 2017.
- [20] A.M. Wilson, G. Zank, K. Eguchi, W. Xing, and J.R. Dahn. Pyrolysed silicon-containing polymers as high capacity anodes for lithium-ion batteries. *Journal of Power*

- Sources*, 68(2):195 – 200, 1997. Proceedings of the Eighth International Meeting on Lithium Batteries.
- [21] R. Borah, F.R. Hughson, J. Johnston, and T. Nann. On battery materials and methods. *Materials Today Advances*, 6:100046, 2020.
- [22] Wangda Li, Steven Lee, and Arumugam Manthiram. High-nickel nma: A cobalt-free alternative to nmc and nca cathodes for lithium-ion batteries. *Advanced Materials*, page 2002718, Jul 2020.
- [23] Nathiya Kalidas, Joakim Riikonen, Wujun Xu, Katja Lahtinen, Tanja Kallio, and Vesa Pekka Lehto. Cascading use of barley husk ash to produce silicon for composite anodes of li-ion batteries. *Materials Chemistry and Physics*, 245:122736, 2020.
- [24] Gilpin R. Robinson, Jane M. Hammarstrom, and Donald W. Olson. Graphite. *Professional Paper*, 2017.
- [25] Xiao-Guang Yang, Teng Liu, Yue Gao, Shanhai Ge, Yongjun Leng, Donghai Wang, and Chao-Yang Wang. Asymmetric temperature modulation for extreme fast charging of lithium-ion batteries. *Joule*, 3(12):3002–3019, Dec 2019.
- [26] Anna Tomaszewska, Zhengyu Chu, Xuning Feng, Simon O’Kane, Xinhua Liu, Jingyi Chen, Chenzhen Ji, Elizabeth Endler, Ruihe Li, Lishuo Liu, Yalun Li, Siqi Zheng, Sebastian Vetterlein, Ming Gao, Jiuyu Du, Michael Parkes, Minggao Ouyang, Monica Marinescu, Gregory Offer, and Billy Wu. Lithium-ion battery fast charging: A review. *eTransportation*, 1:100011, 2019.
- [27] A.G. Macdiarmid, J.C. Chiang, A.F. Richter, and A.J. Epstein. Polyaniline: a new concept in conducting polymers. *Synthetic Metals*, 18(1):285 – 290, 1987. Proceedings of the International Conference of Science and Technology of Synthetic Metals.
- [28] Jeffrey Lopez, David G. Mackanic, Yi Cui, and Zhenan Bao. Designing polymers for advanced battery chemistries. *Nature Reviews Materials*, 4(5):312–330, Apr 2019.
- [29] Brian B. Jing and Christopher M. Evans. Catalyst-free dynamic networks for recyclable, self-healing solid polymer electrolytes. *Journal of the American Chemical Society*, 141(48):18932–18937, Nov 2019.
- [30] Simon Muench, Andreas Wild, Christian Friebe, Bernhard Häupler, Tobias Janoschka, and Ulrich S. Schubert. Polymer-based organic batteries. *Chemical Reviews*, 116(16):9438–9484, Aug 2016.

- [31] Terri C. Lin, Yan Yan, Sophia C. King, Chun-Han Lai, and Sarah H. Tolbert. Fast-charging cathodes from polymer-templated mesoporous LiVPO_4F . *ACS Applied Materials and Interfaces*, 12(30):33775–33784, Jul 2020.
- [32] Alexandru Vlad, Kevin Arnould, Bruno Ernould, Louis Sieuw, Julien Rolland, and Jean-François Gohy. Exploring the potential of polymer battery cathodes with electrically conductive molecular backbone. *Journal of Materials Chemistry A*, 3(21):11189–11193, 2015.
- [33] Agus Purwanto, Cornelius Satria Yudha, U Ubaidillah, Hendri Widiyandari, Takashi Ogi, and Hery Haerudin. Nca cathode material: synthesis methods and performance enhancement efforts. *Materials Research Express*, 5(12):122001, Sep 2018.
- [34] Shi-chao Zhang, Lan Zhang, Wei-kun Wang, and Wen-juan Xue. A novel cathode material based on polyaniline used for lithium/sulfur secondary battery. *Synthetic Metals*, 160(17-18):2041–2044, Sep 2010.
- [35] Yujie Yao, Hongming Zhang, and Xianhong Wang. Polyaniline: an effective suppressor against diffusion and dissolution of polysulfides in li-s battery. *Journal of Solid State Electrochemistry*, 23(8):2559–2567, Jul 2019.
- [36] Ander Reizabal, R. Gonçalves, A. Fidalgo-Marijuan, C.M. Costa, Leyre Pérez, Jose-Luis Vilas, and S. Lanceros-Mendez. Tailoring silk fibroin separator membranes pore size for improving performance of lithium ion batteries. *Journal of Membrane Science*, 598:117678, 2020.
- [37] Rui F.P. Pereira, Ricardo Brito-Pereira, Renato Gonçalves, Marco P. Silva, Carlos M. Costa, Maria Manuela Silva, Verónica de Zea Bermudez, and Senentxu Lanceros-Méndez. Silk fibroin separators: A step toward lithium-ion batteries with enhanced sustainability. *ACS Applied Materials and Interfaces*, 10(6):5385–5394, Feb 2018.
- [38] R LINFORD and W SCHLINDWEIN. Medical applications of solid state ionics. *Solid State Ionics*, 177(19-25):1559–1565, Oct 2006.
- [39] Stefano Passerini and Boone B Owens. Medical batteries for external medical devices. *Journal of Power Sources*, 97-98:750 – 754, 2001. Proceedings of the 10th International Meeting on Lithium Batteries.
- [40] Arwa T. Kutbee, Rabab R. Bahabry, Kholod O. Alamoudi, Mohamed T. Ghoneim, Marlon D. Cordero, Amani S. Almuslem, Abdurrahman Gumus, Elhadj M. Diallo,

- Joanna M. Nassar, Aftab M. Hussain, and et al. Flexible and biocompatible high-performance solid-state micro-battery for implantable orthodontic system. *npj Flexible Electronics*, 1(1), Oct 2017.
- [41] Thang Phan Nguyen, Quang Vinh Nguyen, Van-Huy Nguyen, Thu-Ha Le, Vu Quynh Nga Huynh, Dai-Viet N. Vo, Quang Thang Trinh, Soo Young Kim, and Quyet Van Le. Silk fibroin-based biomaterials for biomedical applications: A review. *Polymers*, 11(12):1933, Nov 2019.
- [42] Alexander Stukowski. Visualization and analysis of atomistic simulation data with ovito—the open visualization tool. *Modelling and Simulation in Materials Science and Engineering*, 18(1):015012, Dec 2009.
- [43] Kenji Okuyama, R. Somashekar, Keiichi Noguchi, and Syuji Ichimura. Refined molecular and crystal structure of silk i based on ala-gly and (ala-gly)₂ser-gly peptide sequence. *Biopolymers*, 59(5):310–319, 2001.
- [44] Tetsuo Asakura and Yu Suzuki. Silk fibroin. *Encyclopedia of Polymeric Nanomaterials*, pages 1–7, 2014.
- [45] Cong-Zhao Zhou, Fabrice Confalonieri, Michel Jacquet, Roland Perasso, Zhen-Gang Li, and Joel Janin. Silk fibroin: Structural implications of a remarkable amino acid sequence. *Proteins: Structure, Function, and Genetics*, 44(2):119–122, 2001.
- [46] Tetsuo Asakura, Keiko Okushita, and Mike P. Williamson. Analysis of the structure of bombyx mori silk fibroin by nmr. *Macromolecules*, 48(8):2345–2357, Apr 2015.
- [47] Tetsuo Asakura, Tsutomu Yamane, Yasumoto Nakazawa, Tsunenori Kameda, and Kazuto Ando. Structure of bombyx mori silk fibroin before spinning in solid state studied with wide angle x-ray scattering and ¹³c cross-polarization/magic angle spinning nmr. *Biopolymers*, 58(5):521–525, 2001.
- [48] Tetsuo Asakura, Jun Ashida, Tsutomu Yamane, Tsunenori Kameda, Yasumoto Nakazawa, Kosuke Ohgo, and Kohei Komatsu. A repeated beta-turn structure in poly(ala-gly) as a model for silk i of bombyx mori silk fibroin studied with two-dimensional spin-diffusion nmr under off magic angle spinning and rotational echo double resonance edited by m. f. summers. *Journal of Molecular Biology*, 306(2):291–305, Feb 2001.

- [49] Richard E. Marsh, Robert B. Corey, and Linus Pauling. An investigation of the structure of silk fibroin. *Biochimica et Biophysica Acta*, 16:1 – 34, 1955.
- [50] Ed Vitz. Secondary protein structure in silk, June 2019.
- [51] Jack D. Dunitz. Pauling’s left-handed alpha-helix. *Angewandte Chemie International Edition*, 40(22):4167–4173, Nov 2001.
- [52] M. Boulet-Audet, F. Vollrath, and C. Holland. Identification and classification of silks using infrared spectroscopy. *Journal of Experimental Biology*, 218(19):3138–3149, Sep 2015.
- [53] Edwin Kamalha, Yuan Sheng Zheng, Yong Chun Zeng, and Mutua N. Fredrick. Ftir and waxd study of regenerated silk fibroin. *Advanced Materials Research*, 677:211–215, Mar 2013.
- [54] Masashi Sonoyama, Mitsuhiro Miyazawa, Gen Katagiri, and Hideyuki Ishida. Dynamic ft-ir spectroscopic studies of silk fibroin films. *Applied Spectroscopy*, 51(4):545–547, Apr 1997.
- [55] Freddi G, Monti P, Nagura M, Gotoh Y, and Tsukada M. Structure and molecular-conformation of tussah silk fibroin films - effect of heat-treatment. 35(5):841–847, 1997.
- [56] P. Monti, P. Taddei, G. Freddi, T. Asakura, and M. Tsukada. Raman spectroscopic characterization of bombyx mori silk fibroin: Raman spectrum of silk i. *Journal of Raman Spectroscopy*, 32(2):103–107, 2001.
- [57] Wiley spectrabase, August 2020.
- [58] G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, Jul 1963.
- [59] Scott A. Hollingsworth and P. Andrew Karplus. A fresh look at the ramachandran plot and the occurrence of standard structures in proteins. *BioMolecular Concepts*, 1(3-4):271–283, Oct 2010.
- [60] Sven Hovmöller, Tuping Zhou, and Tomas Ohlson. Conformations of amino acids in proteins. *Acta Crystallographica Section D Biological Crystallography*, 58(5):768–776, Apr 2002.

- [61] Tetsuo Asakura, Masashi Minami, Reiko Shimada, Makoto Demura, Minoru Osanai, Teruaki Fujito, Mamoru Imanari, and Anne S. Ulrich. 2h-labeling of silk fibroin fibers and their structural characterization by solid-state2h nmr. *Macromolecules*, 30(8):2429–2435, Apr 1997.
- [62] D. Van Neck T. Verstraelen. Study of charge transfer in ionic liquids with constrained density functional theory. Master’s thesis, Ghent University.
- [63] E. Scarpellini, M. Ortolani, A. Nucara, L. Baldassarre, M. Missori, R. Fastampa, and R. Caminiti. Stabilization of the tensile strength of aged cellulose paper by cholinium-amino acid ionic liquid treatment. *The Journal of Physical Chemistry C*, 120(42):24088–24097, Oct 2016.
- [64] Xiaoteng Jia, Yang Yang, Caiyun Wang, Chen Zhao, R. Vijayaraghavan, Douglas R. MacFarlane, Maria Forsyth, and Gordon G. Wallace. Biocompatible ionic liquid–biopolymer electrolyte-enabled thin and compact magnesium–air batteries. *ACS Applied Materials and Interfaces*, 6(23):21110–21117, Nov 2014.
- [65] Ksenia S. Egorova, Evgeniy G. Gordeev, and Valentine P. Ananikov. Biological activity of ionic liquids and their application in pharmaceuticals and medicine. *Chemical Reviews*, 117(10):7132–7189, Jan 2017.
- [66] Guangyao Wu, Haopeng Pang, Prasanna Ghimire, and Guobing Liu. 1h magnetic resonance spectroscopy and diffusion weighted imaging findings of medulloblastoma in 3.0t mri: A retrospective analysis of 17 cases. *Neural Regeneration Research*, 7:2554 – 2559, 2012.
- [67] P. Khomhoi, W. Sangprasert, Vannajan Lee, and Piyarat Nimmanpipug. Theoretical study of the bombyx mori silk surface functionalization: Quantum mechanical calculation of the glycine-alanine unit reacting with fluorine and molecular dynamic simulation of wettability. *Chiang Mai Journal of Science*, 37:106–115, 01 2010.
- [68] Mrinal Patel, Devendra K. Dubey, and Satinder Paul Singh. Phenomenological models of bombyx mori silk fibroin and their mechanical behavior using molecular dynamics simulations. *Materials Science and Engineering: C*, 108:110414, 2020.
- [69] Myeongsang Lee, Junpyo Kwon, and Sungsoo Na. Mechanical behavior comparison of spider and silkworm silks using molecular dynamics at atomic scale. *Physical Chemistry Chemical Physics*, 18(6):4814–4821, 2016.

- [70] B. M. Rode, C. F. Schwenk, and B. R. Randolf. Classical versus quantum mechanical simulations: The accuracy of computer experiments in solution chemistry. *Novel Approaches to the Structure and Dynamics of Liquids: Experiments, Theories and Simulations*, pages 41–52, 2004.
- [71] Andreas Bäck and Nikola Marković. Comparison of classical and quantum dynamics for collinear cluster scattering. *The Journal of Chemical Physics*, 122(14):144711, Apr 2005.
- [72] Stefan Grimme, Jens Antony, Stephan Ehrlich, and Helge Krieg. A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu. *The Journal of Chemical Physics*, 132(15):154104, Apr 2010.
- [73] Thomas M. Henderson, Benjamin G. Janesko, and Gustavo E. Scuseria. Range separation and local hybridization in density functional theory†. *The Journal of Physical Chemistry A*, 112(49):12530–12542, Dec 2008.
- [74] Erin R. Johnson and Axel D. Becke. A post-hartree-fock model of intermolecular interactions: Inclusion of higher-order corrections. *The Journal of Chemical Physics*, 124(17):174104, May 2006.
- [75] Michael E. Foster and Karl Sohlberg. Empirically corrected dft and semi-empirical methods for non-bonding interactions. *Phys. Chem. Chem. Phys.*, 12(2):307–322, 2010.
- [76] Axel D. Becke and Erin R. Johnson. A density-functional model of the dispersion interaction. *The Journal of Chemical Physics*, 123(15):154101, Oct 2005.
- [77] Erin R. Johnson and Axel D. Becke. A post-hartree-fock model of intermolecular interactions. *The Journal of Chemical Physics*, 123(2):024101, Jul 2005.
- [78] O. Anatole von Lilienfeld and Alexandre Tkatchenko. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *The Journal of Chemical Physics*, 132(23):234109, Jun 2010.
- [79] Koichi Momma and Fujio Izumi. Vesta: a three-dimensional visualization system for electronic and structural analysis. *Journal of Applied Crystallography*, 41(3):653–658, May 2008.
- [80] John P. Perdew. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

- [81] W. Kohn. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, 1965.
- [82] K. Kuchitsu, editor. *Structure Data of Free Polyatomic Molecules*, volume 23. Springer-Verlag Berlin Heidelberg, 1995.
- [83] Thomas Kühne, Marcella Iannuzzi, Mauro Ben, Vladimir Rybkin, Patrick Seewald, Frederick Stein, Teodoro Laino, Rustam Khaliullin, Ole Schütt, Florian Schiffmann, Dorothea Golze, Jan Wilhelm, Sergey Chulkov, Mohammad Hossein Bani-Hashemian, Valéry Weber, Urban Borštnik, Mathieu Taillefumier, Alice Jakobovits, Alfio Lazzaro, and Jürg Hutter. Cp2k: An electronic structure and molecular dynamics software package i. quickstep: Efficient and accurate electronic structure calculations. 03 2020.
- [84] September 2020.
- [85] Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics*, 127(11):114105, Sep 2007.
- [86] Tetsuo Asakura, Yu Suzuki, Yasumoto Nakazawa, Gregory P. Holland, and Jeffery L. Yarger. Elucidating silk structure using solid-state nmr. *Soft Matter*, 9(48):11440, 2013.
- [87] B. Lotz and H.D. Keith. Crystal structure of poly(l-ala-gly)ii: A model for silk i. *Journal of Molecular Biology*, 61(1):201 – 215, 1971.
- [88] J. O. Warwicker. The crystal structure of silk fibroins. *Transactions of the Faraday Society*, 52:554, 1956.
- [89] Chunyang Peng. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.*, 17(1):49–56, 1996.
- [90] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, Aug 1959.
- [91] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, Sep 1994.
- [92] Marwa El-Azazy. Introductory chapter: Infrared spectroscopy - a synopsis of the fundamentals and applications. *Infrared Spectroscopy - Principles, Advances, and Applications*, Mar 2019.

- [93] Giuliano Freddi, Patrizia Monti, Masanobu Nagura, Yohko Gotoh, and Masuhiro Tsukada. Structure and molecular conformation of tussah silk fibroin films: Effect of heat treatment. *Journal of Polymer Science Part B: Polymer Physics*, 35(5):841–847, Apr 1997.
- [94] G. Schaftenaar and J.H. Noordik. Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design*, 14(2):123–134, 2000.
- [95] Shweta, Eram Khan, Poonam Tandon, Rakesh Maurya, and Padam Kumar. A theoretical study on molecular structure, chemical reactivity and molecular docking studies on dalbergin and methylalbergin. *Journal of Molecular Structure*, 1183:100 – 106, 2019.
- [96] J. Hafner. Materials simulations using vasp—a quantum perspective to materials science. *Computer Physics Communications*, 177(1):6 – 13, 2007. Proceedings of the Conference on Computational Physics 2006.
- [97] Olivia Lynes, Jonathan Austin, and Andy Kerridge. Ab initio molecular dynamics studies of hydroxide coordination of alkaline earth metals and uranyl. *Physical Chemistry Chemical Physics*, 21(25):13809–13820, 2019.
- [98] Julien Bloino and Vincenzo Barone. A second-order perturbation theory route to vibrational averages and transition properties of molecules: General formulation and application to infrared and vibrational circular dichroism spectroscopies. *The Journal of Chemical Physics*, 136(12):124108, Mar 2012.
- [99] Marco Caricato. Exploring potential energy surfaces of electronic excited states in solution with the eom-ccsd-pcm method. *Journal of Chemical Theory and Computation*, 8(12):5081–5091, Jul 2012.
- [100] Jon Baker and Matt Shirel. Ab initio quantum chemistry on pc-based parallel supercomputers. *Parallel Computing*, 26(7):1011 – 1024, 2000.
- [101] William Humphrey, Andrew Dalke, and Klaus Schulten. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33 – 38, 1996.
- [102] Ping Zhou, Guiyang Li, Zhengzhong Shao, Xiaoyun Pan, and Tongyin Yu. Structure of bombyx mori silk fibroin based on the dft chemical shift calculation. *The Journal of Physical Chemistry B*, 105(50):12469–12476, Dec 2001.

- [103] Jun Magoshi, Yoshiko Magoshi, Shigeo Nakamura, Nobutami Kasai, and Masao Kakudo. Physical properties and structure of silk. v. thermal behavior of silk fibroin in the random-coil conformation. *Journal of Polymer Science: Polymer Physics Edition*, 15(9):1675–1683, 1977.
- [104]
- [105] Mário Túlio S. Rosado, Maria Leonor R.S. Duarte, and Rui Fausto. Vibrational spectra (ft-ir, raman and mi-ir) of alpha- and beta-alanine. *Journal of Molecular Structure*, 410-411:343 – 348, 1997. *Molecular Spectroscopy and Molecular Structure* 1996.
- [106]
- [107] Fred Elhi, Karl Karu, Pille Rinne, Kadi-Anne Nadel, Martin Järvekülg, Alvo Aabloo, Tarmo Tamm, Vladislav Ivaništšev, and Kaija Põhako-Esko. Understanding the behavior of fully non-toxic polypyrrole-gelatin and polypyrrole-pvdf soft actuators with choline ionic liquids. *Actuators*, 9(2):40, May 2020.

Evaluating a *Bombyx mori* silk fibroin model using a range of atomistic simulation tools

Evaluating a *Bombyx mori* silk fibroin model using a range of atomistic simulation tools

M. Haskew^{a,b}, B. Deacon^a, C. Yong^c, J. Hardy^b and S. T. Murphy^a

^aEngineering Department, Lancaster University, Bailrigg, Lancaster, LA1 4YW, UK

^bChemistry Department, Lancaster University. Bailrigg, Lancaster LA1 4YW, UK

^cDaresbury Laboratory, Warrington WA4 4FS

Abstract

Bombyx mori (B. mori) silk fibroin (SF) (a biopolymer) is a ubiquitous material due to its attractive properties (e.g. biocompatibility and biodegradability), and is utilised in a variety of applications, for instance, biomedical devices, drug delivery scaffolds and biodegradable batteries. Furthermore, transient implantable medical bionics (TIMBs), such as, biodegradable batteries that disappear after their operation, are gaining attention because they potentially facilitate the deployment of novel instructive biomaterials for regenerative medicine. In the wider context, the generation of degradable electronics potentially addresses problems associated with e-waste and these materials can influence biological processes in a controllable manner, (e.g. tissue regeneration and drug delivery via electrical stimulation). Utilising the information from experimental and computational SF studies, an appropriate SF model has been produced (alanine-glycine [Ala-Gly]_n crystal structure). By incorporating molecular dynamic (MD) and density functional theory (DFT) techniques, a unique insight of the SF material is achieved. The secondary structure of the computational model has been evaluated using Ramachandran plots under different environments (e.g. different temperatures and ensembles). In addition, mean square displacement (MSD) of water incorporated into the SF model was investigated. The diffusion coefficients, activation energies, most and least favourable positions of water, and trajectory of water diffusion through the SF model is obtained. With further

computational study and combination with experimental data, the behaviour/degradation of SF (and similar biomaterials) can be elucidated. Consequently, a greater control of the aforementioned devices may be achieved and positively affect their useful and practical applications.

1. Introduction

SF from the *B. mori* silk worm is an Ala-Gly-rich protein, which is spun from aqueous solution to produce strong and tough fibres.^{1, 2} Furthermore; SF has excellent biocompatibility, making it a much-desired biomaterial.^{3, 4} Many attempts have been made to mimic the natural process of producing robust silk filament under mild environmental conditions.^{5, 6, 7, 8} However, this has proven challenging and many of the resultant fibres have been weaker than native silk.⁹ Therefore, a greater understanding of the chemistry and properties of natural silk fibres (e.g. SF) is essential, because this would optimise the utilisation of silk for useful and practical applications, such as, TIMBs. SF has already seen use in various applications, for instance, a biodegradable *B. mori* SF can assume two distinct structures in the solid state,¹ silk I and silk II (before and after spinning, respectively). Silk I has been reported to assume a repeated β -turn type II conformation, whereas, silk II adopts a heterogeneous structure but mainly antiparallel β -sheets.¹ Silk fibres, inherently, are non-crystalline or semi-crystalline biopolymers possessing a mixture of secondary structures (e.g. β -sheets, helices, β -turns and random coils).¹² A common challenge when analysing silk via X-ray diffraction or electron diffraction studies, is silk's readiness to convert from silk I into silk II.¹³ Other experimental techniques have also been applied, such as, solid-state nuclear magnetic resonance (NMR), which is advantageous as the silk I form can be analysed without orientation or crystallisation (and simultaneous conversion into silk II).¹² Further detail has been obtained using atomistic simulations on SF structures derived from solid-state NMR methods, such as, 2D spin diffusion NMR under off magic angle spinning, rotational echo double resonance (REDOR) and ¹³C chemical shift data, as well as, X-ray diffraction data of a poly(Ala-Gly) sample.¹

The *B. mori* SF macromolecule is comprised of three segments (Heavy Chain ca. 350 kDa, Light Chain ca. 26 kDa and P25 gene ca. 25 kDa) in a ratio of 6:6:1.^{15, 16} The Heavy Chain is connected to the Light Chain via a single disulphide link, while the P25 gene has non-covalent interactions with the Heavy Chain and Light Chain.¹⁷ Furthermore, the Heavy Chain is made up of 5263 residues where glycine (Gly) is present in 45.9 %, alanine (Ala) in 30.3 %, serine (Ser) in 12.1 %, tyrosine (Tyr) in 5.3 %, valine (Val) in 1.8 % and 4.7 % of the other amino acids.¹⁵ The Heavy Chain possesses twelve repetitive domains that are Gly rich (ca. 45 %) forming the crystalline regions, separated by short linker domains (42 to 43 residues). The short linker domains are non-repetitive and form amorphous regions.¹⁸ Whereas, the repetitive domain is predominantly formed of Gly-X repeats (ca. 94 % of the repetitive domain), where X is Ala (64 %), Ser (22 %), Tyr (10 %), Val (3 %), threonine (Thr) (1.3 %).^{15, 18} The structural features of *B. mori* SF have been conveniently studied using the synthetic peptide, (Ala-Gly)_n, as a model for the crystalline regions,^{13, 19} because the lack of Ser in the model peptide (Ala-Gly)₁₅ does not affect the ¹³C cross-polarisation magic angle spinning (CP-MAS) NMR chemical shifts of the Ala and Gly residues in the repeated sequence (Ala-Gly-Ser-Gly-Ala-Gly)_n of native SF.^{12, 20, 21} From X-ray and electron diffraction studies *B. mori* SF the periodic copolypeptide (Ala-Gly)_n, has been shown to have an orthorhombic crystal structure with unit cell dimensions, $a = 4.65 \text{ \AA}$, $b = 14.24 \text{ \AA}$, $c = 8.88 \text{ \AA}$.^{12, 13, 22, 23} Within the simplified (Ala-Gly)_n model the repeat β -turn type II structure is stabilised by intramolecular hydrogen bond interactions. The overall planar sheets are held together by intermolecular hydrogen bonding interactions, involving the central amide-bond of the β -turn, perpendicular to intramolecular interactions. Although, such an (Ala-Gly)_n structural model vastly simplifies the overall structure of *B. mori* silkworm's SF, it makes it convenient and less computationally demanding.

Prior to spinning the *B. mori* SF is stored in the middle silk gland (30 % in water) and undergoes a conformational change when under tensile and shear stress and loss of water.¹ The rate of degradation of SF is directly related to the content of secondary β -sheet crystalline structure present within the bulk material.²⁴ The β -sheet content from regenerated SF can be modified through the use of various processing methods (e.g. water content and drying methods).²⁵

Previously, MD simulations has been used to investigate the mechanical behaviour of *B. mori* SF¹⁸ and the conformational change of its silk I form into silk II.¹ The transformation of silk I into silk II is brought on by either external forces and in an aqueous environment (the silk gland and spinneret).¹ To simulate this structural change, a (Ala-Gly)_n model (at 298 K) was stretched (application of both shear (ca. 0.5 GPa) and tensile (ca. 0.1 GPa) stress) and the torsion angles of the residues evaluated. The resulting secondary structures showed a good agreement with existing solid-state NMR information indicating the potential of atomistic simulation techniques. The computationally produced silk II structure possessed ca. 75 % β -sheet and ca. 25 % β -turn content, comparable with experimental values of 73 % β -sheet and 27 % β -turn content.¹

Despite extensive investigation of *B. mori* SF structure, discussed above, the fingerprint structural parameters for silk I and silk II remain mostly unexamined. The primary structure of *B. mori* SF has been determined to be predominately (Ala-Gly)_n,²⁷ although, the SF can exist as either in silk I or silk II form and their structural confirmations are less clear. It is generally accepted^{27, 28, 29} that the basic features of silk II are the orderly packing of antiparallel β -sheets, however, other studies suggest that silk II form consists of randomly stacked antiparallel β -sheets.^{30, 31} As for silk I, the structural parameters remain unclear because this less stable conformation is susceptible to transformation into the silk II conformation, leading to difficulty in performing X-ray diffraction experiments. As a result, multiple models exist for the silk I form (e.g. crankshaft model with Ala and Gly residues close to β -sheet and α -helix confirmations²¹, loose 4-fold helical confirmation³² and four-residue β -turn structure³³). The possible structural models of *B. mori* SF (in silk I and silk II form) have been examined using DFT to determine the NMR chemical shifts. The DFT approach incorporated a similar (Ala-Gly)_n model mentioned previously, then calculated the ¹³C chemical shielding tensors using the theory-gauge independent atomic orbital (DFT-GIAO) with the BLYP exchange-correlation functional.³⁴ The results obtained indicated that the silk I structure did not entirely agree with that characterised by ¹³C NMR experiments. Instead, a ₃₁₀-helix like conformation with torsion angle ranges of $\langle\phi\rangle = -59 \pm 2^\circ$, $\langle\psi\rangle = 119 \pm 2^\circ$ for the Ala residue and $\langle\phi\rangle = -78 \pm 2^\circ$, $\langle\psi\rangle = 149 \pm 2^\circ$ for the Gly residue was suggested. However, the silk II structure agreed well with that characterised by ¹³C NMR experiments and previous descriptions of SF in silk II

form (i.e. the orderly packing of antiparallel β -sheets). The torsion angle ranges are $\langle\phi\rangle = -143 \pm 6^\circ$, $\langle\psi\rangle = 142 \pm 5^\circ$ for both Ala and Gly residues.³⁴

Therefore, in this study, non-hydrated and hydrated SF crystal structure models were studied using both DFT and classical MD to understand how water is accommodated in the silk structures, what impact it has on the silk itself and how it moves within between the protein chains.

2. Methodology

Utilising both DFT and classical MD to study the structure of SF provides a description across a range of time and length scales. While DFT simulations can provide a detailed description of the electronic structure of the silk, the number of atoms accessible is insufficient to accurately assess the secondary structure of the protein chains. By contrast, classical MDs neglect of electronic information enables the simulation of thousands of atoms enabling analysis of the secondary structure and the examination of bulk water transport around the protein chains.

2.1 Preparation of the (Ala-Gly)_n Silk Fibroin unit cell

Construction of the (Ala-Gly)_n Silk Fibroin unit cell followed the methodology described by Yamane *et al*.¹. The initial unit cell was created by arranging four Ala-Gly chains with repeated β -turns according to information from experimental and computational experiments on SF.^{1, 12, 15, 16, 18, 22, 23} To simulate the bulk system of the repeated polymer chains, a periodic boundary condition was implemented, where nitrogen and carbon-terminals were connected to mirror images of themselves and the resulting structure is illustrated in **Figure 1**.

Hydrated SF supercells were created by introducing ca. 7.5 wt % of water molecules, mimicking experimentally reproduced SF films^{35, 36}. Water molecules were placed randomly within the supercells, while ensuring that no water molecules were placed within 1.7 Å of the silk.^{37, 38}

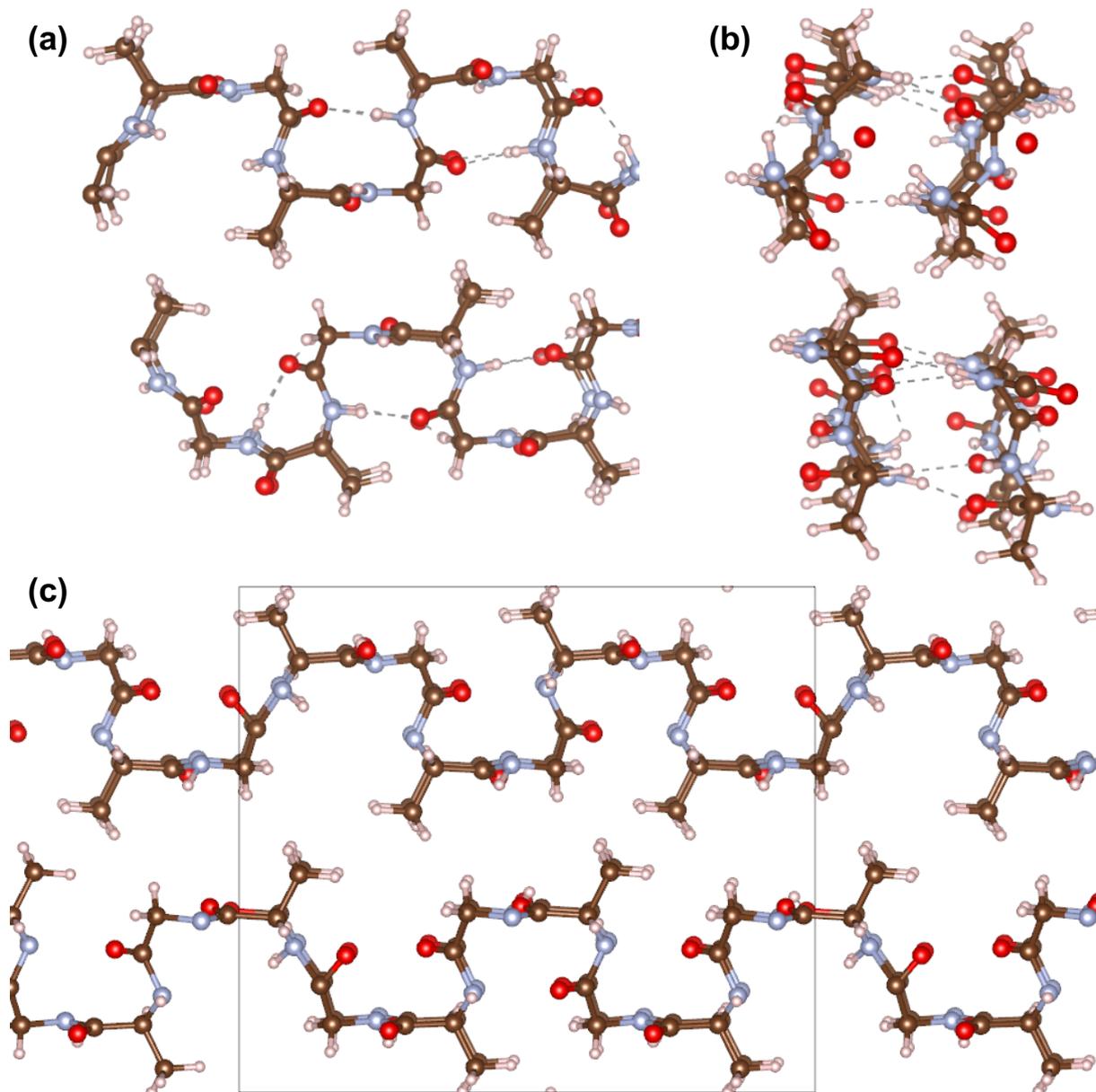


Figure 1. Visualisation of the crystal structure of *B. mori* SF in a silk I form. (a) Is a snapshot of the crystal structure (repeated β -turn type II conformation) from along the z-axis, (b) from along the x-axis and (c) is another snapshot for a unit cell of the periodic *B. mori*. SF crystal structure from along the z-axis. A visual representation of the repeated β -turn type II conformation where the brown ball and sticks depict carbon atoms, light grey the hydrogen atoms, red the oxygen atoms and pale blue the nitrogen atoms. The lattice parameters of the unit cell shown in **Figure 1** are orthorhombic: $a = 17.8 \text{ \AA}$, $b = 15.7558 \text{ \AA}$, $c = 11.4904 \text{ \AA}$. The grey dashed lines represent the hydrogen bond interactions. Figure created using **VESTA**¹⁴.

2.2 Classical molecular dynamic simulations

Classical molecular dynamics simulations utilised a range of supercells, ranging from 2×2×2 to 4×4×4 repetitions of the unit cell containing 2176 and 17,408 atoms (2377 and 19,022 atoms when hydrated), respectively. In each case, the DL_FIELD package using the force field scheme OPLS2005/OPLS-AA (eq. 1, 2, 3 and 4), was used to create the input files for the DL_POLY_4 simulation package.^{39, 40, 41, 42} In addition, the periodic (Ala-Gly)_n crystals were visualised using the Visual Molecular Dynamics (VMD) programme.⁴³

Interactions between the ions in the (Ala-Gly)_n chains were represented using the OPLS-AA force field where the total energy (E_{tot}) of a molecular system is evaluated as a sum of the following components, the non-bonded energy (E_{nb}), bond stretching and angle bending terms (E_{bond} and E_{angle} , respectively), and the torsional energy (E_{torsion}). The non-bonded term is computed as a sum of the Coulomb and Lennard-Jones contributions for pairwise intra and intermolecular interactions.^{35, 44}

OPLS-AA force field where the total energy (E_{tot}) of a molecular system is evaluated as a sum of the following components, the non-bonded energy (E_{nb}), bond stretching and angle bending terms (E_{bond} and E_{angle} , respectively), and the torsional energy (E_{torsion}). The non-bonded term is computed as a sum of the Coulomb and Lennard-Jones contributions for pairwise intra and intermolecular interactions.^{35, 44}

$$E_{\text{nb}} = \sum_{i < j} \left[\frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^6}{r_{ij}^6} \right) \right] f_{ij} \quad (1)$$

where, q_i and q_j are the charges on ions separated by r_{ij} , σ_{ij} and ϵ_{ij} are the Lennard-Jones coefficients. Geometric combining rules for the Lennard-Jones coefficients were employed: $\sigma_{ij} = (\sigma_{ii} \sigma_{jj})^{1/2}$ and $\epsilon_{ij} = (\epsilon_{ii} \epsilon_{jj})^{1/2}$. The summation runs over all of the pairs of atoms $i < j$ on molecules A and B or A and A for the intramolecular interactions. Moreover, in the latter case, the coefficient f_{ij} is equal to 0.0 for any $i - j$ pairs connected by a valence bond (1-2 pairs) or a valence bond angle (1-3 pairs).

$f_{ij} = 0.5$ for 1,4 interactions (atoms separated by exactly three bonds) and $f_{ij} = 1.0$ for all of the other cases. The E_{bond} and E_{angle} are reported in eq. 2 and 3.^{35, 44}

$$E_{bond} = \sum_{bonds} K_r (r - r_{eq})^2 \quad (2)$$

$$E_{angle} = \sum_{angles} K_\theta (\theta - \theta_{eq})^2 \quad (3)$$

where, the subscript eq is used to depict the equilibrium values of bond length r and the angle θ . Finally, the torsional contribution is given in eq. 4.³⁵

$$E_{torsion} = \sum_i \frac{V_1^i}{2} [1 + \cos(\phi_i)] + \frac{V_2^i}{2} [1 - \cos(2\phi_i)] + \frac{V_3^i}{2} [1 + \cos(3\phi_i)] \quad (4)$$

where, ϕ is the dihedral angle and V_1 , V_2 and V_3 are the Fourier coefficients to be optimised.

To represent interactions between water molecules and the silk chains the TIP3P_O force field was employed. The TIP3P_O model uses 12-6-1 (Coulomb plus Lennard-Jones) and 9-6-1 functions for water which yield results of similar quality for the liquid. The three-site model possesses three interaction points corresponding to the three atoms of the water molecule; each site has a point charge, q for electrons and an A and C Lennard-Jones parameters.⁴⁴

$$\Delta E(12 - 6 - 1) = \sum_a^{in A} \sum_b^{in B} \left(\frac{q_a q_b e^2}{r_{ab}} + \frac{A_a A_b}{r_{ab}^{12}} - C_a C_b / r_{ab}^6 \right) \quad (5)$$

The TIP3P_O model utilises Lennard-Jones parameters on the hydrogen atoms, in addition to the oxygen atom.⁴²

In order to examine the mobility of water around the SF the mean squared displacement (MSD) of the oxygen ions in the water molecule were determined. The MSD at time t is defined as an ensemble average, (eq. 6).

$$MSD \equiv \langle |x(t) - x_0|^2 \rangle = \frac{1}{N} \sum_{i=1}^N |x^{(i)}(t) - x^{(i)}(0)|^2 \quad (6)$$

where, N is the number of particles to be averaged, vector $x^{(i)}(0) = x_0^{(i)}$ is the reference position of the i -th particle and vector $x^{(i)}(t)$ is the position of the i -th particle at time t .⁴⁵

Simulation supercells were initially energy minimised under constant volume conditions before being equilibrated at a range of temperatures (10, 150, 273, 298, 310, 373 and 473 K) under the canonical ensemble (NVT) and then isothermal-isobaric ensemble (NPT). For the equilibrations, the Nose-Hoover thermostat and barostats were used with relaxation times of 0.05 ps and 0.1 ps, respectively. Equilibrations last 100,000 timesteps with each timestep corresponding to 0.0005 ps. During each equilibration the atomic configurations were recorded to enable examination of the secondary structure. Dihedral angles were then calculated using the DL_ANALYSER package enabling the creation of Ramachandran plots.^{46, 47, 48}

2.3 Density functional theory simulations

DFT simulations⁴⁹ were carried out using the Quickstep method in CP2K,⁵⁰ with the BLYP exchange correlation functional^{51, 52} A double zeta valence polarised (DZVP) basis set was employed for all calculations.⁵³ The DZVP has been shown to perform well with energy and geometry calculations, providing accurate results while also being computationally efficient.⁵⁴

Due to the greater computational requirements the DFT simulations were performed on supercells smaller than 2x2x2. Hydrated and non-hydrated silk supercells were minimised until the forces on the atoms were below 10⁻³ eV/Å. Simulation supercells were then equilibrated using ab initio molecular dynamics (AIMD) under NPT conditions at the same temperatures studied for using classical MD. AIMD simulations used the thermo and barostats with relaxation times of 60 fs and ran for 3000 timesteps of 0.05 fs.

Lastly, the trajectory of 8 water molecules (ca. 7.5 % water content) through the hydrated periodic (Ala-Gly)₁₆ crystal were also obtained utilising CP2K.⁵⁰ This was possible by running MD simulations using the information obtained from the DFT optimised hydrated periodic (Ala-Gly)₁₆ crystal, however, the NVT ensemble was set for 310 K and 42350 steps while each timestep was 0.05 fs, as a result, the experiment lasted 2117.5 fs.⁴³

3. Results and discussion

3.1 The (Ala-Gly)_n Silk Fibroin crystal models' secondary structure

Figure 2 and **3** show the Ramachandran contour plots of the non-hydrated and hydrated (Ala-Gly)₁₂₈, (Ala-Gly)₁₀₂₄ and (Ala-Gly)₁₆ SF crystal models. Classical MD and DFT simulations were conducted to obtain the torsion angles for residues Ala and Gly, therefore for clarity; only the 298 K NPT ensemble experiments are shown. The torsion angles for (Ala-Gly)₁₂₈ and (Ala-Gly)₁₀₂₄ utilised classical MD simulations, whereas, (Ala-Gly)₁₆ utilised DFT simulations.

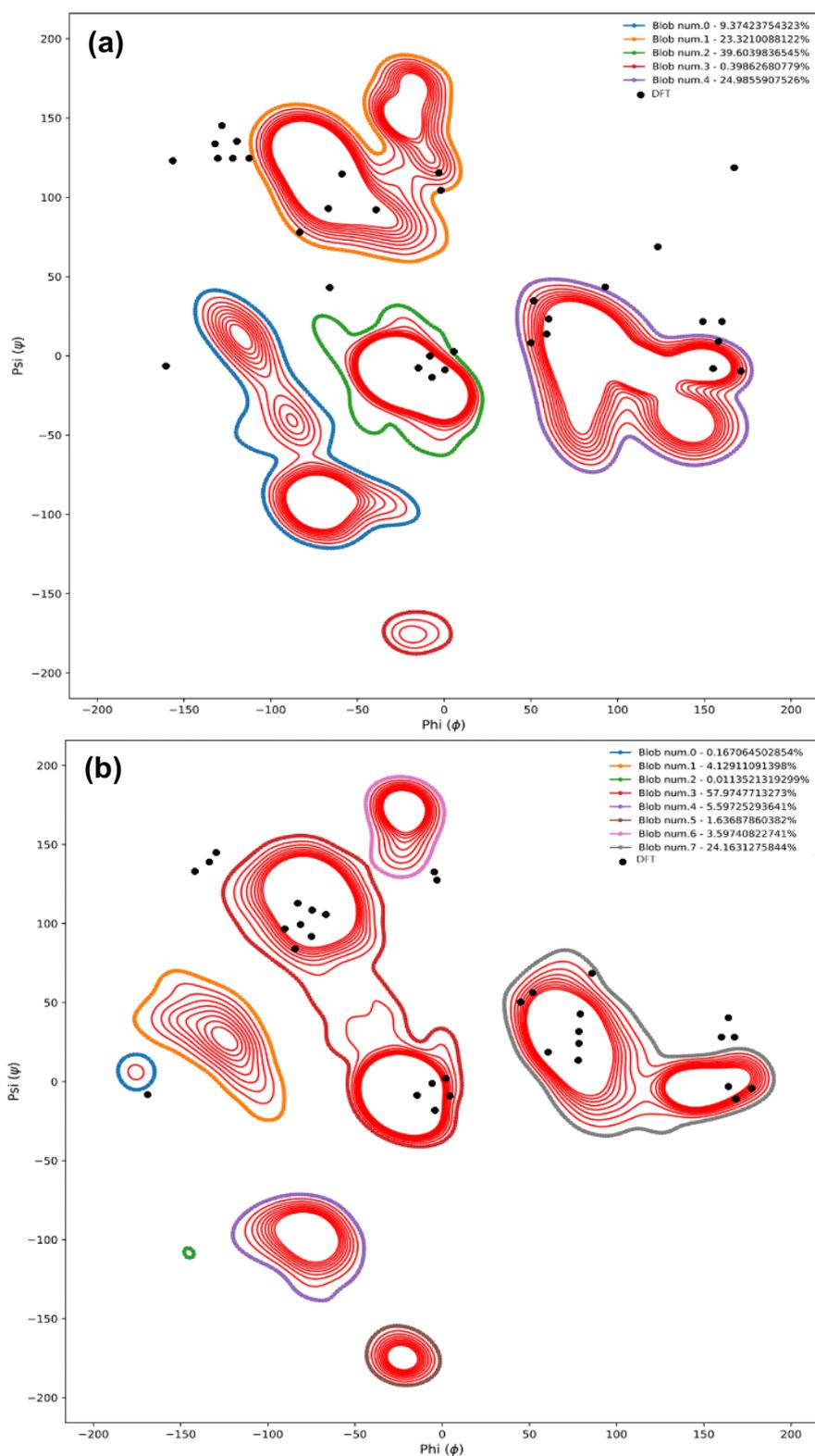


Figure 2. Ramachandran contour plots of the (Ala-Gly)₁₂₈ and (Ala-Gly)₁₆ SF crystal model. (a) Is in the non-hydrated state and **(b)** is in the hydrated state. The legend depicts the percentage of the residues within each region and position of the CP2K (DFT) generated torsion angles of Ala and Gly residues from (Ala-Gly)₁₆.

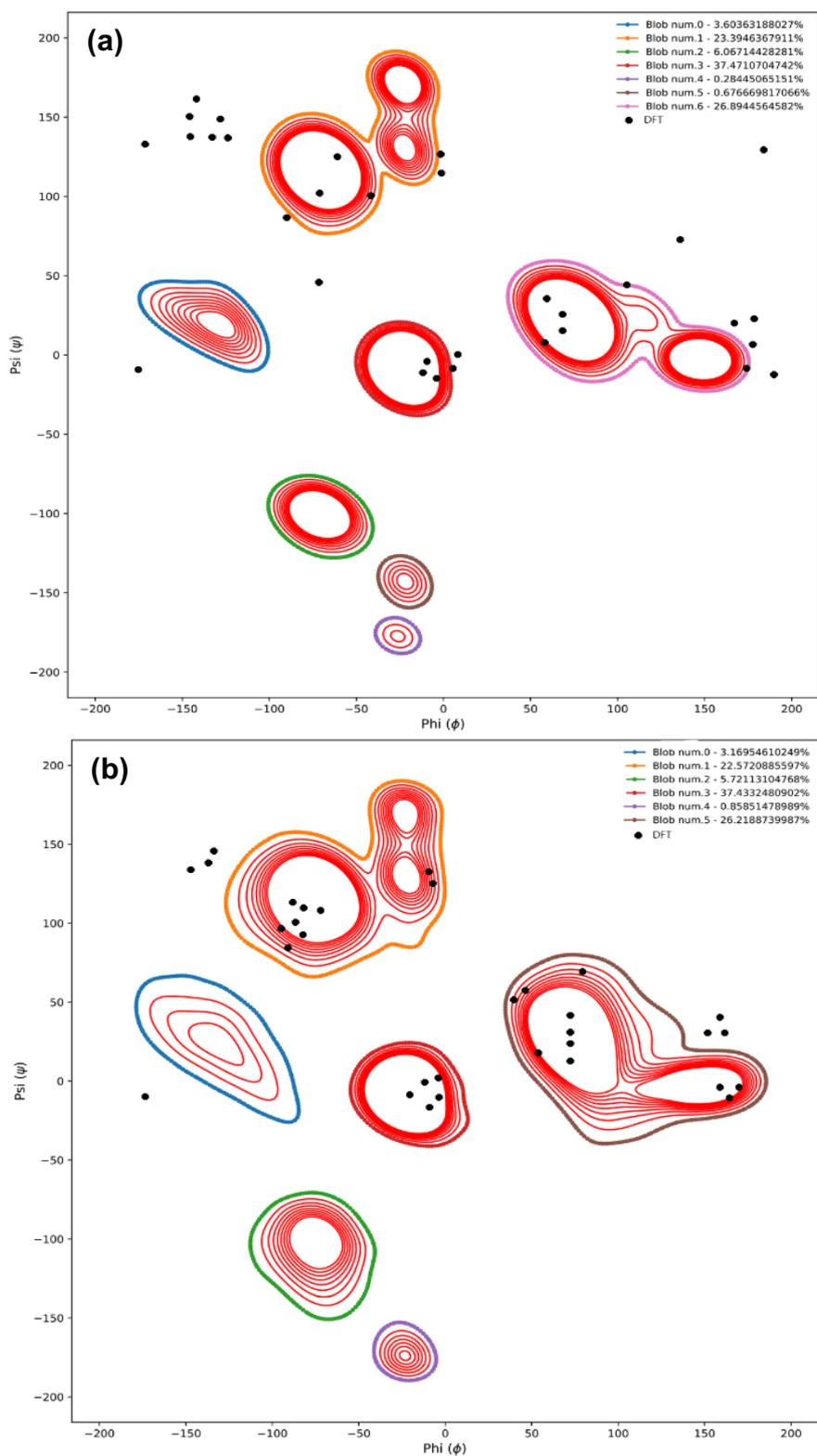


Figure 3. Ramachandran contour plots of the (Ala-Gly)₁₀₂₄ and (Ala-Gly)₁₆ SF crystal model. (a) Is in the non-hydrated state and **(b)** is in the hydrated state. The legend depicts the percentage of the residues within each region and position of the CP2K (DFT) generated torsion angles of Ala and Gly residues from (Ala-Gly)₁₆.

The Ramachandran plots in **Figures 2** and **3**, possess regions that lie within $\langle\phi\rangle = -60^\circ$ and $\langle\psi\rangle = 130^\circ$ and $\langle\phi\rangle = 70^\circ$ and $\langle\psi\rangle = 10^\circ$, which is characteristic of the Ala and Gly residues, respectively, as reported in the literature.^{1, 59, 61} Hereby, implying the SF model utilised in this work, possesses qualities that have been experimentally observed. The (Ala-Gly)_n SF crystal structure is to be considered in the silk I form (i.e. repeated β -turn type II conformation), because β -sheets are not the predominant secondary structure, instead, 3_{10} -helix is the predominant secondary structure (ca. 37%), agreeing with the literature.³⁴ Furthermore, **Figures 2** and **3** indicate that the SF crystal models possesses a heterogeneous structure, evidenced by a left-handed α -helix, 3_{10} -helix, β -sheet (ca. $\langle\phi\rangle = 70^\circ$ and $\langle\psi\rangle = 10^\circ$, ca. $\langle\phi\rangle = -40^\circ$ and $\langle\psi\rangle = -30^\circ$, ca. $\langle\phi\rangle = -60^\circ$ and $\langle\psi\rangle = 130^\circ$, respectively) and random coil structures, which are entirely consistent with the literature.^{1, 9, 12, 13, 15, 16, 22, 23, 34, 61,}

On the other hand, by comparing the non-hydrated states of (Ala-Gly)₁₂₈ (**Figure 2**) with the hydrated states, the torsion angle regions for the Ala and Gly residues appear in slightly different locations. This is likely due to the flexibility of the SF model's polymer backbone (Ala and Gly) chains. It could be suggested that due to greater temperatures, the polymer backbone chains can move more freely, as the difference in the Ala and Gly residue's positions were exacerbated for higher temperatures. With the introduction of more residues and/or water molecules, more hydrogen bond interactions are introduced (e.g. between polymer backbone chains), thus, limiting the potential flexibility of the (Ala-Gly)_n SF crystal model. The difference in the Ala and Gly residue positions for the non-hydrated and hydrated states of the larger (Ala-Gly)₁₀₂₄ SF crystal model (**Figure 3** and **5**) were not observed. This could mean that the (Ala-Gly)₁₀₂₄ SF crystal model possesses a greater initial number of hydrogen bond interactions (intra- and intermolecular), therefore, inherently less flexible than the (Ala-Gly)₁₂₈ SF crystal model.

The torsion angles determined using DFT, have also been included in **Figure 2 (b)** and **Figure 3 (b)**. This data further supports the assertion that the SF crystal model possesses a heterogeneous structure. Furthermore, the DFT technique appears to be more accurate at determining the residues located within the β -sheet region. For instance, a previous SF structure study using DFT chemical shift calculation (mentioned previously),³⁴ reported that the torsion angle ranges for Ala and Gly are $\langle\phi\rangle = -143 \pm 6^\circ$, $\langle\psi\rangle = 142 \pm 5^\circ$. In addition, the torsion angle range of $\langle\phi\rangle = -143$

$\pm 6^\circ$, $\langle \psi \rangle = 142 \pm 5^\circ$ is within the characteristic range for antiparallel β -sheets.^{1, 34} In this work (**Figure 2 and 3**), the (Ala-Gly)₁₂₈, (Ala-Gly)₁₀₂₄ and (Ala-Gly)₁₆ have achieved similar torsion angle values for Ala and Gly residues.

Figure 4. A VESTA¹⁴ visualisation of most (a) and least (b) favourable water molecule positions in the (Ala-Gly)₁₆ SF crystal model. The brown ball and sticks depict the carbon atoms, light grey the hydrogen atoms, red the oxygen atoms and pale blue the nitrogen atoms. The dashed lines depict the hydrogen bond interactions. In **(a)**, the central water molecule (W₆) is represented and in **(b)**, the central water molecule (W₅) is represented.

In order to understand how the water molecules move around the silk protein chains we calculate the MSD in the hydrated (Ala-Gly)₁₀₂₄ supercell using classical molecular dynamics. The MSDs as a function of time are presented in figure 6. Figure 6 indicates that at the water is mobile even at room temperature. From the gradients of the lines presented in figure 6 it is possible to create an Arrhenius plot for water diffusing around the silk as presented in figure 7.

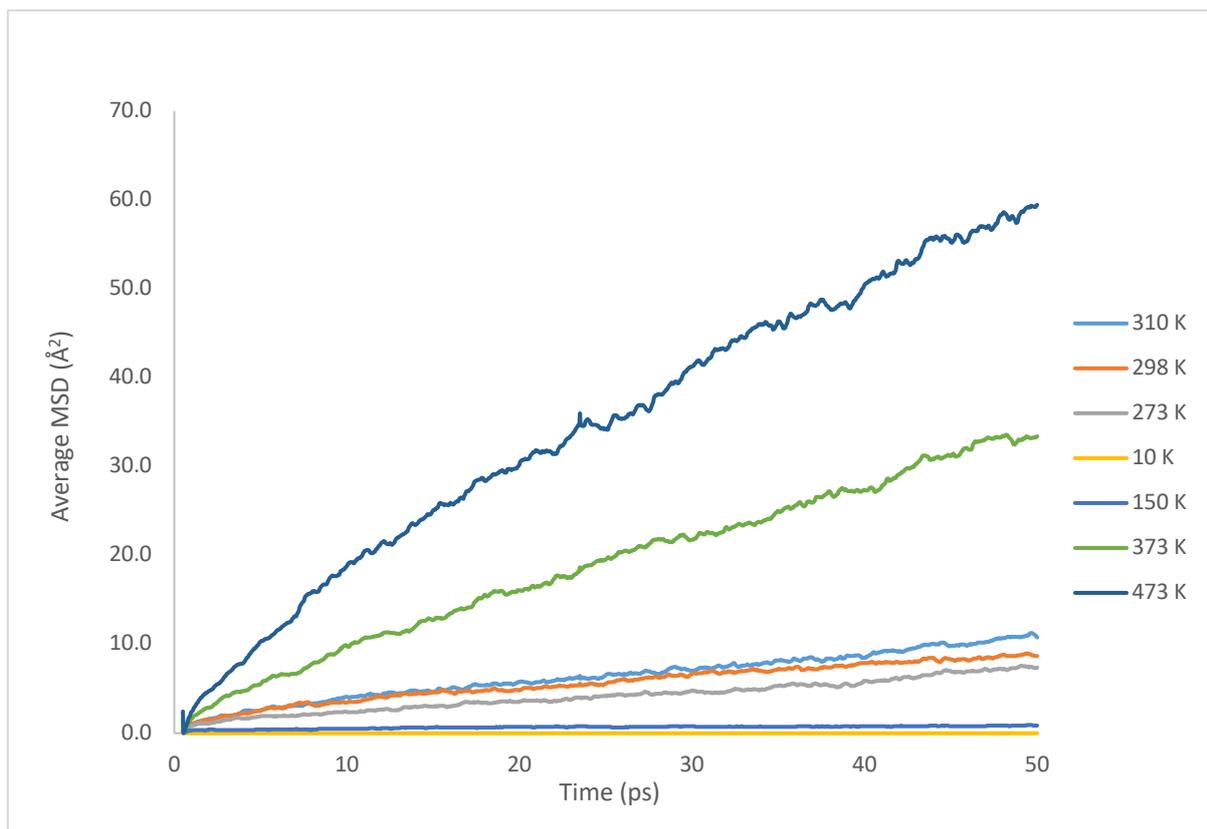


Figure 6. MSD of the water molecules in the (Ala-Gly)₁₀₂₄ SF crystal. The water displacement (in Å²) using the SF crystal model (Ala-Gly)₁₀₂₄, the legend depicts the temperature (in K) of the experiment's system.

The data presented in figure 7 can be compared to a previous experimental study examining the use of SF as an edible coating for perishable food preservation. This study obtained a diffusion coefficient (at room temperature) of $1.05 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ at 58 % β -sheet content, $3.21 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ at 48 % β -sheet content and $5.79 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ at 36 % β -sheet content.⁶¹ By comparison the value predicted here is $2.918 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$. The good agreement between the diffusivities further supports the efficacy of our computational model. However, it is noted that the silk structure used contained between 48 and 58 % as opposed to the ca. 26 % present in the simulations.

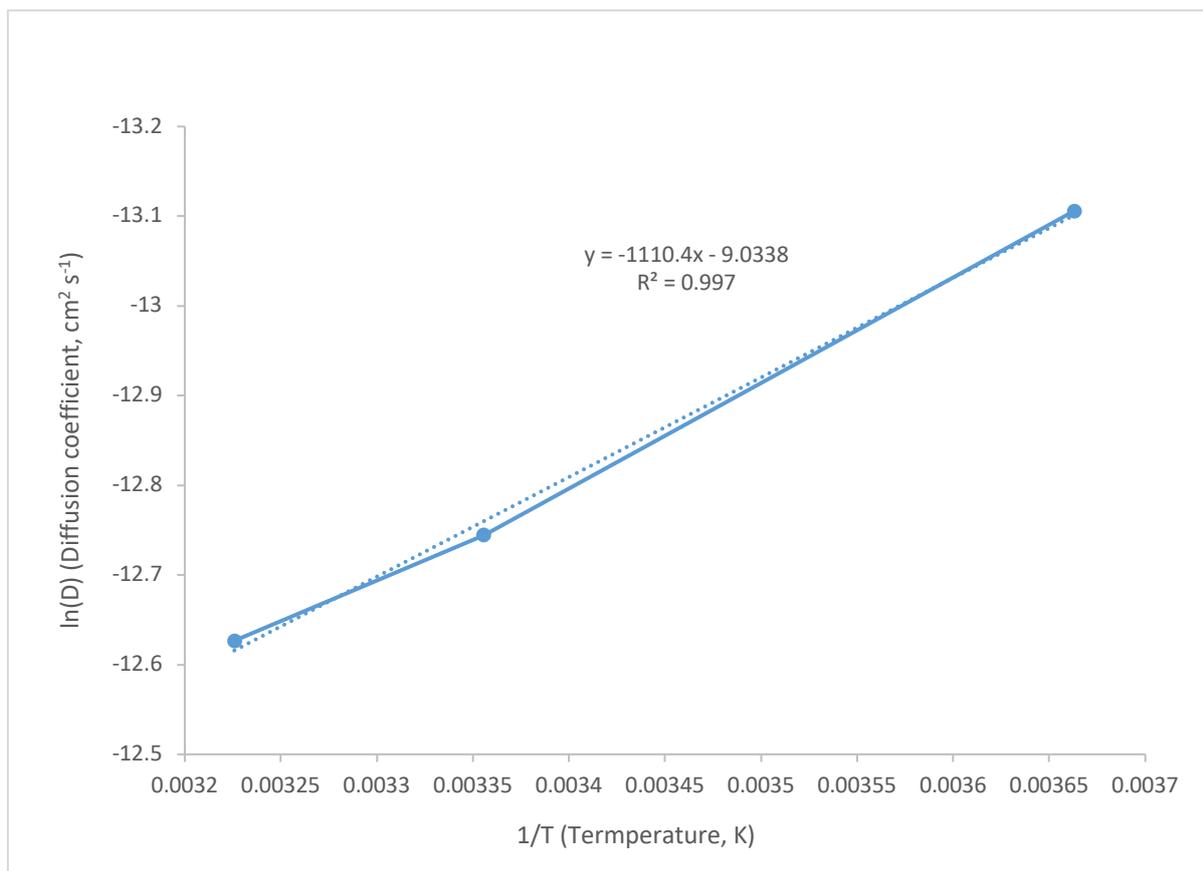


Figure 8. Arrhenius plot of the hydrated (Ala-Gly)₁₀₂₄ SF crystal. By using the gradient of the slope in the above Arrhenius plot, the calculated activation energy is determined as 9.232 kJ/mol.

The molecular dynamics simulations enable the determination of the oxygen pathways through silk crystal providing insight into how exactly the water moves from site to site.

Lastly, the trajectory for the displacement of water molecules in the hydrated (Ala-Gly)₁₆ SF crystal model is visualised in **Figure 9** and **10**.

Furthermore, the displacement of 8 water molecules (ca. 7.5 wt % water content) in the hydrated (Ala-Gly)₁₆ SF crystal model is reported in **Table 3**.

Water molecule	Displacement of water molecules in (Ala-Gly)₁₆ SF crystal model (Angstroms, Å)	Equivalent position of the water molecule before the experiment (Refer to Table 2)	Equivalent position of the water molecule after the experiment (Refer to Table 2)
1	0.559	W ₄	W ₂
2	1.394	W ₉	W ₁
3	1.314	Periodic boundary	W ₈
4	0.062	W ₄	W ₈
5	0.540	W ₅	W ₅
6	0.718	W ₁₀	W ₁₀
7	0.907	W ₃	W ₃
8	2.295	W ₇	W ₇

Table 3. The displacement of 8 water molecules (7.5 wt % water content) in the (Ala-Gly)₁₆ SF crystal model. The displacement of the 8 water molecules incorporated in the (Ala-Gly)₁₆ SF crystal model was recorded at the final timeframe of 2117.5 fs.

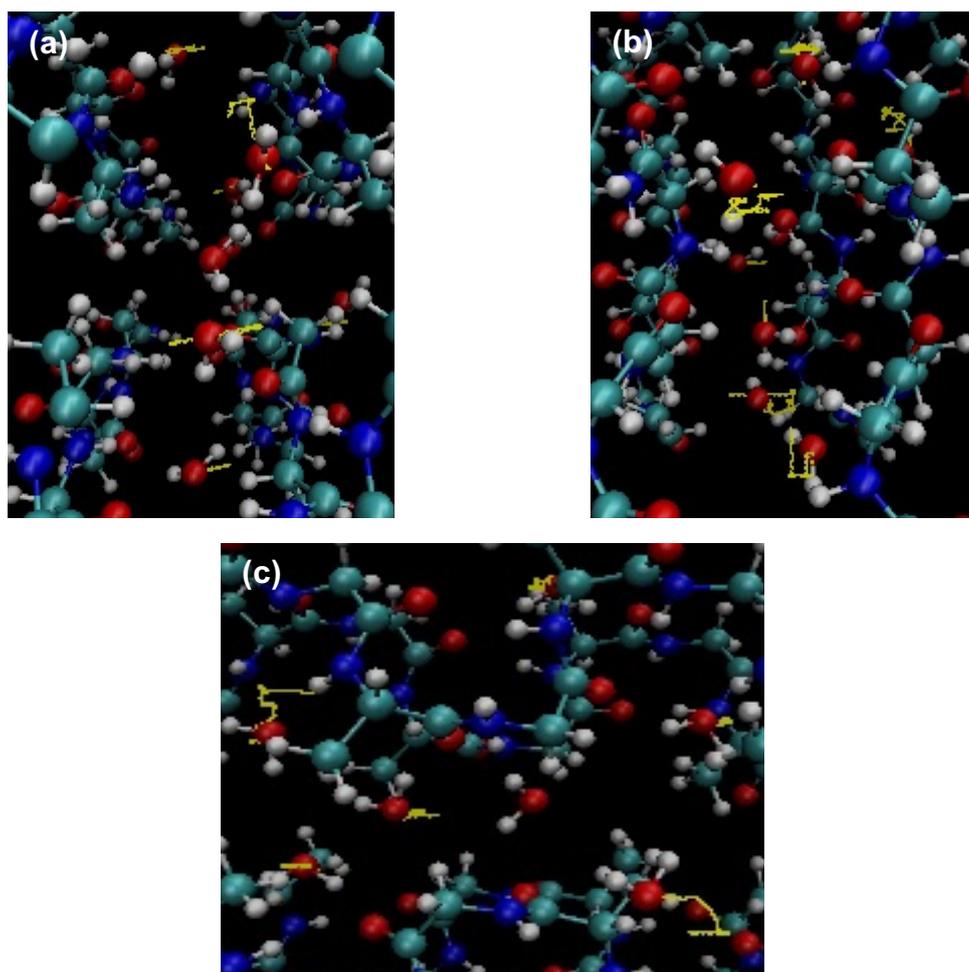


Figure 10. A VMD⁴³ visualisation of the trajectory (yellow lines) of the 8 water molecules in the (Ala-Gly)₁₆ SF crystal structure. (a) Is viewing down the x-axis, (b) is viewing down the y-axis and (c) is viewing down the z-axis. The turquoise ball and sticks depict the carbon atoms, light grey the hydrogen atoms, red the oxygen atoms and blue the nitrogen atoms.

4. Conclusion

A unique insight of the (Ala-Gly)₁₆, (Ala-Gly)₁₂₈ and (Ala-Gly)₁₀₂₄ SF crystal models were achieved by using a combination of MD and DFT computational techniques. The secondary structure was evaluated using the information available in the literature and validated by possessing the characteristic torsion angles for residues Ala and Gly. The DFT simulations provided a more accurate insight for the β -sheet region residues, while the MD simulations provided the contour plots with percentage residues detailing the predominant secondary structures. In addition, the hydrated with water (Ala-Gly)_n SF crystals were investigated, and the displacement of water, diffusion coefficient, activation energy, energy of water positions and trajectory were reported. As a result, an appreciation for combining different techniques to investigate materials is obtained. With regards to future work, many avenues remain open (e.g. mechanical evaluation of the (Ala-Gly)_n SF crystal model) or similar biopolymer, and incorporating new species (e.g. charged ions or ionic liquids) into the crystal models. With continued investigation into materials, like SF, a greater understanding of their properties and key aspects/interactions can be achieved.

References

- 1) T. Yamane, K. Umemura, Y. Nakazawa and T. Asakura, "Molecular Dynamics Simulation of Conformational Change of Poly(Ala-Gly) from Silk I to Silk II in Relation to Fiber Formation Mechanism of Bombyx mori Silk Fibroin", *Macromolecules*, **2003**, 36, 6766-6772.
- 2) J. P. O'Brien, R. F. Stephen, Y. Termonia and K. H. Gardner, *Adv. Mater.* **1998**, 10, 1185-1195.
- 3) G. H. Altman, F. Diaz, C. Jakuba, T. Calabro, R. L. Horan, J. Chen, H. Lu, J. Richmond and D. L. Kaplan, *Biomaterials*, **2003**, 24, 401-416.
- 4) G. Gronau, S. T. Krishnaji, M. E. Kinahan, T. Giesa, J. Y. Wong, D. L. Kaplan and M. J. Buehler, *Biomaterials*, **2012**, 33, 8240-8255.
- 5) D. Porter and F. Vollrath, *Adv. Mater.* **2009**, 487-492.
- 6) J. Yan, G. Zhou, D. P. Knight, Z. Shao and X. Chen, *Biomacromolecules*, **2010**, 11,1-5.
- 7) Y. Jin, Y. Hang, J. Luo, Y. Zhang, H. Shao and X. Hu, *Int. J. Biol. Macromol.* **2013**, 62, 162-166.
- 8) B. Mortimer, J. Guan, C. Holland, D. Porter and F. Vollrath, *Acta Biomater.* **2015**, 11, 247-255.
- 9) T. Asakura, K. Okushita and M. P. Williamson, "Analysis of the Structure of Bombyx mori Silk Fibroin by NMR", *Macromolecules*, **2015**, 48, 2345-2357.
- 10) X. Jia, C. Wang, V. Ranganathan, B. Napoer, C. Yu, Y. Chao, M. Forsyth, F. G. Omenetto, D. R. MacFarlane and G. G. Wallace, "A Biodegradable Thin-Film Magnesium Primary Battery Using Silk Fibroin-Ionic Liquid Polymer Electrolyte", *ACS Energy Letters*, **2017**, 2, 831-836.
- 11) S. Aznar-Cervantes, M. I. Roca, J. G. Martinez, L. Meseguer-Olmo, J. L. Cenis, J. M. Moraleda and T. F. Otero, "Fabrication of conductive electrospun silk fibroin scaffolds by coating with polypyrrole for biomedical applications", *Bioelectrochemistry*, **2012**, 85, 36-43.
- 12) T. Asakura, Y. Suzuki, Y. Nakazawa, G. P. Holland and J. L. Yarger, "Elucidating silk structure using solid-state NMR", *Soft Matter*, **2013**, 9, 11440-11450.
- 13) B. Lotz and F. C. Cesari, *Biochimie*, **1979**, 61, 205-214.

- 14) K. Momma and F. Izumi, "VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data," *J. Appl. Crystallogr.* **2011**, 44, 1272-1276.
- 15) C. Z. Zhou, F. Confalonieri, M. Jacquet, R. Perasso, Z. G. Li and J. Janin, "Silk fibroin: structural implications of a remarkable amino acid sequence", *Proteins Struct. Funct. Genet.* **2001**, 44, 119–122.
- 16) T. Asakura, K. Okushita and M. P. Williamson, "Analysis of the structure of Bombyx mori silk fibroin by NMR", *Macromolecules*, **2015**, 48, 2345–2357.
- 17) S. Inoue, K. Tanaka, F. Arisaka, S. Kimura, K. Ohtomo and S. Mizuno, "Silk fibroin of Bombyx mori is secreted , assembling a high molecular mass elementary unit consisting of H-chain , L-chain , and P25 , with a 6: 6: 1 molar ratio", *J. Biol. Chem.* **2000**, 275, 40517–40528.
- 18) M. Patel, D. K. Dubey and S. P. Singh, "Phenomenological models of Bombyx mori silk fibroin and their mechanical behavior using molecular dynamics simulations", *Mater. Sci. and Eng.* **2020**, 108, 2-20.
- 19) K. Okuyama, R. Somashekar, K. Noguchi and S. Ichimura, *Biopolymers*, **2001**, 59, 310-319.
- 20) J. M. Yao, K. Ohgo, R. Sugino, R. Kishore and T. Asakura, *Biomacromolecules*, **2004**, 5, 1763-1769.
- 21) T. Asakura, K. Ohgo, T. Ishida, P. Taddei, P. Monti and R. Kishore, *Biomacromolecules*, **2005**, 6, 468-474.
- 22) B. Lotz and H. D. Keith, "Crystal structure of poly(I-Ala-Gly)II: A model for silk I", *J. of Molecular Bio.* **1971**, 61, 1, 201-202.
- 23) J. O. Warwicker, "The Crystal Structure of Silk Fibroin", *Acta. Cryst.* **1954**, 7, 565-573.
- 24) B. D. Lawrence, S. Wharram, J. A. Kluge, G. G. Leisk, F. G. Omenetto, M. I. Rosenblatt and D. L. Kaplan, "Effect of Hydration on Silk Film Material Properties", *Macromol. Biosci.* **2010**, 10, 4, 393-403.
- 25) X. Hu, D. Kaplan and P. Cebe, "Effect of water on the thermal properties of silk fibroin", *Thermochimica Acta*, **2007**, 461, 137-144.
- 26) K. Mita, S. Ichimura and M. Zama, *J. Mol. Biol.* **1988**, 203, 917-925.
- 27) M. Demura, M. Minami, T. Asakura and T. A. Cross, *J. Am. Chem. Soc.* **1998**, 120, 1300-1308.

- 28) R. E. Marsh, R. B. Corey and L. Pauling, *Biochim. Biophys. Acta.* **1955**, 16, 1-34.
- 29) R. D. B. Fraser, T. P. Macrae, F. H. C. Stewart and E. Suzuki, *J. Mol. Biol.* **1965**, 11, 706-712.
- 30) Y. Takahashi, M. Gehoh and K. J. Yuzuriha, *Polym. Sci., B: Polym. Phys.* **1991**, 29, 889-891.
- 31) Y. Takahashi and M. Gehoh, *Int. J. Biol. Macromol.* **1999**, 24, 127-138.
- 32) T. Konishi and M. Kurokawa, *Sen'i Gakkaishi*, **1968**, 24, 550-554.
- 33) N. D. Lazo and D. T. Downing, *Macromolecules*, **1999**, 32, 4700-4705.
- 34) P. Zhou, G. Li, Z. Shao, X. Pan and T. Yu, "Structure of Bombyx mori Silk Fibroin Based on the DFT Chemical Shift Calculation", *J. Phys. Chem. B*, **2001**, 105, 12469-12476.
- 35) Q. Lu, X. Hu, X. Wang, J. A. Kluge, S. Lu, P. Cebe and D. L. Kaplan, "Water-Insoluble Silk Films with Silk I Structure", *Acta. Biomater.* **2010**, 6, 1380-1387.
- 36) K. Yazawa, K. Ishida, H. Masunaga, T. Hikima and K. Numata, "Influence of Water Content on the β -Sheet Formation, Thermal Stability, Water Removal, and Mechanical Properties of Silk Materials", *Biomacromolecules*, **2016**, 17, 1057-1066.
- 37) W. L. Jorgensen, "Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water", *J. Am. Chem. Soc.* **1981**, 103, 335-340.
- 38) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations," *J. Comp. Chem.* **1983**, 4, 187-217.
- 39) G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, "Evaluation and Reparameterization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides", *J. Phys. Chem. B*, **2001**, 105, 6474-6487.
- 40) I. T. Todorov, W. Smith, K. Trachenko and M.T. Dove, *J. Mater. Chem.* **2006**, 16, 1911-1918.
- 41) C. W. Yong, *J. Chem. Inf. Model.* **2016**, 56, 1405-1409.
- 42) C.W. Yong and I.T. Todorov, *Molecules*, **2018**, 23, 36.

- 43)W. Humphrey, A. Dalke and K. Schulten, "VMD - Visual Molecular Dynamics", *J. Molec. Graphics*, **1996**, 14, 33-38.
- 44)W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids", *J. Am. Chem. Soc.* **1996**, 118, 11225-11236.
- 45)D. Frenkel and B. Smit, "Understanding molecular simulation: From algorithms to applications 2nd. Ed. *Academic Press*, **2002**, 97.
- 46)A. L. Morris, M. W. MacArthur, E. G. Hutchinson and J. M. Thornton, "Stereochemical Quality of Protein-Structure Coordinates", *Proteins*, **1992**, 12, 345-364.
- 47)G. N. Ramachandran and V. Sasisekharan, "Conformation of polypeptides and proteins", *Adv. Protein Chem.* **1968**, 23, 283-437.
- 48)T. E. Creighton, "Proteins: Structures and Molecular Properties, 2nd ed. **1996**, 172-176, 183. New York: W. H. Freeman & Company.
- 49)S. Grimmea, J. Antony, S. Ehrlich and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu", *J. Chem. Phys.* **2010**, 132, 154104.
- 50)T. D. Kuhne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schutt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borstnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. V. Vondele, M. Krack and J. Hutter, "Cp2k: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations", *J. Chem. Phys.* **2020**, 152, 1-49.
- 51)A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior". *Phys. Rev. A.* **1988**, 38, 6, 3098–3100.
- 52)C. Lee, W. Yang, R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density". *Phys. Rev. B.* **1988**, 37, 2, 785–789.
- 53)N. Godbout, D. R. Salahub, J. Andzelm and E. Wimmer. "Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I.

Boron through neon, optimization technique and validation”, *Can J. Chem.* **1992**, 70, 571-560.

- 54) J. V. Vondele and J. Hutter. “Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases”, *J. Chem. Phys.* **2007**, 127, 11, 114105.
- 55) A. D. Becke and E. R. J. Johnson, “A density-functional model of the dispersion interaction”, *Chem. Phys.* **2005**, 123, 154101.
- 56) E. R. Johnson and A. D. Becke, “A post-Hartree–Fock model of intermolecular Interactions”, *J. Chem. Phys.* **2005**, 123, 024101.
- 57) E. R. Johnson and A. D. Becke, “A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections”, *J. Chem. Phys.* **2006**, 124, 174104.
- 58) A. Lilienfeld and A. Tkatchenko, “Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids”, *J. Chem. Phys.* **2010**, 132, 23, 234109.
- 59) K. Okuyama, K. Takanashi, Y. Nakajima and Y. Hasegawa, “Analysis of Silk I structure by X-ray and electron diffraction methods”, *J. Sericultural Sci. Japan*, **1988**, 57, 1, 23-30.
- 60) T. Asakura, J. Yao, T. Yamane, K. Umemura and A. S. Ulrich, “Heterogeneous Structure of Silk Fibers from *Bombyx mori* Resolved by ¹³C Solid-State NMR Spectroscopy”, *J. American Chem. Soc.* **2002**, 124, 30, 8794-8795.
- 61) B. Marelli, M. A. Brenckle, D. L. Kaplan and F. G. Omenetto, “Silk Fibroin as Edible Coating for Perishable Food Preservation”, *Sci. Rep.* **2016**, 6, 25263.
- 62) S. A. Arrhenius, “Über die Dissociationswärme und den Einfluß der Temperatur auf den Dissociationsgrad der Elektrolyte”, *Z. Phys. Chem.* **1889**, 4, 96–116.