Exploring the Structural Properties of Natural and Synthetic Biological Molecules in Aqueous Solution

Natasha Hazel Rhys

School of Physics and Astronomy

The University of Leeds

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

May 2015

Declaration

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others. Chapter 3 is based on experiments that were devised by Dr Lorna Dougan, and were completed by Dr Lorna Dougan and Dr Alan Soper. Chapter 4 is based on experiments that were devised by Dr Lorna Dougan and the candidate, and were completed by the candidate, Dr James Towey, Dr Lorna Dougan and Dr Alan Soper. The molecule analysed in Chapter 4 was synthesised by Dr Tom McAllister and Dr Michael Webb, with the candidate introduced to the synthesis procedure. Chapter 5 is based on experiments that were devised by the candidate with the support of Dr Lorna Dougan, and were completed by the candidate, Dr James Towey, Dr Lorna Dougan and Dr Alan Soper. The analysis of the data was performed by the candidate with the support of Dr Lorna Dougan and Dr Alan Soper, with the exception of data published in reference 3 of the bibliography where analysis was completed by both Dr Alan Soper and the candidate. The supporting SANS experiment was devised by the candidate and was completed as part of an ISIS SANS Xpress proposal. The SANS experiment was completed by Dr Stephen King and the data analysed by Dr Stephen King and the candidate. The articles were written by the candidate and Dr Lorna Dougan, as well as Dr Alan Soper where relevant. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Bibliography

(1) N.H. Rhys, A.K. Soper and L. Dougan, The hydrogen-bonding ability of the amino acid glutamine revealed by neutron diffraction experiments, Journal of Physical Chemistry B, 116(45), 13308-13319 (2012).

(2) N.H. Rhys and L. Dougan, The emerging role of hydrogen bond interactions in polyglutamine structure, stability and association, Soft Matter, 9, 2359-2364 (2013).

(3) N.H. Rhys, A.K. Soper and L. Dougan, Hydrophilic association in a dilute glutamine solution is persistent with increasing temperature, (Submitted Manuscript).

For my Mum, who told me I could do anything I put my mind to, and to my Dad, who told me that everything I do I should do with a smile on my face.

Acknowledgements

Over the course of my PhD I have developed an endless number of people to thank. Firstly, I wish to thank my supervisor, Dr Lorna Dougan, for taking me on to do a doctorate and giving me the opportunity to do something I have always loved doing, research. To my collaborator, Dr Alan Soper, I remain consistently in your debt for passing on your endless supply of knowledge, your advice and allowing me to 'come and annoy you' on many an occasion. My sincerest thanks goes to my examiners, Dr Johan Mattsson and Professor Jayne Lawrence, who agreed to examine this thesis, allowed me to have a doctorate and provided me with feedback to polish this final version. This project was completed with funding from a European Research Council, granted to Dr Lorna Dougan. The neutron experiments were completed through granted beamtime funded by the STFC.

My project has allowed me to collaborate with numerous people. My gratitude goes to Drs Tom McAllister and Michael Webb for synthesising the molecule that features most in my thesis and for introducing a physicist to chemistry. My thanks goes to the late Dr Julie Fisher and Dr Bruce Turnbull for help with understanding glutamine from an NMR and thermodynamic perspective respectively. To be able to complete the numerous simulations required for this thesis, I thank Dr Peter Hine for access to computing facilities in the school.

In doing a doctorate it is not all about research but also the environment you are in, which keeps you strong. I am grateful for the useful discussions and support from my lab group, as well as to the people in MNP who will never realise how much their presence, dynamic and enthusiasm made working in this research sector an absolute joy. Within my group, I thank whole-heartedly my other member of 'Team Neutron', Dr James Towey; with our ISIS trips, neutron meetings, we developed and maintained a true camaraderie. I'm also immensely appreciative of the help of Drs Simon Connell, Marcelo da Silva and Kasia Tych, who each read substantial amounts of this thesis in the writing and/or correction stages, and whose guidance and care at the end I truly depended on. Others who I wish to acknowledge inside and outside of the university include Radhika Patel, Vida Sudentaite, as well

as Drs Sebastian Busch, Johanna Galloway, Kym Eden-Jones, Jono McKendry, Khizar Sheikh, May Wheeler and Oleksandr Zavalov whose critical eyes and/or friendly ears helped me make this thesis what it is.

Over the course of my doctorate, I have received guidance and useful feedback from external parties. To those who have reviewed my work and even occasionally passed criticism, I shall always remain grateful to them for providing me with more insight into my own project as well as increased confidence in my own scientific judgment. The Water & Aqueous Solutions Gordon conferences have been a huge extra part of my PhD experience and from this I thank Dr Matthias Heyden, as well as Professors Nancy Levinger and Douglas Tobias who I organised the 2014 meetings with. Whilst not really impacting on my project, I will never be able to able to thank them enough them for believing in my capabilities as this was something I could always keep with me at the hardest of times.

Finally, I would never be here were it not for the the support of my family; including my sister and her growing family, my grandparents, but most of all my parents, who have believed in me from day one. From supporting me all the way prior to starting my doctorate and onwards, they have always been there for every positive and negative, and this thesis would not exist without them.

Natasha H. Rhys

Abstract

Peptoids are synthetic mimics of naturally-occurring peptides (the natural building blocks of proteins). These interesting building blocks are more resistant to temperature, pH and solvent denaturation, and possess a more flexible backbone compared to their peptide counterparts. As a result, peptoids are emerging as promising biomimetic materials for a range of applications. Despite this potential, very little is known about the intermolecular interactions which determine their stability and solubility, including hydrogen-bonding, hydrophobic interactions and hydration properties, as well the structural properties of the individual imino acid blocks that make them. This thesis presents neutron diffraction experiments coupled with isotopic substitution and computational modelling to complete a structural study on a model imino acid in aqueous solution. The focus here has been on glutamine, a molecule that is capable of forming multiple hydrogen bonds and is thought to self-assemble through preferential side-chain hydrogen bond interactions, making it an interesting model system. Furthermore, glutamine is important in many biochemical processes and its presence has been associated with a number of neurodegenerative diseases. By probing the structural properties of the model system in aqueous solution and of its amino acid equivalent, L-glutamine, it has been possible to uncover details of important intermolecular interactions that govern the properties of these biomolecules. The interactions of the model imino acid have also been observed with respect to temperature and concentration and compared to a naturally-occurring imino acid, sarcosine, to determine the impact of side chain on imino acid interactions. This work also serves as a reference for completing structural studies on such biomolecules, covering methodological benefits and limitations. This is the necessary first step in building a framework to understand the self-assembly and driving forces in more complex peptoid and peptide structures and serves as a reference for completing neutron studies on natural and synthetic biomolecules.

Contents

List of Figures

List of Tables

Abbreviations

Biological Molecules

Gln	L-glutamine
Ngln	Glutamine imino acid, formally known as N-(2-carboxamidoethyl)glycine
Sar	Sarcosine

General Abbreviations

ATLAS	Analysis of Time-of-flight diffraction data from Liquid and Amorphous Samples
EPSR	Empirical Potential Structure Refinement
IR	Infra Red (light with a wavelength of 700 nm - 1 mm)
ITC	Isothermal Calorimetry
LJ	Lennard-Jones
MC	Monte Carlo
MD	Molecular Dynamics
NEXAFS	Near-Edge X-ray Absorption Fine Structure
NMR	Nuclear Magnetic Resonance
RDF	Radial Distribution Function
SANDALS	Small Angle Neutron Diffractometer for Amorphous and Liquid Samples
SANS	Small Angle Neutron Scattering
SDF	Spatial Density Function

Units

Angular Units

sr	steradian (the solid angle subtended at a centre of a unit sphere by a unit area on its surface).

Concentration Units

M	Molar Concentration (equal to $\mathrm{mol\,L^{-1}}$)

Energy Units

cal	One calorie = 4.184 J
eV	One electron Volt = 1.602×10^{-19}J
J	One Joule

Length Units

m	metres
mm	$\times 10^{-3}$m
μm	$\times 10^{-6}$m
nm	$\times 10^{-9}$m
Å	$\times 10^{-10}$m
barn	$100 \times \mathrm{fm}^2$

Mass Units

g	grams
mg	$\times 10^{-3}$g

Pressure Units

bar	One bar (= 100,000 Pa)

Thermodynamic Parameters

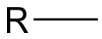
ΔH	Change in Enthalpy
ΔS	Change in Entropy
K	Degree Kelvin (equal to °C + 273.15)
mol	One mole
N_A	Avogadro's Constant (equal to $6.0221 \times 10^{23}\,\mathrm{mol^{-1}}$)

Volumetric Units

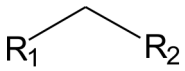
l	Litre (or L)
ml	millilitre

Chemical Groups

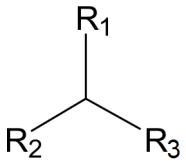
Alkyl/Hydrocarbon Groups



Methyl
(RCH_3)

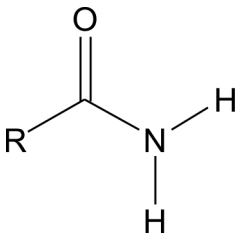


Methylene
(R_2CH_2)

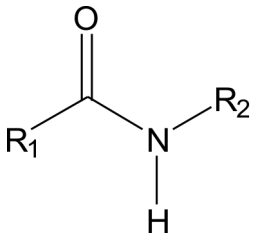


Methanetriyl
(R_3CH)

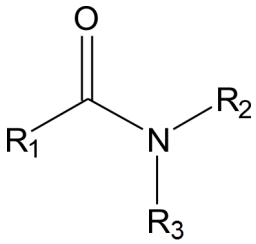
Amide Groups



Primary Amide

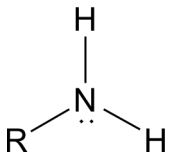


Secondary Amide

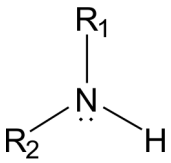


Tertiary Amide

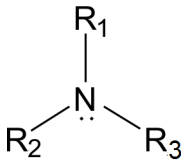
Amine Groups



Primary Amine

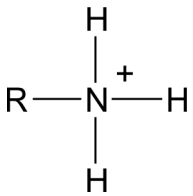


Secondary Amine

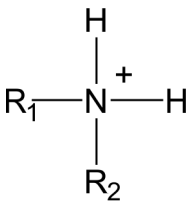


Tertiary Amine

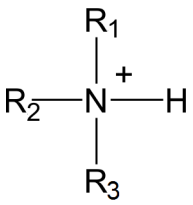
Ammonium Groups



Primary Ammonium

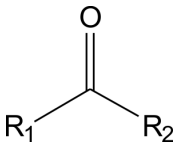


Secondary Ammonium

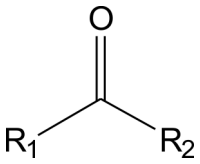


Tertiary Ammonium

Carbonyl Groups

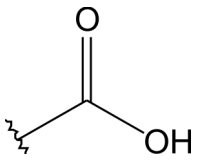


Carbonyl



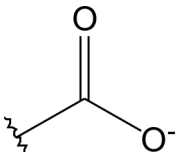
Acyl, where either R_1 or R_2 is an alkyl group

Carboxylic Acid/Carboxyl Groups

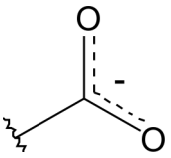


Carboxylic Acid

Carboxylate Groups

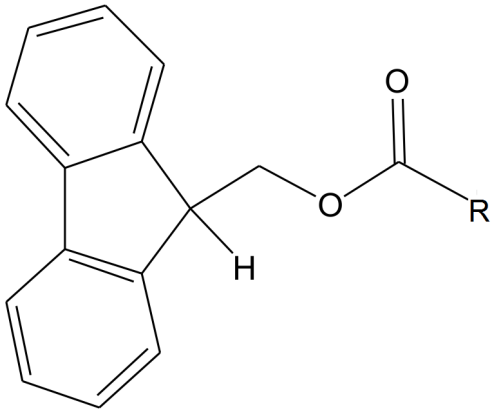


Carboxylate, in Double-Single Bond Form



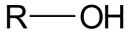
Carboxylate, in Conjugated Form

Fmoc



Fmoc

Hydroxyl



Hydroxyl

Chapter 1

Introduction

This thesis presents structural investigations on the building blocks of a novel class of molecules known as peptoids in aqueous solution. Peptoids are similar in structure to the peptide chains that form protein-like structures, but possess favourable properties such as increased resistance to denaturation and altered solubility compared to their peptide counterparts. Along with other biological entities, including carbohydrates, nucleic acids and lipids, proteins are the machines of the living world. They are responsible for myriad functions, from catalysing metabolic reactions and molecular transport to DNA synthesis and the formation of our tissues (4).

Proteins function by being able to form a specific three-dimensional structure suited to their role. This structure is the result of the sequence of amino acid building blocks they are made up of, which drive a protein's self-assembly and govern how it may interact with its surrounding environment and other entities. Single amino acids are also important for biological homeostasis, playing a role in functions such as neurotransmission and immunity. Indeed, our bodies synthesise many vital amino acids and for a healthy life we are required to obtain others through our diet (5). Right down to the very building blocks, it is thus vital that we understand how these molecules work if we are to acquire deeper knowledge of how nature functions.

This chapter will discuss the background theory behind peptoids and the proteins they mimic, as well as discuss the motivation behind looking at these molecules in solution. Section 1.1 introduces the amino acid, as well as the relevant interactions

and processes involved in forming complex protein structures. The concept of a peptidomimetic will be presented, as well as the structure, properties and applications of peptoids and peptoid-based structures. Section 1.2 covers the theory and methods used to study biomolecules in solution. Finally, Section 1.3 will introduce the project to be presented in this thesis, introducing the aims as well as the three model systems to be studied, L-glutamine, the glutamine imino acid and sarcosine. This chapter serves as a foundation prior to describing the methods (Chapter 2) and the results (Chapters 3, 4 and 5).

1.1 Peptides and their Peptoid Mimics

In order to appreciate how peptoids and other peptidomimetics function, it is important to understand how they compare to their natural peptide versions. Section 1.1 discusses the fundamentals of amino acids and their development into complex protein structures (Section 1.1.1), interactions important for biomolecular association (Section 1.1.2), the development of peptidomimetics (Section 1.1.3), and background information on the structure, synthesis, properties and applications of peptoids (Section 1.1.4).

1.1.1 Proteins and their Amino Acid Building Blocks

1.1.1.1 Amino Acids

Proteins are formed from a sequence of α-amino acids[1]. These amino acids possess a common backbone structure, with a central methanetriyl α-carbon (> CH -) covalently attached to a primary amine (- NH_2) group and a carboxylic acid (- COOH) group. The structure of an amino acid is shown in Figure 1.1 and diagrams of the individual functional groups are presented in the *Chemical Groups* section at the start of this thesis. Attached to the α-carbon of the backbone is a side chain group, also known as the R group of the amino acid. The side chain constitutes a functional group unique to the

[1]An α-amino acid possesses one α hydrocarbon group on the backbone, between the amine and carboxyl groups. If a second hydrocarbon group were present on the backbone in sequence, this would be a β-amino acid.

amino acid.

When the backbone of an amino acid contains an amine and carboxylic acid group, the molecule is in a neutral form (Figure 1.1). The neutral form is applicable to the pure amino acid in solid state. In solution and without any buffering agent, the dominant species is the zwitterion. A zwitterion possesses both positive and negatively charged terminals such that the net charge still equals zero. The amino acid zwitterion is produced by the loss of a hydrogen from the carboxylic acid and the gain of a hydrogen atom on the amine; this produces a carboxylate group (- COO^-) and an ammonium group (- NH_3^+).

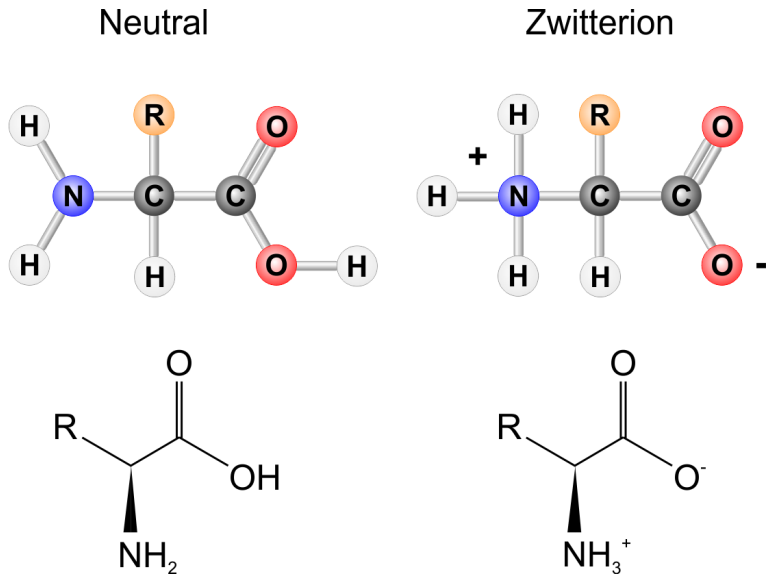


Figure 1.1: The structure of an amino acid. Each consists of a common backbone structure and a side chain or R-group (R) that is unique to a given amino acid. In **neutral** form, the backbone consists of an amine and a carboxylic acid group. In solution and without any buffering agent, the dominant species is the **zwitterion**, whereby a hydrogen atom has been transferred between termini, forming a a charged ammonium and carboxylate group.

The pH at which the charges balance for all amino acids in the system is known as the isoelectric point (pI). The pI of a molecule can be calculated using the following equation:

$$pI = \frac{pK_a + pK_b}{2} \tag{1.1}$$

Here, pK_a and pK_b are acid dissociation constants and signify the transition pH when a particular electronic species becomes dominant in the system (Figure 1.2). For a pH lower than pK_a, the amino acid molecules are predominantly positive, with a positive ammonium group and a neutral carboxylic acid group, due to there being an excess of H^+ ions in the system. For a pH higher than pK_a, the amino acid molecules are predominantly negative, with a negative carboxylate group and a neutral amine group, due to there being an excess of OH^- ions in the system.

A carboxylate group is typically portrayed with a distinct double and single-bonded oxygen atom present. Realistically, this is a conjugated system whereby each bond fluctuates between double and single-bonded, with the two states effectively indistinguishable. The conjugated nature of this group thus means the carboxylate group can be represented in the manner shown in the list of *Chemical Groups* given prior.

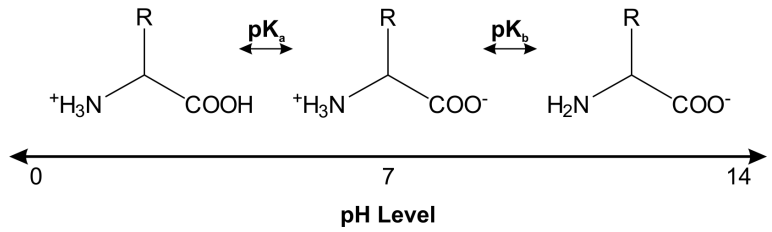


Figure 1.2: Amino acid species across the pH scale. The isoelectric point is the midway point between two acid dissociation constants, pK_a and pK_b, which each mark a transition between whether a neutral or charged version of the amino acid is dominant.

The asymmetry of the α-carbon, attached to four distinct groups, renders an amino acid chiral. The molecules can exist in two enantomeric forms, which have the same composition but are mirror-images of each other (Figure 1.3). Though similar in structure, the properties of each enantiomer may differ (6). Amino acids can either be in left configuration (L-) or right configuration (D-). The amino acids found in naturally-occurring proteins are of L-configuration.

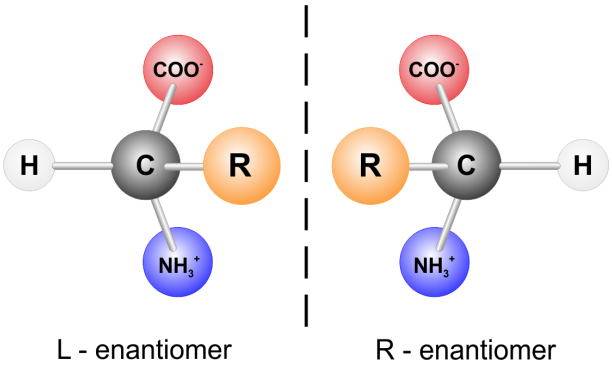


Figure 1.3: The chiral amino acid. As the alpha-carbon is asymmetrical, this means an amino acid can exist as either one of two enantiomers, which are superimposed mirror images of each other. An amino acid may either be in left (L-) and right (D-) configuration. Proteins are made of chains of L-amino acids.

1.1.1.2 From Amino Acids to Proteins

There are 23 standard amino acids found in proteins, 20 of which are directly and regularly encoded by DNA (7). The 20 DNA-encoded amino acids are shown in Figure 1.4, along with their respective three-letter and one-letter notation. Their functional groups vary in terms of polarity (presence of dipoles), hydrophilicity (ability to bond with water molecules) and whether they are charged. Those amino acids that not considered standard are referred to as non-canonical. Non-canonical amino acids can be naturally-occurring or synthetic.

A protein is effectively a self-assembled polymer chain of amino acids, and to attain these structures requires the formation of many intermediates. The protein folding process is portrayed schematically in Figure 1.5. The transition from amino acid building blocks to a protein is a multi-stage process. The 20 DNA-encoded amino acids are assembled into a sequence in living systems by being allocated a DNA codon, a 3-nucleotide sequence that labels them (4). A given segment of DNA will code for a specific protein sequence, with the corresponding individual DNA strand able to act as a template for producing more proteins of that kind. The DNA strand, after being unravelled from the double helix by the RNA polymerase enzyme, is transcribed to produce a chain of messenger RNA (mRNA) molecules that are then able to migrate from the nucleus to a cell ribosome. Transfer RNA (tRNA) molecules bring corresponding

amino acids to the ribosome surface from the cytoplasm. By translating the codons in the sequence, the amino acids may join together in a specific order.

In vivo and *in vitro*, a protein sequence is initially formed by a bond forming between each amino acid. For a given amino acid pair, the backbone ammonium and carboxylate groups can react to form an amide linkage; known as a peptide bond. This is a condensation reaction and will also produce a water molecule. This process can be repeated to form a peptide - a short chain - or a polypeptide - a longer chain. This is the primary structure of the protein, and it is the specific combination of amino acid residues that enables a protein to self-assemble into its unique 3-D structure.

The functional groups on the backbone and the residue side chains are capable of non-covalently interacting with each other. The theory behind these non-covalent interactions is described further in Section 1.1.2. The amide groups on the backbone are able to participate in hydrogen-bond interactions with other portions of the molecule. Such backbone-backbone interactions are important in the formation of a protein's secondary structure. The common structures produced at this stage include α-helices and β-sheets, the latter of which is formed from multiple β-strands. A random coil is a portion of a protein sequence not involved in a distinct secondary fold, with the backbone conformation of each unit independent of the conformations of its neighbouring residues.

Subsequently, it is possible for other non-covalent interactions to take place between the side chains of the residues, allowing the protein to take upon its unique 3-D structure. Possible interactions include hydrophobic and electrostatic interactions. The 3-D complex formed is referred to as the tertiary structure of the protein. The intermolecular interactions that can exist for protein-based and other biological molecules is presented in Section 1.1.2. It is possible for a protein to have a quaternary structure if individual protein subunits interact to form a complex, but this level of structure is not present for all proteins.

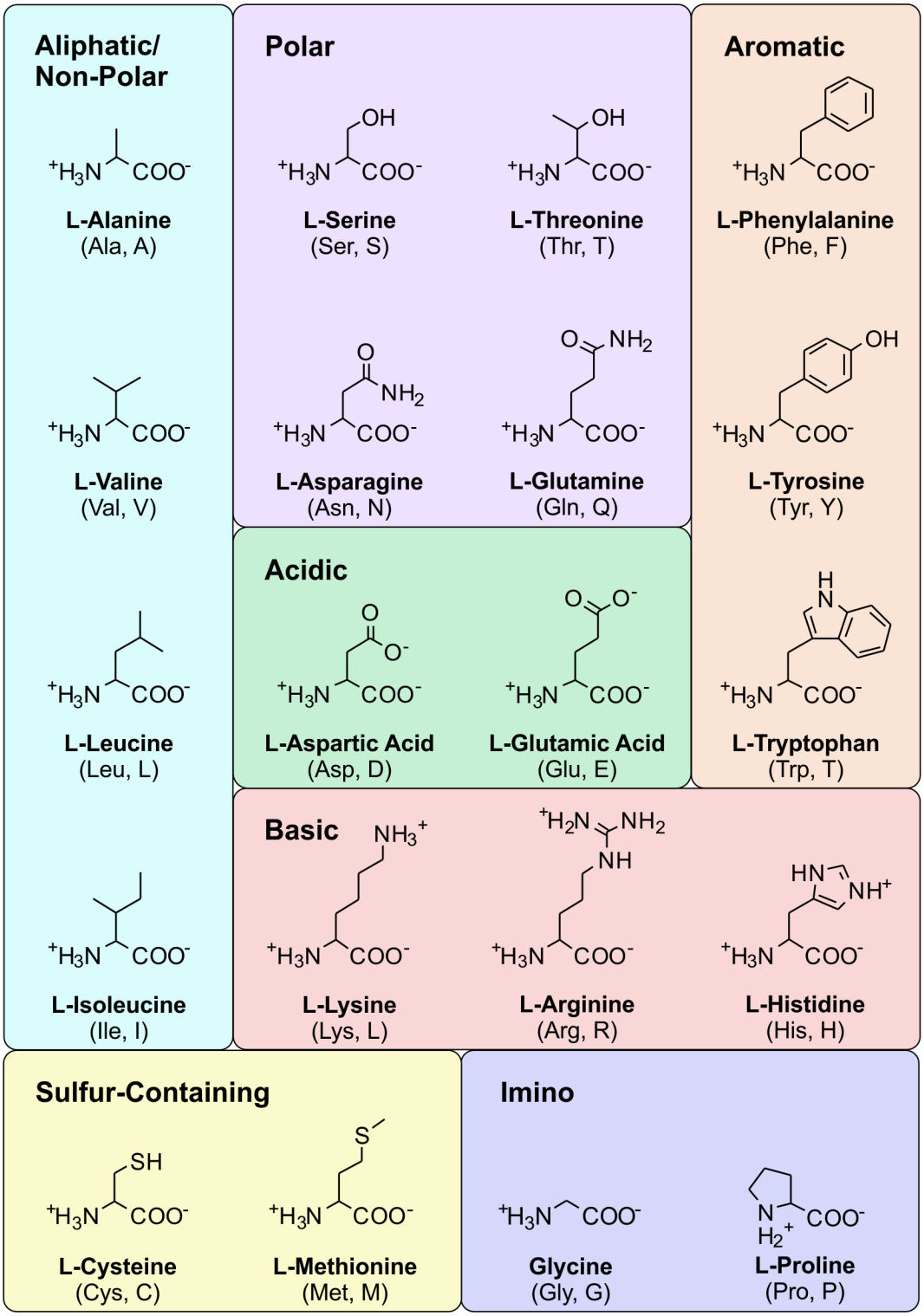


Figure 1.4: The 20 standard amino acids that are encoded by DNA. Each amino acid is shown additionally with its three-letter and one-letter notation. The amino acids categorised as 'imino' can be classed as both imino and amino acids.

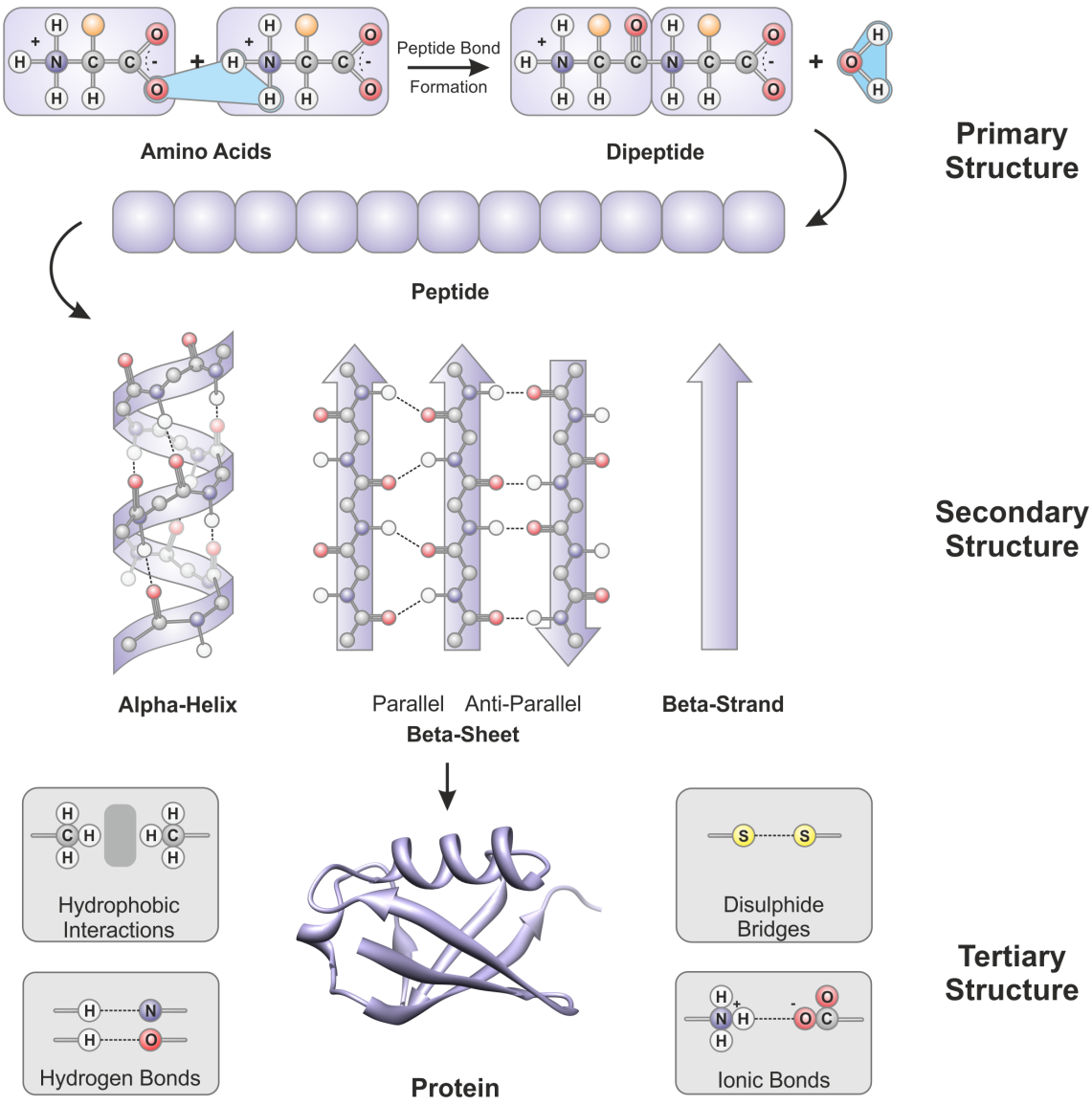


Figure 1.5: The process of producing a protein structure. For the **primary structure**, amino acids joined via a peptide bond to peptides, short chains of amino acid residues. These residues can interact with others in the chain via hydrogen bonding and produce structures such as α-helices, as well as parallel or anti-parallel β-sheets, formed by the interaction between multiple strands. These entities form the **secondary structure** of the protein. These structures may further interact via other interactions, including hydrophobic interactions and ionic bonds, to produce the **tertiary structure** of the protein. The example protein shown is ubiquitin (PDB 1UBQ) (8)

.

1.1.2 Interactions of Protein-based Molecules

What allows any protein-based material to form complex 3-D dimensional structures is down to the molecular interactions it makes on an intra- and intermolecular basis. A summary of these bonds with their comparative strength is given in Figure 1.6.

A sequence of amino acids are bonded together due to a series of peptide bonds being formed through the charged termini of the amino acids (Section 1.1.1.2). This amide link is classed as a covalent bond. A covalent bond is formed by the sharing of electron pairs between atoms and is the the type of interaction that joins atoms together to form a molecule. A simple example of this is the methane molecule shown in Figure 1.6, whereby 4 hydrogen atoms each contribute an electron so as to fill the 8-electron capacity of the carbon electron shell. Beyond this, non-covalent interactions govern how single amino acids interact and peptide chains fold to form proteins, many of which depend on the polarity of the molecule. The key non-covalent interactions that will be discussed in this thesis will be discussed in detail in the subsections to come.

A non-covalent interaction is characterised by two potential parameters; the Lennard-Jones potential and the Coulombic Potential (9). The Lennard-Jones (LJ) potential describes the interaction between two uncharged particles, atoms or molecules, with respect to distance. A schematic of the LJ potential is presented in Figure 1.7. When two uncharged particles are close to each other a repulsive force acts between them, due to the overlapping and distortion of their electron clouds at shorter distances. The potential is mildly attractive as two uncharged particles approach one another from larger separations, due to the presence of induced dipoles; a process described in more detail in Section 1.1.2.1.

The LJ potential for two atoms is described by the equation:

$$U_{LJ}(r) = 4\epsilon[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}] \tag{1.2}$$

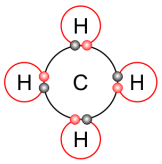
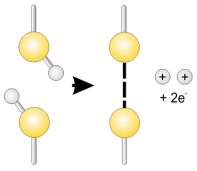
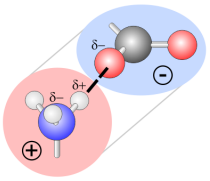
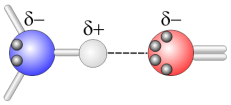
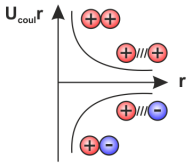
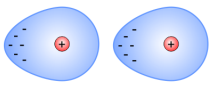
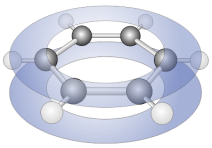
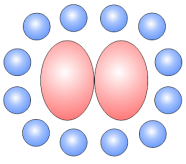
Interaction	Definition	Diagram	Typical Distance (Å)	Strength (kJ mol^{-1})
Covalent Bond	A bond between two atoms that involves the sharing of electrons		0.6 - 3.0, depending on the covalent radius	~100 - 1000
Disulphide Bond	A covalent bond formed between two reduced S-H groups		2.2	167
Salt Bridge	A special case of a hydrogen bond formed, where the donor and acceptor atoms are fully charged.		< 3.5 (typically 2.8)	~12.5 - 17, up to 30 for fully or partially buried salt bridges
Hydrogen Bond	A non-covalent interaction between a donor atom, bound to an electropositive hydrogen atom, and an electronegative acceptor atom.		< 3.5 between donor and acceptor atoms (typically 3)	~2 - 6, for water, 12.5 - 21 if either donor or acceptor is fully charged
Long-range Electrostatic Interaction	A non-covalent Coulombic interaction between charged atoms or groups. This falls off with a 1/r dependence.		Variable	Environment dependent, strong in non-polar ones, weak in polar ones
van der Waals Interaction	A weak attractive force between two atoms or groups arising from fluctuations in electron distributions.		< 3.5	4 on a surface, 17 in a protein interior
Aromatic Interaction	An aromatic ring interacting with an ion (cation-π), polar group lone pair (n–π*) or a ring (stacking), due to its conjugated π-orbital electrons.		< 7 depending on interaction and orientation of aromatic ring	< 20 depending on interaction and orientation of aromatic ring
Hydrophobic Force	The tendency of hydrophobes to associate and reduce contact with water, governed by thermodynamics		Variable	< 40

Figure 1.6: A summary of the non-covalent interactions in biological molecules. The values are taken from (7; 9–11).

In Equation 1.2, ϵ is the energy value for the distance at which the two particles are at equilibrium, whilst σ is the the distance at which the intermolecular potential between the two particles is zero, also known as the van der Waals radius. Both ϵ and σ are specific to the particles in question. The repulsive part of the potential is related to the separation by $1/r^{12}$ and the attractive part by $1/r^{6}$.

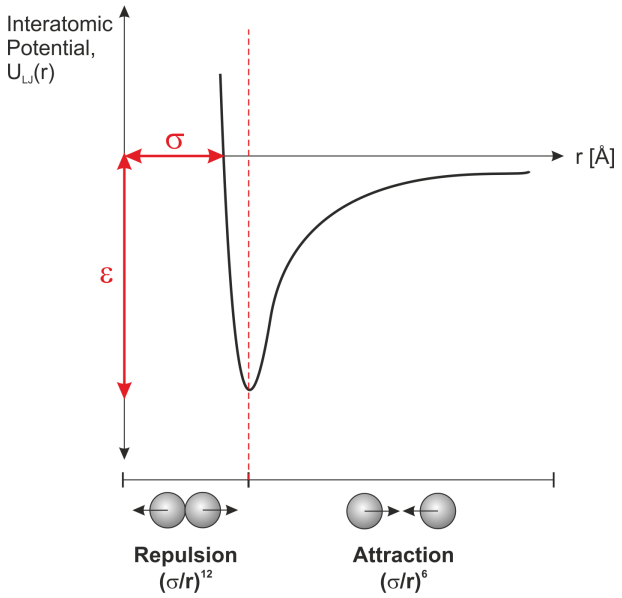


Figure 1.7: Schematic of the Lennard-Jones Potential used to define non-covalent interactions.

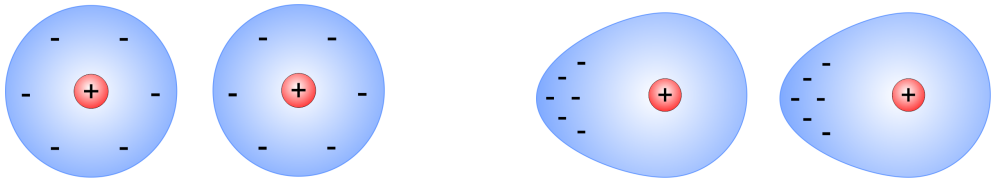
The Coulombic potential describes forces due to electrical charges. Atoms typically possess a partial charge and it is these that determines if a molecular species is charged or polar. It is an effective pair potential that describes the interaction between two point charges and is given by the equation:

$$U_C(r) = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r} \tag{1.3}$$

Here, ϵ_0 is the electrical permittivity of space and r the distance between two charges, q_1 and q_2. It is the balance of Coulombic potentials, as well as LJ potentials that govern the interactions between atoms, which are discussed next.

1.1.2.1 Dipoles and van der Waals Interactions

An atom within any molecule consists of a positive nucleus of neutrons and protons, surrounded by a negative electron cloud. The idealistic view of this cloud is that the electrons are distributed evenly around the nucleus, as shown in Figure 1.8(a). Realistically, the movement of electrons is random and it is common for instantaneous regions of higher electron density to be present (Figure 1.8(b)). As these electrons fluctuate around a nucleus, they create time-varying dipoles.



(a) (b)

Figure 1.8: The instantaneous formation of dipoles in van der Waals type interactions. Whilst the distribution of electrons in the electron cloud may be considered uniform 1.8(a), it is possible for there to be regions of greater electronegativity, which brings about a dipole moment. These can happen on a molecular level and if in sync with dipole moments from another molecule, it is possible for an interaction to be present.

For covalently-bonded atoms, the electron clouds overlap. As the distribution of electrons is uneven, this allows certain atoms to become more electronegative and others electropositive at a given moment. For two molecules, an instantaneous dipole on one molecule can attract that of another, to form an interaction. Whilst the dipole created at that point may be instantaneous, the electron clouds of each molecule may oscillate cooperatively in order to maintain an interaction. Such attractive forces are collectively known as van der Waals forces, named after Johannes Diderik van der Waals who not only proposed that molecules existed at a time when this was still under debate but that they may attract each other (12). It is these dipoles that are responsible for the LJ potential attractive term in Equation 1.2.

Van der Waals interactions have the potential to occur in any chemical group, though as shown in Figure 1.6 they are relatively weak. Interactions that form between two

instantaneous dipoles are sub-classed as London Dispersion forces (13). A functional group may also be able to interact if the dipole has been induced by another species possessing a permanent dipole; an effect known as a Debye force (14). It is possible for certain groups to have a permanent dipole depending on the organisation of covalent bonds and relative electronegativity of the individual atoms. When two permanent dipoles interact, this is a type of Keesom force (15).

1.1.2.2 Hydrogen Bonding

An important example of a permanent dipole driven interaction found in biological systems is the hydrogen bond. This is the interaction between an electropositive hydrogen atom and an electronegative atom, typically oxygen or nitrogen for biological systems. The electropositivity of the hydrogen is sustained by being attached to another electronegative atom. An atom can be permanently electronegative if it has negative lone pairs of electrons. If a comparison is made with the covalently-bonded methane molecule (Figure 1.6), the methane carbon is not electronegative as all pairs consist of one electron from carbon and another from a hydrogen atom.

One of the most common examples of hydrogen bonding occurring in nature is that for water, as shown in Figure 1.9. Each water oxygen atom has two lone pairs of electrons that a hydrogen atom from another molecule can bond to. Hydrogen bonds can also occur between protein-based molecules and these interactions are key in the formation of secondary structures, as detailed in Section 1.1.1.2. Interactions can for instance take place between the amide hydrogen and oxygen atoms on the backbone of peptides (Figure 1.6), which are crucial in forming the common β-sheet structures found in proteins.

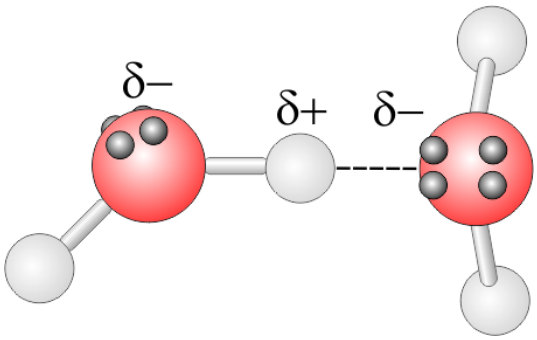


Figure 1.9: Hydrogen bonding in water. Hydrogen bonds are found between molecules in bulk water and are possible by the presence of two lone electron pairs (grey) on the oxygen (red), making this atom electronegative and the hydrogen atoms (white) electropositive.

1.1.2.3 Electrostatic Interactions

If molecules possess a net charge, it is also possible for them to interact through electrostatic interactions. One of the simplest examples of this is the attraction between ions. A salt such as sodium chloride is produced via a redox reaction (6), whereby an electron is transferred between the atoms to create a positive sodium anion and negative chloride cation, which are attracted to each other.

In biological matter, it is possible for salt bridges to form between charged functional groups. A salt bridge is a hydrogen bond that forms between charged pairs. A salt bridge thus constitutes a hydrogen bond and an electrostatic interaction, producing one of the strongest forces between groups. In proteins, a salt bridge is most commonly formed between negatively charged glutamic acid and positively charged lysine residues, as shown in Figure 1.10.

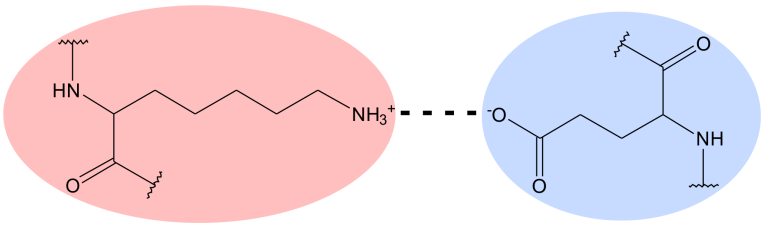


Figure 1.10: A strong interaction present for protein-based matter is the salt bridge, whereby a hydrogen bond can form between a charged species. In proteins, it is common to find salt bridges forming between positive lysine and negative glutamic acid residues.

Salt bridges typically form between ionic species. It is also possible for a similar kind of interaction to form between groups with localised partial charges, where the presence of charges strengthens the hydrogen bonds formed. An example of this is between the partially charged N and C termini of amino acids (Figure 1.6), which strengthens the hydrogen bond interactions present, despite the molecule being neutral in charge. Furthermore, hybrid interactions may occur between ionic species and polar groups. Whilst hybrid interactions are common between functional groups containing a permanent dipole, ion-induced dipole interactions are less frequently found as residues commonly possess a permanent dipole (16).

1.1.2.4 Hydrophobic Interactions and the Gibb's Free Energy

For species with charges and fixed dipoles, their electronic state renders them hydrophilic. Hydrophilic molecules readily form bonds with water molecules. Non-polar entities such as alkyl or hydrocarbon groups are hydrophobic and tend to aggregate in solution to separate themselves from water. Whilst van der Waals forces can act to bind hydrophobic species, this tendency of hydrophobes to aggregate is driven by hydrophobic interactions.

The source of the hydrophobic interaction is a thermodynamic one and concerns the change in the Gibb's free energy (9). The Gibb's free energy is defined as the energy associated with a reaction or interaction that can be used to do work. The change in

the Gibb's free energy, ΔG is defined as:

$$\Delta G = \Delta H - T\Delta S \tag{1.4}$$

In this equation, H is the enthalpy, T the temperature and S the entropy. The enthalpy is a measure of the heat in a system under constant pressure. If the change in enthalpy is greater than zero, $\Delta H > 0$, heat is absorbed and the reaction to cause this change is class as being 'endothermic'. When $\Delta H < 0$, the reaction is a heat-releasing 'exothermic' one. The entropy quantifies the disorder in the system. A system becomes more disordered when $\Delta S > 0$, whilst $\Delta S < 0$ is present in a system with increasing order. For an interaction to be spontaneous, it is necessary that the entropy and enthalpy difference balances in a way that ΔG is negative. The possible conditions that allow this are presented in Table 1.1.

Table 1.1: Thermodynamic conditions required for a spontaneous reaction to take place, where $\Delta G < 0$.

ΔH	ΔS	Reaction Type	Conditions
+	-	Non-spontaneous	Always
-	+	Spontaneous	Always
-	-	Non-spontaneous	$\lvert\Delta H\rvert < \lvert T\Delta S\rvert$
		made spontaneous by	Decreasing T
+	+	Non-spontaneous	$\lvert\Delta H\rvert > \lvert T\Delta S\rvert$
		made spontaneous by	Increasing T

The earliest studies on hydrophobic interactions date back to the work of Meyer and Overton from 1899 onwards in attempts to relate the potency of anaesthetic gases to their hydrophobicity (17–19). Following studies on the connecting between the heat of hydration and entropy (20) as well as the formation of ice-like structures around non-polar molecules (21), it was Walter Kauzmann in 1959 (22) whose work recognised that the hydration entropy was the driving force behind the hydrophobic interaction and the clustering of non-polar species.

When non-polar species aggregate in solution, they form as few a contacts with water as possible, as shown in Figure 1.11. Specifically, water-hydrogen bonds will break in order to accommodate the hydrophobe, but bonds will not be made directly with it. Instead, the water molecules arrange themselves so as to form a clathrate cage around the hydrophobe. The formation of this cage introduces order into the system and thus the entropy change is negative. This is a non-spontaneous interaction.

An interaction that is spontaneous is that between the individual non-polar species. For this, the enthalpy has to increase due to parts of the clathrate cage having to be broken. The entropy difference will also be positive as disrupting the clathrate cage will introduce disorder. With a small enthalpy change and a large entropy difference, the free energy will consequently be negative, meeting the requirements for a spontaneous interaction.

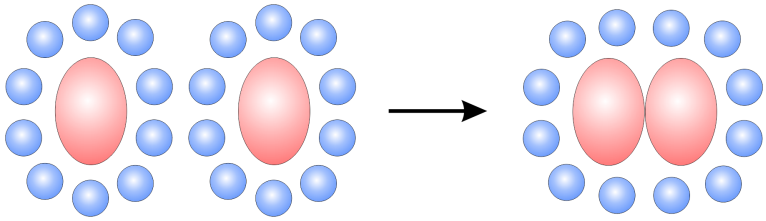


Figure 1.11: A schematic of the hydrophobic effect, which was first observed when noted that hydrophobes (red) cluster in water (blue). By doing this, they can form fewer contacts with water than if they were separate. This is possible due to a positive increase in the change in entropy, which makes this interaction spontaneous.

The interaction of hydrophobic groups is thought to have a major role in the self-assembly of biomolecules, such as proteins. However, much is still unknown about hydrophobic effect, with research continuing to further understand the kinetics, properties and relative importance of these forces compared to interactions through hydrophilic groups (23). As part of this thesis, the prevalence of association through hydrophilic and hydrophobic functional groups will be discussed.

1.1.3 Peptidomimetics

As well as their copious existence in nature, peptides are being exploited in biological, pharmaceutical and materials applications. Peptides possess many desirable properties, including cell permeability, low toxicity, low immunogenicity and the ability to form complex structures.

In certain *in vivo* and *in vitro* applications, peptides are not suitable. They are not resistant to proteolysis, the breakdown of peptide chains by enzymes. Certain peptides are poorly miscible in solution, have the propensity to aggregate or need to be short and highly lipophilic to penetrate the cell membrane (24–26).

Peptidomimetics are synthetic mimics of peptides, the natural building blocks of proteins. The development of peptidomimetics is a growing field of science (24). This field aims to create structures that mimic those found in nature, but with enhanced, tailored properties (24–26).

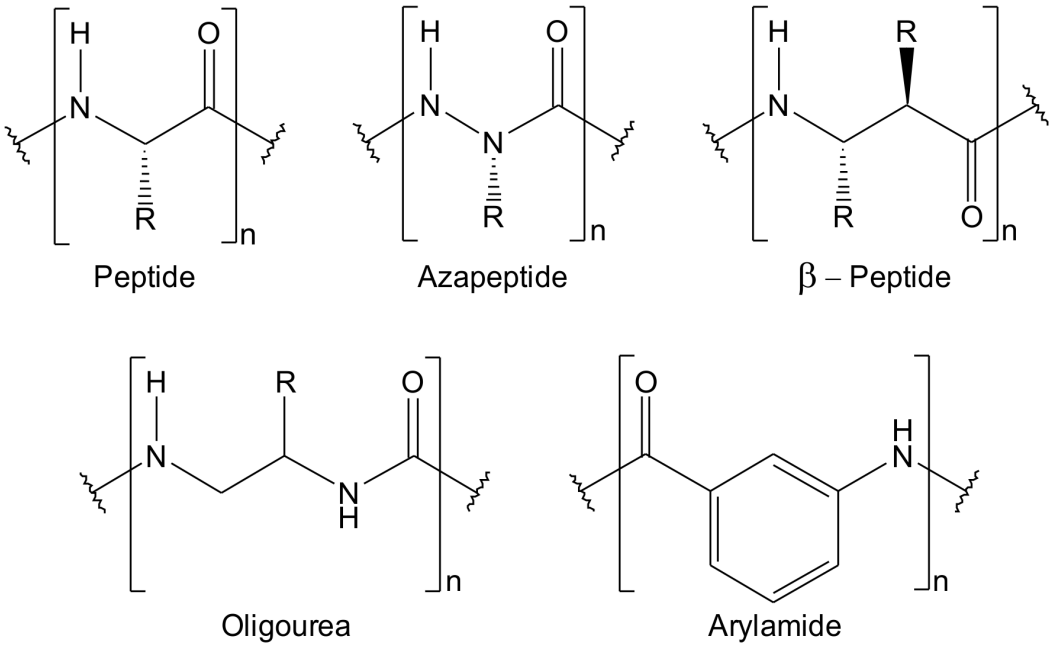


Figure 1.12: Some of the commonly exploited peptidomimetics used in biomaterial design. Changes that can be incorporated range from extending the backbone (e.g. β-peptides), to moving functional groups (e.g. peptoids), or replacing specific atoms (e.g. azapeptides).

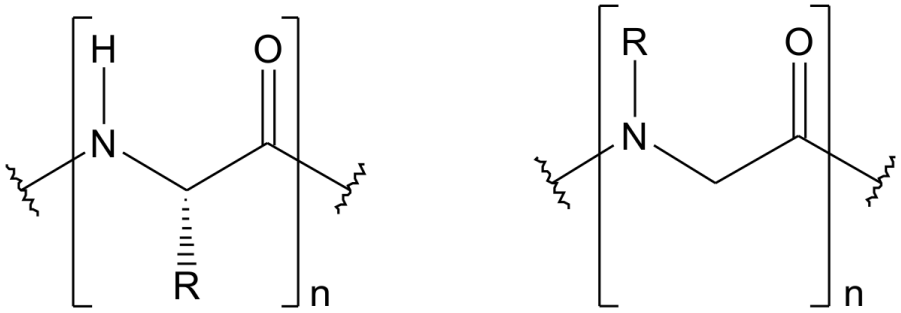
A peptidomimetic is a material considered in the broader class of bioinspired materials. Reviews such as those by Rawlings *et al* (27) define a bioinspired material as that which imitates the function of a material in nature. This term may also be applied to materials that take inspiration from nature but are not necessarily biological (27; 28). A biomimetic is a structure similar to a natural material, but with altered and often more favourable properties.

The most common route to creating a peptidomimetic is to make small structural changes to a peptide, whilst still trying to form structures that possess a resemblance to the peptide backbone, side chains, or both (24–26; 29). In altering the backbone, many synthetic structures have been produced. Some of the most common changes include lengthening the backbone length so as to increase flexibility. Structures that incorporate this include β-peptides. Atoms or functional groups may also be replaced in order to alter the electronegativity of the backbone, seen for example with azapetides mimics. A selection of peptidomimetic structures are presented in Figure 1.12. Non-canonical amino acids may be incorporated instead of altering the backbone. Research by the Korley lab group for example incorporate the synthetic amino acid benzyl-L-glutamate to enhance mechanical stability and produce materials that act similarly to elastin and resilin (28; 30). This thesis will explore the properties of building blocks from a specific peptidomimetic class, peptoids.

1.1.4 Peptoids

Peptoids are synthetic mimics of peptides in which the side chain of the molecule is attached to the backbone amine nitrogen instead of the central α-carbon (Figure 1.13) (31). The term 'peptoid' was first coined to these peptidomimetics by Farmer and Ariëns (32), with the claim that they could mimic the biological function of peptides, while being structurally distinct. The use of peptoids as possible peptidomimetics with an application was first proposed in 1992 by Simon and co-workers (31). This paper discussed the possibility of these biomolecules being used as pharmaceuticals due to their resistance to proteolysis. The use of peptoid-based technology was eventually pioneered by Ronald Zuckermann, when he developed a method to synthesise these materials

that would be both simple and cost-effective (33) (see Section 1.1.4.2). This synthesis method has opened up a whole area of research into this molecular class, with the groups of Zuckermann and many others frequently publishing on literature into the properties and uses of peptoids (34–37). This section will go into detail about peptoids, their monomer building blocks and the applications for which they are commonly exploited.



(a) (b)

Figure 1.13: The structure of a peptide (a) versus a peptoid (b). For a peptoid, the side chain is attached to the backbone nitrogen instead of the α-carbon.

1.1.4.1 Peptoid Nomenclature

Peptoid chains, or oligomers, are made up of sequences of peptoid monomers. These monomers, the equivalent of amino acids to peptides, are known as N-substituted glycines (NSGs) (31), as their structure is that of a glycine molecule (see Figure 1.5 for structure of glycine) with a functional group attached to the backbone nitrogen, which is not merely a hydrogen atom. NSGs are also known more simply as imino acids. It should be noted that imino acids are also defined by IUPAC as a molecule containing a carboxylic acid and an imino ($> C = NH$) group (38), but this designation is also applicable to NSG's. For simplicity, the term 'imino acid' will be used to refer to a peptoid monomer version of a molecule for the rest of this thesis.

As shown in Figure 1.4, the standard amino acids have both a three-letter and one-letter notation. For the imino acid equivalents of these, the letter N is put on the front of the original notation (34; 39). For example, the imino version of the amino acid L-serine

(Ser, S), has the three-letter notation of Nser and the one-letter notation of nS. This terminology will be used in certain figures in this thesis.

1.1.4.2 Peptoid Synthesis

In the literature by Simon and co-workers in 1992 (31), the first peptoids were synthesised using Fmoc-based molecules (see *Chemical Groups*), the same as that used for producing peptide chains. This method though requires the production of large stocks of imino acid monomers (33).

It became easier to study peptoid structures by the development of a solid-state synthesis method, first proposed by Zuckermann and co-workers (33). This now commonly used method instead requires a variety of amine-based compounds, which are readily available and thus makes the process inexpensive compared to α-peptide synthesis. The starting product consists of a Rink-amine linker connected to a polystyrene bead. The amine portion is first acylated using bromoacetic acid, mediated with N,N-diisopropylcarbodiimde (DIC). In this acylation reaction, the hydroxyl ($-OH$) group is removed from the bromoacetic acid and the remaining acyl group attached to the amine. An amination reaction is then used to replace the bromine with a primary amine group that will also possess the side chain of interest. This procedure is repeated until the desired chain has been synthesised, by which the polystyrene portion is cleaved using trifluoroacetic acid, producing a chain with amide-capped termini.

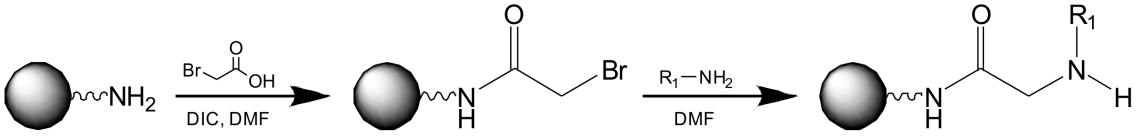
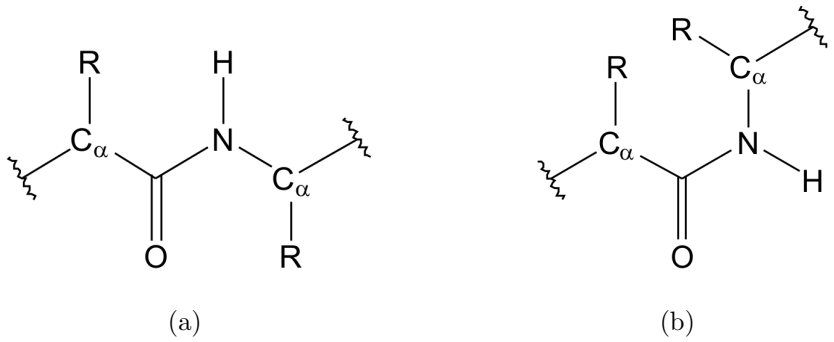


Figure 1.14: The solid-state synthesis method for producing peptoid chains. This two-step procedure requires acylation of a Rink-amine linker, followed by an amination reaction to introduce the side chain and can be repeated with ease until the desired chain length is reached.

1.1.4.3 Peptoid Properties

The simple change of the side chain being attached to the nitrogen induces notable changes to the properties of peptoid molecules. For each residue, the α-carbon is bonded to two hydrogen atoms and consequently renders them achiral. This change alone reduces the conformational rigidity of peptoids. The flexibility increase is also attributable to the backbone amide groups which are also tertiary and are able to isomerise between trans- and cis- conformations more readily than the secondary amides in α-peptides. The trans- and cis-backbone isomers are shown in Figure 1.15. In peptides, the amide group between residues will nearly always take on trans-conformation as the energy barrier between the cis- and trans- forms is high. The conversion to a tertiary amide lowers the energy barrier between the two states (40). Therefore, it is common to find a high probability of cis-bonds present in peptoid structures. Whilst the increased flexibility means a greater diversity of structures can be produced, it may prove a weakness in certain applications (Section 1.1.3).



(a) (b)

Figure 1.15: Cis-trans isomerisation on the peptide backbone. The peptide bond between most standard amino acids takes on trans-configuration (a). If in cis-configuration (b), this would bring together two α-carbons, and consequently two side chains, where steric forces would cause them to repel each other.

Nevertheless, the change to the amide also means the loss of a donating hydrogen on the backbone, preventing them from undergoing proteolysis (31; 41; 42). This also influences the structures that can be formed, with those such as β-sheets that depend typically on hydrogen bonding with the backbone less frequently reported. Any complex structures made are dependent on the side chain functional groups (36; 37).

Whilst hydrogen-bonding is observed selectively in peptoid structures, peptoid assembly is thought to be driven by hydrophobic, electrostatic and n-π^* interactions, some of which were detailed in Section 1.1.2. Often, the diversity of 3-D structures produced is increased by attaching a functional group to the α-carbon so as to make it chiral. Peptoids have the propensity to form helical structures and able to form more complex entities such as nanosheets (43) and superhelices (44). A schematic of folding process so far observed in peptoids is given in Figure 1.16.

The dependence on the side chain means that the properties of peptoids are tunable to their application (36; 37). Peptoid structures have been noted for their increased resistance to thermal, pH and solvent denaturation (34; 45) compared to peptides, much of which has been attributed to the loss of the backbone hydrogen-bonding group.

Comparisons have been made that compare peptide and peptoid structures and suggest that their properties are attributable to more than merely the loss of the backbone amide hydrogen. Tan *et al* compared similar sequences and found that peptoids not only possessed increased cell permeability due to fewer hydrogen bonding sites, but also were more hydrophilic (46). Emphasis was placed on the importance of hydrophilic and hydrophobic interactions in determining peptoid properties, with increased hydrophilicity noted more so in shorter peptoid sequences. Research has suggested the increased hydrophilicity is seen down to the level of individual residues. Tang and Deber (47) completed chromatography experiments on peptide and peptoid residues that suggest that certain peptoid residues may be even more hydrophilic than even the least hydrophobic residue, lysine, potentially due to differing interactions with the backbone nitrogen. Indeed, imino acids have been found to be more soluble than their amino acid counterparts, suggesting greater hydrophilicity (48). Reasons for this property have included different levels of interaction with the backbone nitrogen, though a quantitative description is not yet known.

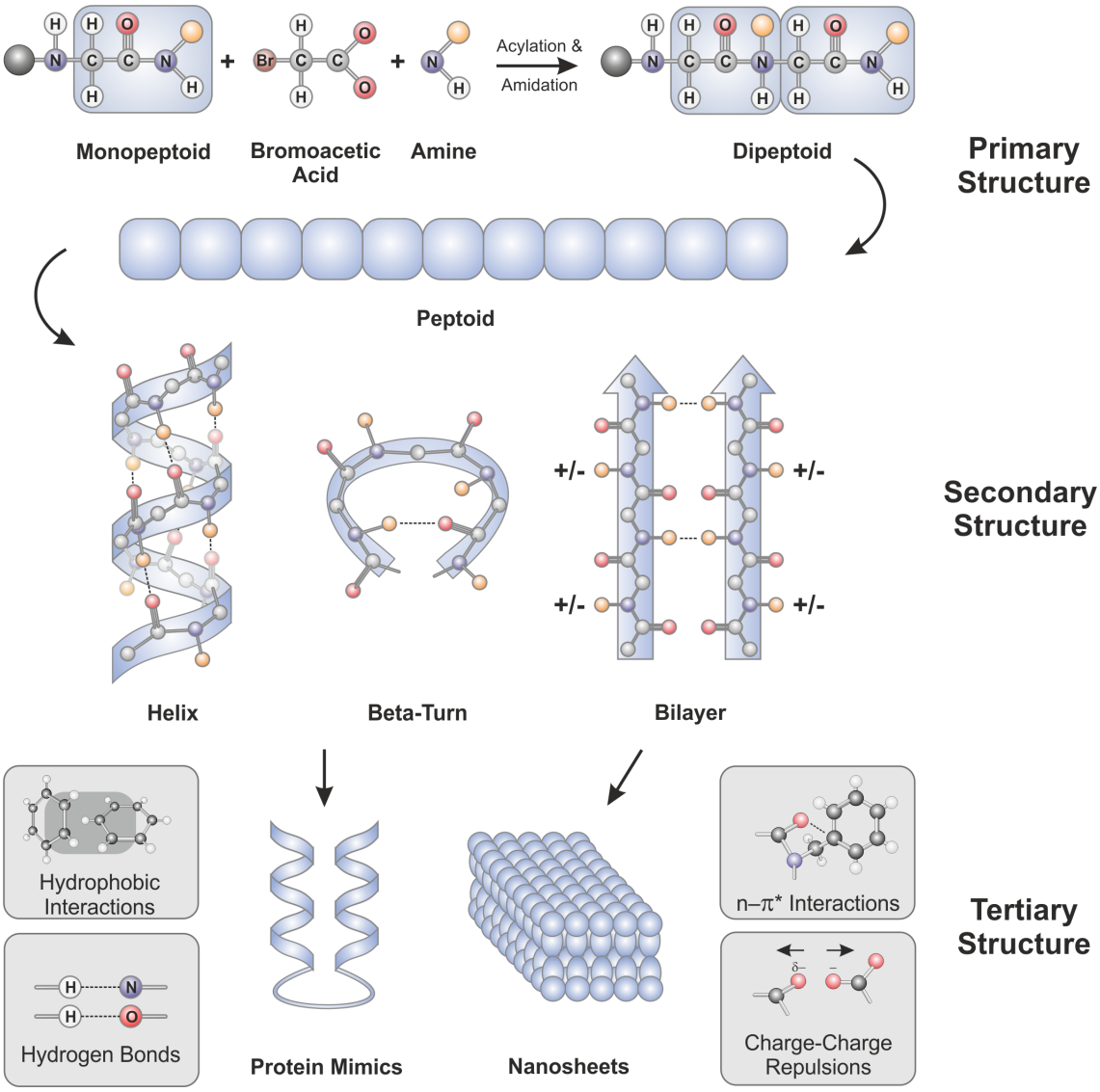


Figure 1.16: The assembly of a peptoid chain. For the **primary structure**, chains of residues are formed by a repetitive two-step process of acylation and amidation. These residues can interact with others in the chain via hydrophobic interactions, produce alpha-helices, as well as bilayer structures. If hydrogen-forming groups are present on the side chain, then structures such as β-turns may form, which are seen in proteins. These entities form the **secondary structure** of the peptoid structure. These structures may further interact via other interactions, including n$\rightarrow \pi^*$ interactions, charge-charge repulsions and steric interactions to produce the **tertiary structure**.

1.1.4.4 Applications: Long-Chain Peptoids

With the pioneering of this subclass of biomolecules in the 1990's, peptoids are emerging as tools for a range of applications. Figure 1.17 portrays just some of these applications, for which the structures span the length-scales, from Ångstroms to micrometres.

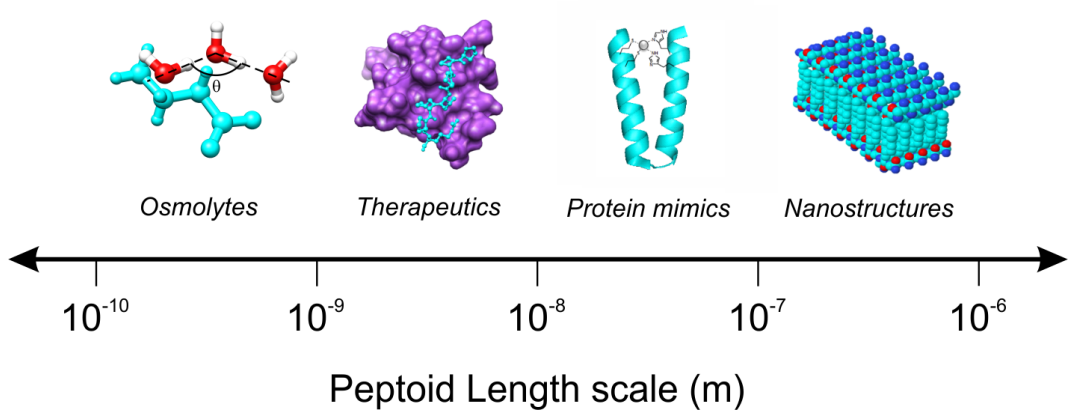


Figure 1.17: The application of peptoids across the length-scales. Whilst the smallest monomer units can function as protecting osmolytes, longer chains may self-assemble into structures involved in mimicry and nanotechnology.

With the first intention for peptoids being as a tool in the pharmaceutical industry, many peptoid molecules are being exploited as potential drug candidates for the treatment of neurological disorders, cancer and circulatory ailments (35; 49; 50). Many of these examples have determined structures with improved specificity and toxicity compared with peptide versions (50; 51) and even resistance to parasites (52). Since the 1990's, structures have been synthesised for a greater diversity of therapeutic applications, including lung surfactants (53; 54) and antimicrobial agents (55; 56). Whilst functioning as therapeutics, peptoids may be be exploited as a diagnostic tool. These peptidomimetics received much interest when a peptoid sequence was found be be able to detect and bind to Alzheimer's fibrils (57). Peptoids are now also being used to in the design of peptoid-derived artificial proteins. From the building of complex helical structures, progress has now been made towards designing zinc-binding mimics (58); a schematic of this being the Protein Mimics picture shown in Figure 1.17.

A rising area in this field is the use of peptoids in nanotechnology; a prime example being the production of nanosheets (43). Two types of chain, charged and hydrophobic, can associate in solution to form a continuous bilayer structure. Whilst only 1 nm in thickness, the size of the sheets spanned out to millimetre lengths. As the folding process of peptoids is more readily known, it may be possible to create more diverse structures for use in the nanotechnology field (35; 43).

1.1.4.5 Applications: Short-Chain Peptoids and Imino Acids

Whilst long chain peptoids are more commonly exploited, research has also being completed using shorter chains. Indeed, chains of three residues length are thought to function as potential pharmaceuticals (36).

As discussed previously in Section 1.1.4.3, it was ascertained that the results for long-chains can be attributed to the changes seen in single amino acid residues (46; 47). Whilst the backbone amide groups are not present in peptoid chains, the hydrophilicity of residues compared to their peptide equivalent is increased (46; 47). The hypothesis behind this is related to steric packing around the group and the possibility of bonds being made with the nitrogen atom. Due to this residue-based behaviour, it is common to incorporate single imino acid residues into peptide chains and produce hybrid 'peptomers', where the replacement of odd residues alone overcomes limitations the original molecule may possess (59).

If the chains are scaled down to tripeptoid level or lower, they are still hydrophilic, with the greatest difference in hydrophilicity in peptide and peptoids observed for the shortest of peptoids (46). The well-known use for such building blocks are as osmolytes. Osmolytes are molecules that affect the process of osmosis and oppose the effect of concentration changes by influencing the structure of water (60–63). Such molecules can also contribute to the protein folding process and prevent these structures from denaturing. They are utilised in multiple applications, from lowering the freezing point of water to preventing the aggregation of biological molecules. How they function can depend on the balance of functional groups within a molecule. The ability to lower the freezing point of water has typically been attributed to molecules containing hydroxyl groups,

which can readily hydrogen bond with water. Research by Huang and co-workers (64) looked at the propensity of serine residue-containing molecules to prevent ice formation and determined that the peptoid version could lower the freezing temperature much more than the peptide version. This suggests it is not only hydroxyl groups that are important in dictating the osmolytic behaviour of organic molecules.

Single imino acids also possess osmolytic capabilities. L-proline (see Figure 1.4), a hybrid amino and imino acid, with the side chain attached to both the α-carbon and the backbone nitrogen, is known to counteract osmotic stress in organisms (60; 62) and prevents the aggregation of disease causative proteins such as those involved in Huntington's disease (65; 66). The imino acid sarcosine has been recognised as being able to counteract the effects of denaturing molecules such as urea (67).

This thesis will study the properties of single imino acids. By studying the properties of individual building blocks it will be possible to elucidate information the interactions that govern their hydration and association, as well as speculate on the hydrophilic and osmolytic properties these molecules are thought to possess. This will provide a fundamental understanding of the interactions important for longer chain peptoids.

1.2 Studying Biomolecules in the Liquid Phase

Since the Simon *et al* paper (31) that first promoted peptoid structures, numerous methods have been exploited to understanding the behaviour of peptoid molecules. Such techniques include circular dichroism (68), Raman spectroscopy (69), atomic force microscopy (43), small-angle neutron scattering (SANS) (70) and fluorescence techniques (47; 58). Many of the techniques exploited provide indirect measurements of the structural properties or give insight into large-scale structures and folding.

The concept of the structure-function relationship is an increasingly common and popular theme in peptoid research (34). In the pharmaceutical field, it is commonplace for all drug candidates to undergo Structure-Activity Relationship studies, which determine the impact of the three-dimensional structure of the molecule on its biological impact. Such studies are time-consuming, but also do not give fundamental, quantita-

tive details about how these molecules interact with matter. In the literature, *X*-ray crystallography (71) or computational studies (72) are typically utilised (34–36) to extract structural information. The limitation of such studies is the absence of water. Water is ubiquitous on Earth and is key as both a solvent and reactant in numerous biological, chemical, and physical processes. *In vivo*, biomolecules exist in an aqueous medium, so an understanding of how molecules interact with water, and with other solute molecules, is vital. By completing studies in the liquid phase, it is possible to begin to attain a more realistic picture of how these molecules function in living systems.

This section will introduce how structural information can be obtained in aqueous solution (Section 1.2.1) and the techniques that can be used to extract structural information on aqueous biomolecules (Section 1.2.2). As the work in this thesis presents studies of the properties of amino and imino acid building blocks, techniques to study small biomolecules in solution will be covered. The section will end in 1.2.3 by focusing on the information that neutron diffraction, the technique to be used for this thesis, has obtained on amino acids and other biomolecules to date.

1.2.1 Examining Structure in the Liquid State

Structural information about biological molecules has been predominantly obtained by the use of *X*-ray crystallography. Crystallography exploits the tendency of molecules to form symmetric patterns and hold a given entity within an ordered structure. Such units are typically held within a lattice of unit cells, where a unit cell represent the small volume by which all the structural information to build up the macroscopic picture if translated (73). Any motion experienced is generally the result of mild thermal fluctuations. Being held within a lattice, it is possible to obtain repeating structural information over a long length-scale. The largest crystals are metres in length. However, for liquid structure, structure may only repeat up to a length of 1-2 nm (74; 75), which complicates the process of obtaining structural insight. Thus, the approach of a unit cell cannot be utilised.

In an experiment studying liquid structure, structural information can be obtained through a radial distribution function, $g(r)$. Radial distribution functions (RDFs) pro-

vide a measure of how the density of particles varies with respect to distance from a reference particle. Particles here represent atoms or molecules. The RDF is a useful way to describe the average structure in disordered systems (75).

The calculation of an RDF requires the construction of a series of concentric spheres around a reference particle, each sphere being separated by a distance Δr, as shown in Figure 1.18. The number of particles is then counted and the process repeated for other particles in the system. By determining the average number of particles in each shell, $n(r)$, it is possible to produce a histogram of the average number of atoms found in each volume segment with increasing distance. By normalising this by the product of the shell volume, $4\pi r^2 \Delta r$, and the particle number density, ρ, the final function is produced.

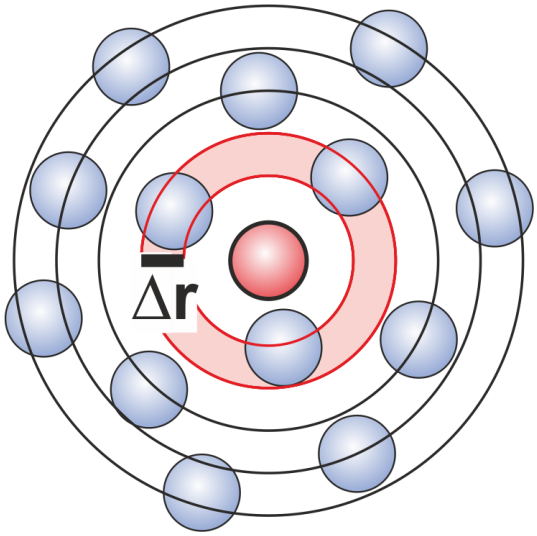


Figure 1.18: Calculating the radial distribution function, g(r), for a reference particle, which requires determining the number of particles in each sphere within a distance Δr.

This can be expressed as:

$$g(r) = \frac{n(r)}{4\pi\rho r^2 \Delta r} \tag{1.5}$$

The function is usually calculated for pairs of atoms and thus would describe the density of atom types β around atom type α, $g_{\alpha\beta}(r)$. Figure 1.19 shows g_{OwHw}(r) for pure water, portraying the density of water hydrogen atoms around water oxygen atoms.

Each peak is defined as a coordination shell, with the area underneath each peak related to the number of hydrogen atoms found in this shell around oxygen. The area shown in red defines the first coordination shell of the function. The peak at 1.8 Å indicated by the red dotted line, represents the most probable intermolecular hydrogen bond distance between water molecules.

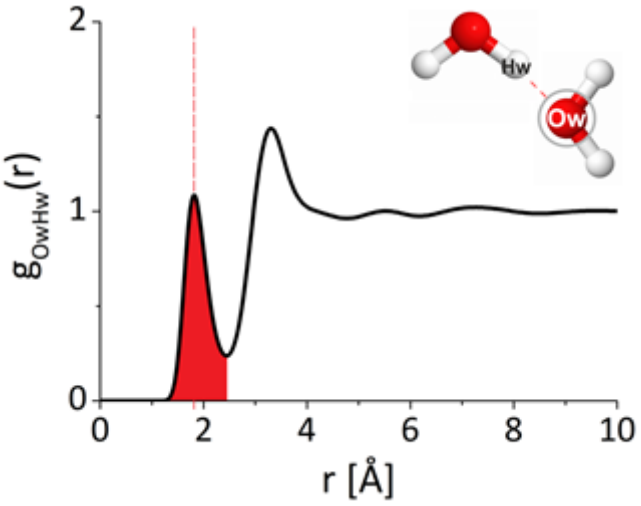


Figure 1.19: The intermolecular water oxygen (Ow) - water hydrogen (Hw) radial distribution function for pure water. The red dashed line signifies the most probable hydrogen bond length to find between the atoms. The peak in red represents the first coordination shell of the function, where the closest interactions between atoms take place. The area underneath is correlated to the number of Hw counted around Ow in the distance range of the shell.

Using a structural technique such as neutron or X-ray diffraction, it is possible to extract $g(r)$ from the data. This discussed in more detail in Section 2.1.3.

1.2.2 Determining Biological Structure in the Liquid Phase

Given the ability to obtain structural insight into a system in the liquid state, there are multiple techniques that can be used to extract this information. This subsection will introduce both experimental (Section 1.2.2.1) and computational (Section 1.2.2.2) methods that can be used to study small biomolecules in solution.

1.2.2.1 Experimental Techniques

A number of experimental techniques have been used to study the structural properties of small biomolecules. However, many of the methods used can only gain an indirect measure of atomic-level structure in solution. Spectroscopy methods such as Raman have been used to study small biomolecules in solution, including imino acids (69). Raman spectroscopy is sensitive to the vibrational and rotational modes of molecules (76), with signal arising from the inelastic scattering of photons, and thus indirectly probes structure. The resulting spectrum provides structural information on a system, appearing as bands at a known frequency, as well as on interactions between the probed molecule and its surroundings, manifested as shifts of the spectral bands. Atomic separations cannot usually be predicted without the use of complimentary simulations (Section 1.2.2.2) or advanced deconvolution techniques.

Terahertz spectroscopy utilises radiation in the far infra-red region, with frequencies in the region of 10^{-12} Hz. Here, organic and biological molecules exhibit strong absorption through vibrational and rotational modes. Terahertz spectroscopy has been able to provide indirect structural details of both water (77) and biological molecules (78). Studies on amino acids have provided hydration shell coordination numbers on a molecular level, but does not provide information on interatomic distance atomic pairwise coordination numbers (79).

X-ray absorption spectroscopy (XAS) is used to explore the electronic states of a system (80). When a *X*-ray strikes an atom, a transfer of energy can lead to the excitation or ejection of an electron, the latter known as the photo-electric effect which is measured by XAS. The energy needed to eject an electron determines the type of spectroscopy utilised, including Edge *X*-ray Absorption Fine Structure (EXAFS) and Near-Edge *X*-ray Absorption Fine Structure (NEXAFS). XAS is successful at qualitatively determining the location of interactions based on electronic information, as well as the conformation and flexibility of a molecule (81). However, no details on coordination numbers are usually retrieved. NEXAFS has been used to complete the first structural comparison between an amino and imino acid building block (82), which will be discussed in more detail in Chapters 4 and 5.

Nuclear Magnetic Resonance (NMR) spectroscopy monitors the influence of a magnetic field on nuclear spins (83). NMR spectroscopy is sensitive to nuclei with non-zero spins such a hydrogen, as well as isotopes of atoms found in organic molecules, including carbon (^{13}C), nitrogen (^{15}N) and potassium (^{31}P), making it a useful method to investigate peptide-based molecules. NMR spectra can be used to study both molecular structure and changes in the molecular environment. Diffusion NMR can also be used to determined self-diffusion coefficients of molecules in solution, as well as insight into molecular association (83). Quantitative details on aspects such as coordination numbers are not typically sought from NMR spectroscopy, though limited efforts have been made to extract information on the hydration shell around biomolecules (84). In the peptoid field, solution-phase studies of peptoid nonamers in acetronitrile have been completed to determine the intramolecular bonds present between residues (85).

In an applied electric field, the electronic spins of hydrogen atoms align with the field, producing a secondary field to oppose the applied one. If the applied field is adjusted until a group's resonance state is observed. The shift required is specific to the functional group. As the system is usually studied in deuterium oxide, it is only non-exchangeable protons that can be studied. A variety of NMR experiments can be completed that can measure the conformation and diffusion of molecules in solution, as well as insight into molecular association (83). Quantitative details on aspects such as coordination numbers are not typically sought from NMR spectroscopy, though limited efforts have been made to extract information on the hydration shell around biomolecules (84). In the peptoid field, solution-phase studies of peptoid nonamers in acetronitrile have been completed to determine the intramolecular bonds present between residues (85).

The quantitative information that has been missing from many of the experimental techniques above, including coordination numbers and atomic separations, is achieved using X-ray and neutron scattering. As this thesis discusses the results from neutron diffraction data, the theory behind diffraction methods will be discussed in detail in Chapter 2.

1.2.2.2 Computational Techniques

An *in silico* study aims to analyse structure and function by producing a computational model of the system. Computational modelling for biomolecular systems was pioneered in the 1960's and has continued to rise in popularity over the years (86), often complementing experimental measurements. The increase in the complexity of systems being studied via computational methods has been made possible by drastic improvements seen in computational performance over the decades (86). This development was recognised by the awarding of the 2013 Nobel Prize in Chemistry to Karplus, Levitt and Warshel, who laid the foundation for the computational methods used to understand and predict chemical processes.

A simulation provides a computational representation of the potential energy surface of the system being investigated, defined by covalently and non-covalently bonded parameters. The potential energy, U, of the system is given by the equation:

$$U = U_{Bond} + U_{Angle} + U_{Dihedral} + U_{LJ} + U_{Coulombic} \tag{1.6}$$

The first three terms in Equation 1.6 refer to intramolecular, covalently-bonded parameters, namely bond lengths (U_{Bond}), bond angles (U_{Angle}) and dihedral angles ($U_{Dihedral}$). The concept of a dihedral angle is introduced in Section 2.2.3.5 and considers rotational and out-of-plane movements. These are incorporated into a simulation by specifying the lengths and angles for a given molecule. A good estimation of these can be retrieved from quantum mechanical models such as AM1 (87) or from chemical software-derived models such as ghemical (88), often predicting conformation in the gas or solid phase. The final two terms are the intermolecular, non-covalently bonded parameters and correspond to the LJ (U_{LJ}) and Coulombic ($U_{Coulombic}$) potentials for each atom. The definitions of the LJ and Coulombic potentials are given in Section 1.1.2. Within a molecular simulation, force fields such as CHARMM (89), AMBER (90) and OPLS (91; 92) have been developed to calculate the values required for biological and chemical molecules. For water, standard models also exist, including SPC/E, SPC (93) and TIP models (94; 95).

There are multiple types of simulation commonly used to study biomolecular structure. In a Molecular dynamics (MD) simulation, Newton's laws of motion are solved to determine the time-dependent behaviour of molecules in the system (96). Monte Carlo (MC) simulations are used to model the effects of modifications to the system, including rotational and translational movements. A given modification is accepted if the move causes the potential energy of the system to be lowered (97; 98). The movements incorporated into MC simulations are not necessarily physical and thus dynamical properties cannot be accessed. Monte Carlo simulations are discussed in more detail in Section 2.2.1.5, being incorporated as part of the EPSR modelling system used for this thesis.

A simulation provides insight into a system at a molecular or even atomic-level, allowing for the prediction of properties and behaviour from the molecular structure. Such methods allow research to be completed that could never take place with experimental techniques, such as exploring systems under extreme conditions. Simulations are often completed to interpret experimental results and can help in improving the acquisition of such data.

The time taken to run a simulation is dependent on the complexity of the system. This can mean that to run a simulation can be computationally expensive and can introduce errors into the force field that leads to erroneous results (99). More sophisticated methods can be used to determine rare events and reduce the computational cost of running the simulation, including coarse-grained modelling, though such methods do not allow for structural information to be extracted at atomic level (96).

1.2.3 Neutron Diffraction and EPSR Modelling on Aqueous Biomolecules

For this thesis, neutron diffraction and EPSR modelling have been used to study biomolecules in aqueous solution. The theory and motivation behind using these methods is presented in Chapter 2. This combination of techniques is tailored to the research of liquids and amorphous materials.

The use of EPSR for elucidating the structural properties of biological molecules was reviewed by Soper in 2000 (100), with extensive studies being completed on an

array of organic molecules (101–118), salts (119–125) and water (126). Post-2006, the range of studies carried out on biomolecules diversified, with literature now existing for various amino acids (127–129), dipeptides (130; 131), sugars (132; 133), lipids (134–136), and neurotransmitters (137; 138). Whilst these methods are considered ideal for small biomolecules, as discussed further in Section 2.1.4.4, the size of the biomolecules investigated via this method continues to increase, with tripeptides (139–141) and micelles (142) now being studied. One of these tripeptides is glycyl-prolyl-glycinamide, a sequence also considered to have imino acid residues as shown in Figure 1.4. This sequence was found in solution to produce β-turns, formed by water molecules facilitating interactions between the residues.

In the study of amino acids and peptides, neutron diffraction and EPSR modelling have been crucial in highlighting the relative importance of hydrophilic interactions for biomolecular association, in comparison with hydrophobic contacts. In studies by McLain *et al* (131), the self-association for three dipeptides, glycyl-L-alanine, glycine-L-proline and L-alanyl-L-proline, was compared at a molecular ratio of 1 dipeptide to 29 water molecules. All of the dipeptides contain the same hydrophilic backbone groups, whilst their side chains are considered on scales typically as hydrophobic, though to varying degrees. With the side chains having different hydrophobicity levels, it was possible to analyse the effect of dipeptide hydrophobicity on self-association. The hydrophobicity pattern of the dipeptides was L-alanyl-L-proline > glycine-L-proline > glycyl-L-alanine. The work suggested that association was driven through interactions between the carboxylate and ammonium terminals groups, as opposed to the hydrophobic groups. For glycyl-L-alanine, clusters of up to 50 dipeptides were formed through association of hydrophilic groups, whilst the maximum size was only 15 for hydrophobic groups. Furthermore, it was elucidated that the more hydrophilic the dipeptide, the greater the association. Compared to the 50-mers found in glycyl-L-alanine solutions, the maximum cluster size through hydrophilic groups was 35 for glycine-L-proline and 25 for L-alanyl-L-proline. Whilst hydrophobic interactions have typically been thought to be the driving force in association and self-assembly of biomolecules, charged-based interactions were recognised as being the dominant force in the association of the building blocks of complex biomolecules. The role of aromatic tryptophan in proteins has

been indirectly elucidated using indole as a model system (118). This study highlighted the potential importance of X-H$\cdots\pi$ bonding in biological systems and interestingly also determined that the hydrocarbon pyrrole ring can coordinate water molecules.

The single, zwitterionic imino acids studied using these techniques include L-glutamic acid and L-proline. Capped amino acids have also been studied extensively (143–145), though in combination with MD simulations instead of EPSR modelling. In charged amino acids such as L-glutamic acid, the carboxylate and ammonium functional groups readily form hydrogen bonds with water, disrupting the tetrahedral network of water (127). Research completed on L-proline aimed to study the impact on the hydration from varying the biomolecular concentration, as well as elucidate the mechanisms by which it may act as a protecting osmolyte (128). At ratios of 1:10, 1:15 and 1:20 L-proline molecules to water molecules, there were subtle changes in hydration caused by increasing the concentration. Whilst the mechanisms for osmolytic function are dependent on the molecule, it has been suggested that the stacking of its pyrolidine rings accounts for its properties. The results suggested that there was a low probability of any stacking or clustering, with clusters with up to 8 molecules present at any given moment. The lack of any long-lived association was also supported by small angle neutron scattering (SANS) studies. Subsequent studies by Busch *et al* (146) determined that the L-proline clusters present may stack so as their hydrophilic and hydrophobic portions are aligned separately, showing the prevalence of both interactions types in solution.

1.3 Aims and Objectives

The aim of this thesis is to obtain a structural insight into a model imino acid building block in aqueous solution. For this, neutron diffraction and EPSR modelling will be used to obtain a quantitative understanding of the interactions these molecules make with its aqueous environment and how it associates. This thesis takes a bottom-up approach to the understanding of peptoid structures (Figure 1.20); by determining the behaviour of single building blocks, it may be possible to learn more about peptoid structures from a fundamental perspective.

By taking a bottom-up approach, this work can be considered a first step to understanding the properties of peptoid structures. One limitation of this method is the inability of account for emergent properties; cooperative properties that arise from the interaction of a population of elements but which are not characteristic of any single element, such as the folding behaviour of these constructs, which can only be probed when studying specific chains (147). Whilst it is possible to recognise interactions potentially common to both imino acids and peptoid chains of interest, it is not possible to fully extrapolate the folding behaviour from them. Nevertheless, this does not invalidate the present study. The bottom-up approach allows us to simplify the system to point where current state-of-art modelling methods and atomic-level physical characterisation can be done. This in turn opens the possibility to study the properties that could lead to the emergent behaviour of more complex systems, such as hydration levels and self-interaction propensity. Furthermore, the study of single imino acids building blocks is an infrequently explored area in research (82; 128), despite the recognised benefits of amino acid building blocks for many applications. Amino acids and their non-canonical variants have a role in areas such as metabolism and neurotransmission (5), with Section 1.1.4.5 highlighting already the use of related imino acid building blocks as osmolytes. It is thus interesting to consider how the properties of such building blocks could be enhanced to fulfil their function better, as well as explore novel molecules that could be beneficial for various applications.

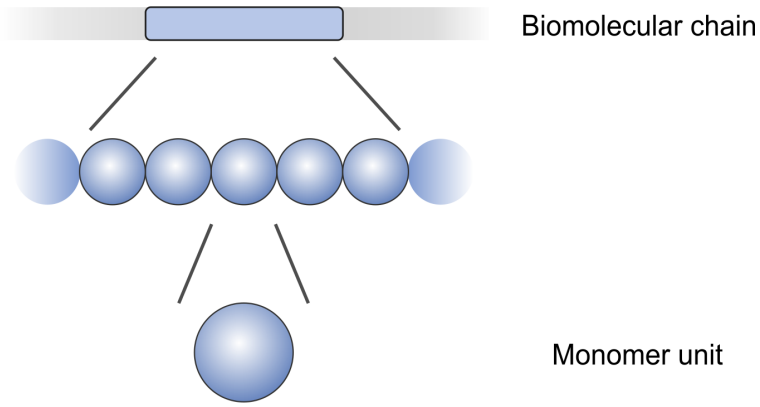


Figure 1.20: The bottom-up approach to understanding the properties of biomolecular structures. By studying the structure and interactions of single building blocks, it may be possible to determine reasons for the properties observed in complex chains.

This thesis will present structural studies on the hydration and association of three biomolecules in aqueous solution, the glutamine imino acid that is serving as the model imino acid, its amino acid equivalent L-glutamine and the imino acid sarcosine that has been subject of studies previously. The structures of these three molecules are shown in Figure 1.21 and the molecules are separately introduced in the following subsections.

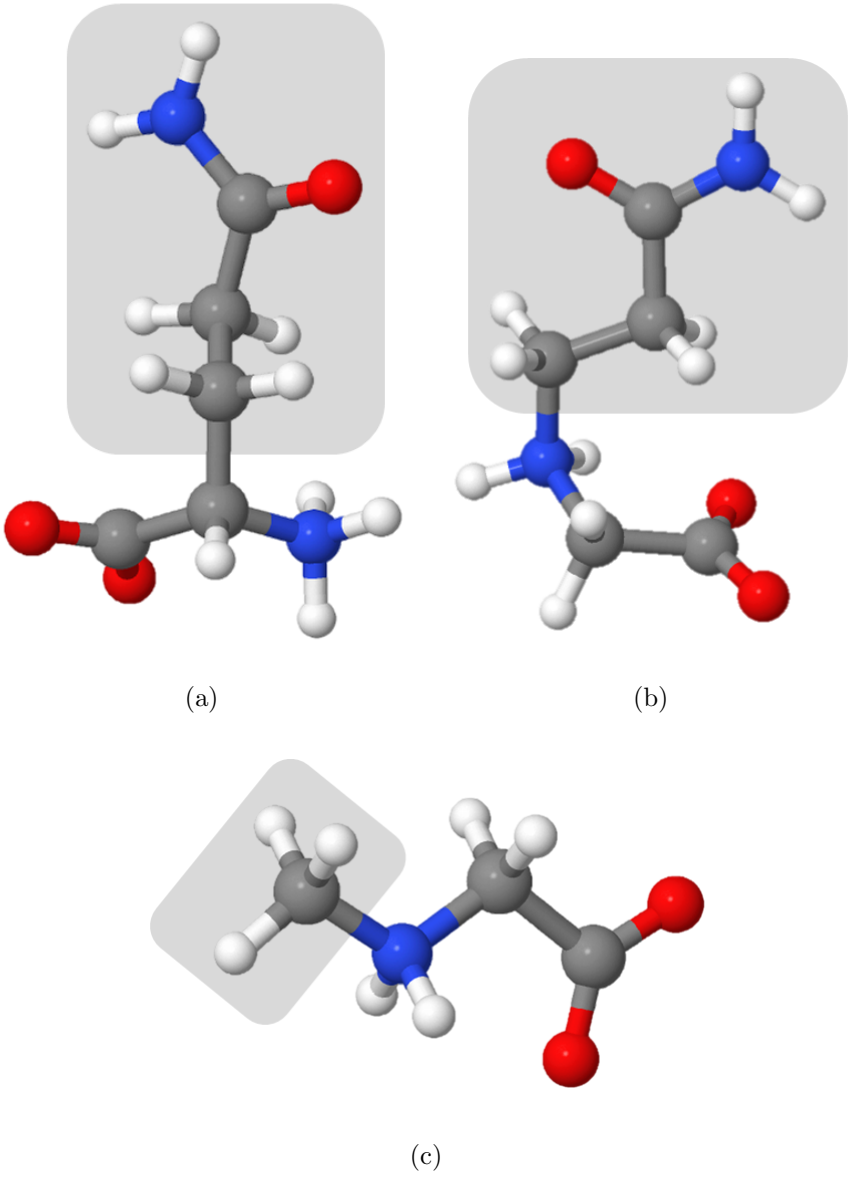


Figure 1.21: The biomolecules (a) L-glutamine, (b) the glutamine imino acid (Ngln) and (c) sarcosine that are being studied for this thesis. The side chain group for each is labelled in grey. Carbon atoms are grey, hydrogen atoms in white, nitrogen atoms in blue and oxygen atoms in red. C=O double bonds are not shown in these diagrams.

Chapter 3 present the structural results for L-glutamine and considerations needed for obtaining structural information on these molecules. Chapter 4 covers the properties of the glutamine imino acid, including a comparison with L-glutamine to determine if the structure causes the differing properties between the molecules. Chapter 5 compares the glutamine imino acid with sarcosine to determine how the side chain group alters imino acid interactions, concluding with a detailed description of how both imino acids impact on the structure of bulk water.

1.3.1 Glutamine

L-glutamine is a non-essential amino acid and is one of the most abundant found in the human body (148). It is one of few molecules able to cross the blood-brain barrier and is present in large quantities within our blood and tissues. L-glutamine is thought to be vital for proper digestion, brain function and immunity, and is often prescribed to combat various diseases to protect from stress (149). L-glutamine possesses a side chain of a primary amide attached to two methylenes groups (- CH_2 - CH_2 - $CONH_2$). The amide group renders the amino acid polar, hydrophilic and without net charge, but is also able to form hydrogen bonds.

Within proteins, glutamine is one of two residues that undergoes deamidation as part of the thermal denaturation process (150). The other residue is asparagine, which also contains a primary amide group. The process causes the amide group to be lost from the side chain, with it instead attaching to the backbone. This is one of the first processes thought to happen in thermal denaturation as a result of energy transfer between the aqueous environment and the amide group being facilitated (151). Nevertheless, favourable side chain-side chain interactions between amide-based residues are thought to play a role in the stabilisation of helical structures (152).

Glutamine has been implicated in having a role in neurodegenerative disease. Homopolypeptide chains of 5 or more glutamine residues can be found in 1.4% of all proteins and most commonly found in those responsible for DNA and RNA synthesis (153). Extensive research into polyglutamine chains has led them to be associated with nine neurodegenerative diseases, including Huntington's disease and spinobulbar

muscular atrophy (3). In Huntington's disease, specific areas such as the basal ganglia of the brain are damaged, causing movement difficulties, behavioural and cognitive changes; and eventual mortality (154). Polyglutamine chains originate in neurodegenerative disease from specific proteins, which are found in the neuronal cytoplasm. It is thought that a genetic mutation causes the polyglutamine chains in these proteins to increase in length. At a length of > 36 residues or more, these pathogenic chains have a tendency to collapse to form compact, mechanically stable structures (2; 155). These structures, though randomly collapsing, may reorder to form secondary-like structure folds, with the most commonly speculated structure for these chains including β-sheets (156–158), though other structures such as random coils (159), β-hairpins (160; 161) and β-helices (162) have been suggested. These aggregates cannot be digested by the ubiquitin-proteosome system (UPS), which usually digests mutated proteins (163). Instead polyglutamine is left to aggregate with other collapsed chains, once free from the rest of the protein. They form fibrils, micelles and plaques (164) within the cytoplasm, but will also migrate to the nucleus, to form nuclear inclusions (165). The plaques present are suspected to have a toxic effect, including the disruption of neurogenesis, the dysregulation of Ca^{2+} homeostasis, and damage to cell organelles (155). Force spectroscopy (166), small-angle neutron scattering (167) and computational (168) studies have also suggested that monomer polyglutamine chains will also readily collapse. It has been speculated the stability of these collapsed chains is attributable to an extensive hydrogen bonding network, where intramolecular-bonds outweigh interactions with the solvent (2; 166). Currently, there is a lack of experimental structural evidence to support these hypotheses.

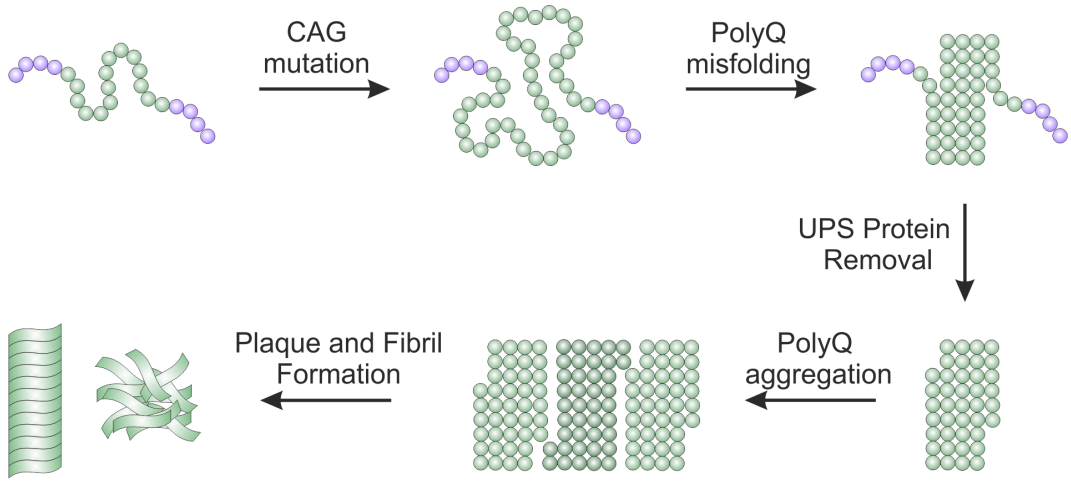


Figure 1.22: The aggregation of polyglutamine in disease. Whilst polyglutamine (PolyQ) sequences are common in proteins, a mutation can cause the length of these chains to aggregate. If over 36 residues, these chains are thought to collapse to form compact, stable structures. These structures cannot be digested by the ubiquitin-proteosome system (UPS) that removes mutated proteins, with these polyQ sections aggregating to form the fibrils and plaques observed in neurodegenerative disease.

The propensity to aggregate is seen also for L-glutamine. The solubility of L-glutamine is $42\,\mathrm{mg\,ml^{-1}}$ at $25\,^{\circ}\mathrm{C}$ (48; 169), lower than the limit of certain hydrophobic amino acids. Considering the nature of this molecule, including in the context of disease, it is interesting to explore the intermolecular interactions this amino acid forms in solution to elucidate what drives association.

1.3.2 Glutamine imino acid

The molecule that will feature most in this thesis is the glutamine imino acid. The chemical name for this imino acid is N-(2-carboxamidoethyl)glycine or N-(2-carbamylethyl)glycine. Its abbreviation is Ngln and this term will be used throughout this thesis. The synthesis method for this imino acid is presented in Appendix A.

Research on this molecule as a peptoid residue is limited. Discussion of this residue featured in work on forming peptoid superhelices by Murnen *et al* (44). Interestingly, when chains with the glutamic acid site chain (see Figure 1.4) were replaced with those of glutamine, superhelical structures were no longer formed. Such research has led to the

hypothesis that hydrogen bonding is not as important in the self-assembly of peptoid structures. Nevertheless, the Ngln residue has been used in producing antimicrobial peptoids (170) and neuropeptide mimics (171).

No research has been completed using the single Ngln imino acid. With not even a crystal structure characterising this substance, the work presented in this thesis is the first ever study to be completed to understand the structural properties of this molecule. This molecule acts as a clear example of the impact that relocating the side chain can have on the properties of an imino acid compared to its amino acid analogue. As part of the work for this thesis, this molecule was readily miscible in water at a concentration of $300\,\mathrm{mg\,ml^{-1}}$, approaching $\times 10$ the solubility of L-glutamine. Based on this difference alone, it is interesting to study this molecule to understand the conformation and interactions it makes. If compared to L-glutamine, it is also possible to determine structural reasons for Ngln's enhanced solubility.

1.3.3 Sarcosine

Sarcosine is the only naturally-occurring, pure imino acid. It has a side chain of a single methyl group (- CH_3) and is the imino acid version of the amino acid L-alanine. Whilst the three-letter notation for sarcosine could be Nala, it also has notation based on its colloquial name, Sar. Sarcosine is thought to be one of the basic molecules that existed on prebiotic Earth and is found in meteorite samples (172). It has been recognised as a possible biomarker for prostate cancer (173) and is an intermediate in the choline oxidation pathway (174). Sarcosine is recognised biochemically as an osmoprotectant, capable of inhibiting the effects of denaturants such as urea (61; 67; 175), preserving protein structure and preventing ice formation (64). Compared to L-alanine, which has a solubility of 166 mg ml of water at 25 °C (48), the solubility for sarcosine is higher at 481 mg ml (48).

As mentioned in Section 1.2.2.1, sarcosine and L-alanine were candidate molecules in the first structural comparison between monopeptide and monopeptoid (82), completed by Uejio *et al.* Using NEXAFS spectroscopy, they were able to determine that sarcosine more readily formed hydrogen bonds with water, which could potentially explain the

cause of its increased solubility and hydrophilicity. This study will be described in more detail in Chapter 4. However, the results of this investigation were qualitative and did not give details as to the exact number of bonds that this molecule forms with other molecules. Thus, the work presented in this thesis both determines the effect of the side chain on the interactions of imino acids and expands upon the Uejio work and other studies.

Chapter 2

Methods

Chapter 1 discussed the techniques utilised to study the properties of peptoids and small biomolecules. In the peptoid field, there has been heavy reliance on the use of indirect experimental methods, crystallography and computational modelling to probe structural properties (34–36), with limited information being obtained from direct, solution-based, experimental sources (70; 176). This chapter describes the theory behind neutron diffraction and EPSR modelling, the methods used for this thesis, which can be used to elucidate the structural behaviour of such biomolecules in solution.

The stages at which neutron diffraction and computational modelling are applied in the experimental procedure is summarised in the flowchart shown in Figure 2.1. A neutron diffraction experiment is first completed using biomolecular solutions to obtain structural information about the system. Neutron diffraction is used as it is particularly sensitive to the location of hydrogen atoms (Stage 1). The theory and experimental procedure for a neutron diffraction experiment is described in Section 2.1. This section will cover the theory of diffraction (Section 2.1.1), the motivation behind using neutrons (Section 2.1.2), the structural information that can be obtained (Section 2.1.3) and how the experiment is completed (Section 2.1.4). To extract the atomic-level detail, it is necessary to initially create a computational model of the system (Stage 2) and then drive the model to fit the experimental data (Stage 3). The initial model, known as the reference potential, is described in Section 2.2.1, and the process of driving the simulation to fit the experimental data is covered in Section 2.2.2. Auxiliary routines

that can be implemented in the EPSR method to obtain extra quantitative information are discussed in Section 2.2.3. The outcome of following this procedure is atomic-level detail of the structure of a system (Stage 4). The results discussed in subsequent chapters are based on the information obtained at Stage 4.

To supplement select neutron diffraction and EPSR studies, small-angle neutron scattering (SANS) studies have also been completed to further understand biomolecular association in aqueous solution (Supplementary Methods). The SANS-specific theory and experimental details are detailed in Section 2.3.

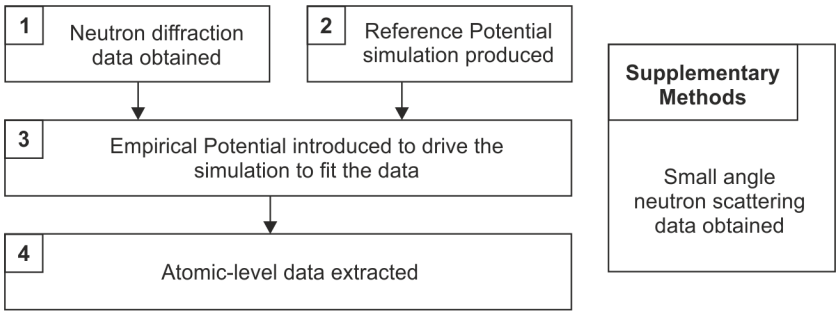


Figure 2.1: Flowchart of the core components for obtaining atomic-level information (Stage 4) from neutron diffraction data and EPSR modelling. Upon completion of the neutron diffraction experiment (Stage 1), a reference model with computationally-derived parameters is made up of the same molar ratio and atomic number density as the system (Stage 2). The data and simulation are then compared, allowing an empirical potential to be introduced that drives the simulations to fit the data (Stage 3). For this thesis, SANS data has also been obtained to support certain results determined at Stage 4 (Supplementary Methods).

2.1 Neutron Diffraction

Scattering is a process by which incoming radiation is deflected or diffused from its linear trajectory due to non-uniformities in the medium through which it transverses. How radiation of a given kind scatters can provide specific information about a system. This thesis focuses on the use of diffraction to understand aqueous solutions. Diffraction occurs due to elastic, coherent scattering of radiation and is sensitive to the structure of a system. This section aims to cover the key theory that governs the diffraction

phenomenon. It is not possible to cover all of the aspects of diffraction and scattering in detail. However, many references are available that are dedicated to explaining further scientific details of these processes, including references (73–75).

2.1.1 Theory of Diffraction

2.1.1.1 Diffraction as a Probe of Structure

A key part of diffraction is the process of interference. This takes place when waves within the same spatial region interact with each other (177). These interactions are governed by the Principle of Superposition. For this phenomenon, the amplitude of the resultant wave is the sum of the amplitudes of the waves that were superposed. The two principle scenarios observed for the superposition phenomenon are shown in Figure 2.2.

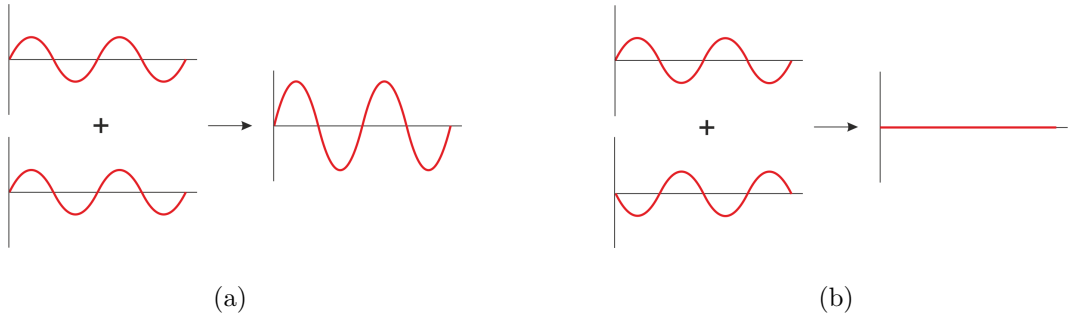


Figure 2.2: The Principle of Superposition. (a) If two waves arrive in phase, they will superpose to produce a wave that has an amplitude equalling the sum of the amplitudes of the original two; this is constructive interference. (b) If two waves arrive out of phase, then their amplitudes will cancel to produce no wave; this is destructive interference. It is necessary for waves to interfere constructively in order to produce a diffraction pattern.

The resulting amplitude for each scenario is dependent on the phase of the waves. If two waves meet that are in phase, the resultant wave will have an amplitude equalling the sum of the amplitudes of the original ones. This is known as constructive interference (Figure 2.2(a)). However, if the waves are exactly out of phase, then the waves interfere destructively and cancel (Figure 2.2(b)). Both interference processes can be described

by the following equations:

$$\begin{aligned} &\text{Constructive Interference} && n\lambda = X \\ &\text{Destructive Interference} && n\lambda = X + \lambda/2 \end{aligned} \tag{2.1}$$

In these relations, λ is the wavelength, n an integer and X a derivable quantity. For a diffraction experiment, the aim is to produce a diffraction pattern, with constructive and destructive interference required to produce this. It is possible to derive the quantity for X by considering the diffraction of waves through a plane.

Figure 2.3 shows the diffraction of two waves from planes of atoms. X equates to the phase or path difference between the waves interacting with each plane, which is shown in red. Half of this path length can be found via trigonometry. Deriving this distance ($dsin\theta$) and multiplying it by two produces the following equation:

$$n\lambda = 2dsin\theta \tag{2.2}$$

Here, d is the distance between the planes and θ the scattering angle. This equation is known as Bragg's Law and highlights that diffraction can be used to find distances within a system (178). This relation holds for the case of Fraunhofer diffraction (179), where the distance between the diffraction and resultant pattern is much greater than d and λ, rendering the radiation as travelling near parallel.

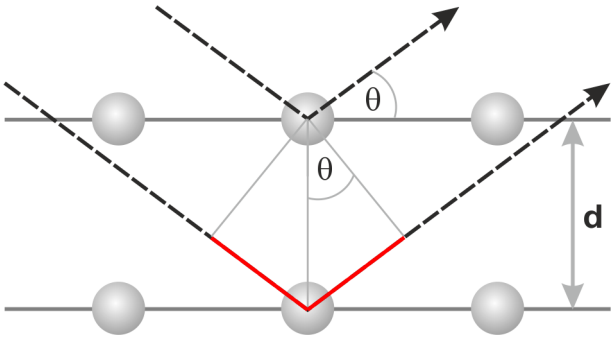


Figure 2.3: Bragg diffraction of waves from a plane. Bragg's Law allows information on the structure of a sample to be obtained from diffraction experiments.

A diffraction experiment aims to measure the angles at which constructive interference of waves with a known wavelength takes place. In a diffraction experiment, the angle measured is 2θ, which is the difference in angle between the incident and scattered waves. The largest angle possible for scattering is π radians, or 180°, which is equivalent to the wave being scattered in the direction from which it originated. Consequently, it is possible to complete an experiment to a θ angle of π radians. Diffraction instruments are typically categorised as either being small-angle or wide-angle, according to the angular range they probe. Equation 2.2 signifies that d and θ are inversely proportional to one another. Therefore, if small angles are studied, then information on large-scale structures can be obtained, whilst a wide-angle diffractometer can span a much larger range of angles and probe very short distances. Another requirement of the diffraction experiment is that the wavelength is known. The importance of the wavelength is discussed in the next section.

2.1.1.2 The Wave-like Nature of Matter

Equation 2.2 stated that it is possible to determine the length scales within a system assuming the wavelength of the probe is known. This means that there is a requirement for a diffraction probe to behave like a wave so as it can undergo interference.

As well as electromagnetic radiation, it is possible for particles to behave as waves and have wave-like properties. This principle is known as wave-particle duality, where in return waves can also act as particles. A particle can be thought of as a matter wave, which has oscillations characteristic of its wave-like nature, whilst the envelope that the wave spans is due to its particle like nature (Figure 2.4). The wave envelope will have waves of a wavelength λ and be of length Δ. The particle can be found somewhere within this envelope.

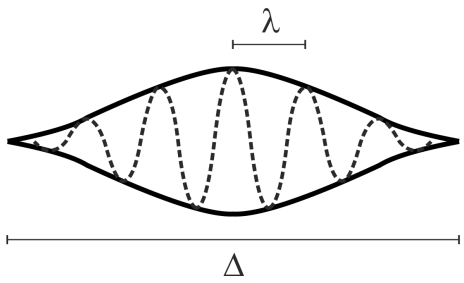


Figure 2.4: The matter wave, which possess oscillations of wavelength λ due to its wave-like nature and spans an envelope of length Δ because of its particle-like nature.

A kinematic understanding of the wave-particle duality phenomenon was determined by de Broglie (180):

$$\lambda = \frac{h}{p} = \frac{h}{mv} \tag{2.3}$$

The de Broglie relation relates the wavelengths of any wave, electromagnetic or matter, to its momentum, p, and Planck's constant, h. The momentum is the product of the mass, m and velocity, v. The equation signifies that any radiation with a momentum, even those considered particles, has wave-like nature and can interfere.

It should be noted that when the separation between components in a sample to be studied is much greater than the wavelength of the diffraction probe, it will follow a classical behaviour and no sign of interference will be present. If the separation and wavelength become comparable in distance, this is the quantum world, where interference can take place. To see an interference pattern, the wavelength must be tailored to be comparable to the size of the sample.

2.1.1.3 Elastic Scattering

Another property that is important for diffraction is the difference in energy between the incident and scattered waves of the probe. For diffraction, it is necessary that the waves scatter elastically, where the energy of the incident and scattered waves is the same. If any energy were transferred, this would be inelastic scattering. If the energy

differed in a diffraction experiment, this would affect the wave properties of the probe. This would affect the ability of the probe to produce a diffraction pattern.

The incident and scattered waves can also be described in terms of wavevectors. The wavevector, k, is defined as:

$$k = \frac{2\pi}{\lambda} \tag{2.4}$$

Equation 2.3 showed that energy is related to the wavelength, as energy is proportional to the momentum. If the wavelength does not change, neither can the magnitude of the wavevector. For elastic scattering, the modulus of the incident (k_i) and scattered (k_f) wavevectors remains the same:

$$|k_i| = |k_f| \tag{2.5}$$

2.1.1.4 Reciprocal Space

The wavevectors involved in a scattering experiment are shown schematically in Figure 2.5. The value Q in this figure is the exchanged wavevector and a measure of the difference between k_i and k_f. Q is a quantity used regularly in diffraction experiments and we can extend this diagram further to understand what Q represents.

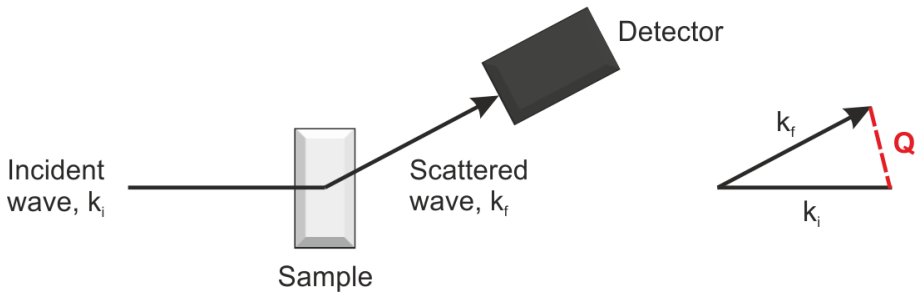


Figure 2.5: Diagram showing the scattering of waves in a diffraction experiment. Q is the exchanged wavevector; the difference between the incident k_i and scattered k_f wavevectors.

Figure 2.6 shows an Ewald sphere. This geometric construction was first developed by Paul Peter Ewald (181) as a way to prove Bragg's Law (Equation 2.2) and predict diffraction patterns. The construction requires drawing a sphere around the crystal and using trigonometry to determine the length of Q. Given the angle between k_i and k_f is 2θ and both wavevectors are both $2\pi/\lambda$:

$$\begin{aligned} Q &= k_i - k_f \\ &= 2(\frac{2\pi sin\theta}{\lambda}) = \frac{4\pi sin\theta}{\lambda} \end{aligned} \tag{2.6}$$

Incorporating the definition of the distance, d, from Bragg's Law[1] (Equation 2.2):

$$Q = \frac{2\pi}{d} \tag{2.7}$$

This equation shows that Q is inversely-related to d. In experiments, Q is used to represent the distance in reciprocal space. Due to the length-scales probed in diffraction experiments, on the order of $10^{-9} - 10^{-10}$ m, it is extremely difficult to extract these in real-space. Obtaining results in reciprocal space is thus a convenient way to extract information on these small length-scales.

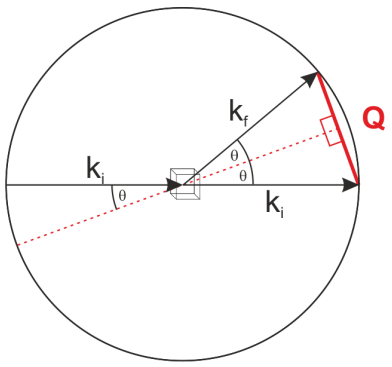
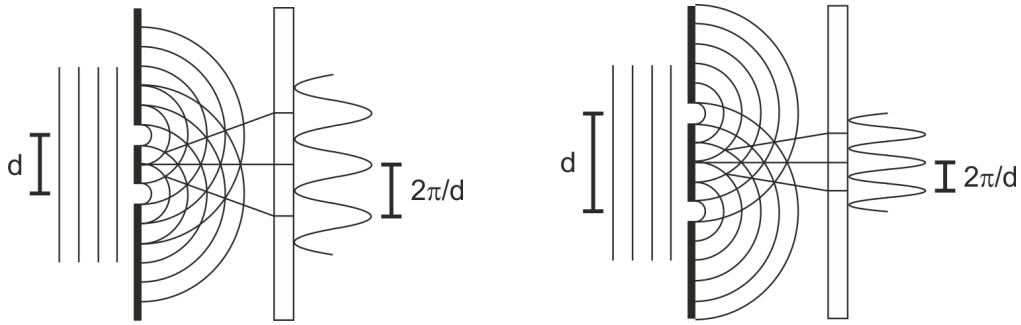


Figure 2.6: The Ewald sphere, a geometrical construction that demonstrates the relationship between the wavevectors of incident and scattered waves, the angle of incidence and the system in reciprocal space.

[1]For the results presented in this thesis, any distance in real-space will referred to as r, instead of d

A practical example showing this phenomenon is 'Young's double-slit experiment'. This experiment was first performed by Thomas Young (182) and was pivotal in recognising light as a wave. A schematic of this experiment is shown in Figure 2.7.

In the experiment, light is passed through two slits. The emerging waves from each slit are able to interfere, with the diffraction pattern appearing on a screen. When the slit spacing decreases (Figure 2.7(a)) the distance between the diffraction peaks increases, whilst if the spacing is increased (Figure 2.7(b)) the distance between the diffraction peaks decreases, proving the inverse relation of the diffraction pattern to the slit spacing. Furthermore, the distance between the diffraction peaks equates to Q, as described in Equation 2.7, signifying that the diffraction is a picture of the system in reciprocal space.



(a) (b)

Figure 2.7: Young's double-slit experiment. The experiment is important in showing the reciprocal nature of the diffraction pattern produced when scattering light through two slits. When the slit spacing is small, as in (a), the distance between the peaks is large, whilst for (b) a large spacing decreases the distance between constructive peaks.

The amount of light that passes through the slits at a given position, x, can be described by the Aperture function, $A(x)$. For Young's double-slit experiment, if it is assumed that the slits have a separation of $d/2$ either side of the centre, x represents the linear position along the screen, and the δ-function is utilised, *A(x)* is defined as:

$$A(x) = \delta\left(x - \frac{d}{2}\right) + \delta\left(x + \frac{d}{2}\right) \tag{2.8}$$

The final diffraction pattern is the summation of all the waves coming from the slits, as dictated by the principle of superposition (see Section 2.1.1.1). The amplitude of wave between x and $x + \Delta$ is proportional to $A(x)\Delta x$. The phase, ϕ, is dependent on x, the angle of propagation, θ, and the time t, with the phase of this wave contribution being:

$$\Delta\phi = \left(\frac{2\pi}{\lambda}\right) xsin\theta \tag{2.9}$$

Where $xsin\theta$ is the associated path length. The complex wavefunction for this contribution is:

$$\Delta\psi = \psi_0 \int_{-\infty}^{\infty} A(x)e^{iqx}dx \tag{2.10}$$

ψ_0 is a constant of proportionality, taking into account any time-dependent variation, whilst $q = 2\pi sin\theta/\lambda$. The diffracted wave, $\psi(q)$, is the sum of wave contributions and takes the form:

$$\psi(q) = \psi_0 \int_{-\infty}^{\infty} A(x)e^{iqx}dx \tag{2.11}$$

Equation 2.11 is the Fourier Transform of the Aperture function. The Fourier Transform can be used to convert between real and reciprocal space. For Young's double-slits, the Fourier Transform of the Aperture function can be expressed as:

$$\begin{aligned} \psi(q) &= \psi_0 \left(e^{iqd/2} + e^{-iqd/2}\right) \\ &= \psi_0\, 2cos\left(\frac{qd}{2}\right) \end{aligned} \tag{2.12}$$

For the observed diffraction pattern, this does not in fact equal $\psi(q)$ but the modulus squared of the complex function or the intensity, $I(q)$. For Young's double-slits, this is:

$$\begin{aligned} I(q) &= |\psi(q)|^2 = \psi(Q)\psi(Q)^* \\ &= I_0\, cos^2\left(\frac{qd}{2}\right) \end{aligned} \tag{2.13}$$

The resulting pattern thus is in the form of equally-spaced cosine fringes, of uniform intensity. Overall, what is shown is that the length-scales used for a function in real-space and its Fourier Transform are inversely related, making reciprocal space the term used for information extracted from the Fourier domain.

2.1.2 The Neutron as a Diffraction Probe

In Section 2.1.1.2, it was established that particles have a wave-like nature and can undergo interference. It is thus possible to use a variety of probes for diffraction experiments, ensuring the wavelength is suitable for the length scale being studied. The waves that are regularly used for diffraction include electromagnetic radiation (e.g. X-rays) and matter waves (e.g. neutrons and electrons). This section shall explain the motivation behind using neutrons as the probe for diffraction experiments of the small biomolecules studied for this thesis.

2.1.2.1 The Neutron

The neutron is a subatomic particle found in every known element, with the exception of hydrogen. Neutrons, along with protons, constitute the nucleus. Supporting the nucleus is the strong nuclear force, which operates to hold all the subatomic particles together in one unit. The neutron is classed in particle physics as a subatomic baryon, a particle made up of three subatomic quarks (177). The neutron is made up of one up and two down quarks. The up quark has a charge of $+2e/3$ and the down quarks a charge of $-e/3$. This renders the neutron electrically neutral.

The neutron was first discovered by James Chadwick in 1932 (183; 184). The reaction that led that led to Chadwick's discovery was:

$$^{9}Be +^{4} He \rightarrow \ ^{13}C \rightarrow \ ^{12}C + \ ^{1}n \tag{2.14}$$

By reacting Beryllium (Be) with alpha particles (^{4}He) sourced from the decay of Pollonium ^{210}Po, it was possible to produce single neutrons. This discovery was key to understanding the structure of the nucleus.

Table 2.1 lists some of the properties of neutrons, as well as X-rays; a commonly used diffraction probe. Neutrons are highly penetrating particles, due to their lack of charge, but interact weakly with matter and thus do not cause damage. It is also possible to exploit the spin of the neutron to undertake magnetism studies. The following subsections will focus on some of the key properties that make the neutron a suitable probe for the study of aqueous biomolecules.

Table 2.1: Properties of neutrons and X-rays.

Property	Neutrons	X-rays
Charge	0	0
Mass	$1.675 \text{ x } 10^{-75}$ kg	0
Spin	$\frac{1}{2}$	1
Magnetic Moment	$1.913\,\mu_N$	0
Energy	$\frac{1}{2}\mathrm{m}v^2 = \frac{\hbar^2 k^2}{2m}$	$\hbar\omega = \frac{hc}{\lambda}$

2.1.2.2 The Atomic Form Factor

For a diffraction experiment, the most appropriate probe to use will be dependent on the structural features that are of most interest in the system. For the work in this thesis, the interest is in the hydrogen bonds that the biomolecules form with their surrounding aqueous environment and when they self-associate. Therefore, the probe needs to be sensitive to hydrogen. What components a given diffraction probe will be sensitive to is governed by the way it interacts with the sample.

In Section 2.1.2.1, the properties of X-rays were compared. In the case of X-rays, they are scattered by the electron clouds of atoms in a sample. The scattering can be described by the atomic form factor, $f(\lambda, \theta)$, which describes the likelihood of a probe of a given wavelength being scattered in a given direction. For X-rays, $f(\lambda, \theta)$ is:

$$f(\lambda, \theta) = Zg(Q)r_e \tag{2.15}$$

Here, *g(Q)* is a factor dependent on the scattered wavevector, Q, and r_e is the classical radius of the electron (2.8 x 10^{-15} m). Furthermore, Z is the atomic number; the number of protons in the nucleus and the number of electrons within the cloud of an element, assuming the atom is neutral. The result of Equation 2.15 is that the atomic form factor is proportional to the number of electrons, as this directly affects the size of the electron cloud. In analyses of X-ray data, the form factor has to be accounted for. As will be detailed further in Section 2.1.2.3, this makes X-rays more sensitive to elements with high-atomic numbers. This makes X-rays unsuitable for studying hydrogen.

When a neutron interacts with atoms in a sample, it it is scattered by the atom's nucleus. This is possible through the strong nuclear force; the same force which holds together the components of the nucleus. The force operates at small length scales compared to the wavelength of the neutron. The nuclear force operates at length scales of 10^{-15} m (177), whilst the wavelength of neutrons in a diffraction experiment is around an Ångstrom, 10^{-10} m. Therefore, the nucleus effectively acts as a point source for neutrons to interact with. When neutrons are scattered, a spherical wave is produced (Figure 2.8). The scattering is uniform around the nucleus, unlike for X-rays, where more of the wave propagates in the forward direction. Whilst the scattering is uniform, over multiple nuclei, the net direction of the wavefront will be k_f. For neutrons, $f(\lambda, \theta)$ is a constant:

$$f(\lambda, \theta) = -b \tag{2.16}$$

The constant b equals the scattering length of the atomic nucleus. The precise value of b is dependent on the element, isotope and spin of the atom, and will be discussed further in Section 2.1.2.3. The negative sign in Equation 2.16re in phase with one another. Compared to Equation 2.15, the form factor is not dependent on Z, making neutrons more advantageous for studying hydrogen atoms in a system.

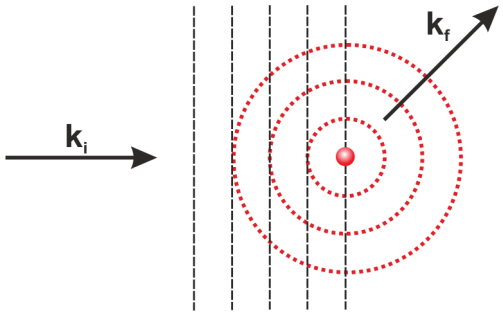


Figure 2.8: Schematic of neutron scattering from a single nucleus. When a neutron interacts with a nucleus, the resultant scattering is uniform around the nucleus. Over many nuclei though, the net scattering will tend towards k_f.

As will be described in Section 2.3.1, the neutron form factor is more complex when completing experiments on larger structures, such as in small angle neutron scattering (SANS) experiments, as the shape and size of the sample has to be accounted for. Therefore, the form factor has to be considered when analysing such data.

2.1.2.3 The Scattering Cross-Section

A neutron scattering experiment aims to measure the proportion of neutrons that are scattered in a given direction. This quantity is known as the differential scattering cross-section, $d\sigma/d\Omega$, which is given by the equation:

$$\frac{d\sigma}{d\Omega} = \frac{R}{N\Phi\Delta\Omega} \tag{2.17}$$

This equation corresponds to the ratio of the rate, R, that neutrons are scattered through a solid angle element, $d\Omega$, over the flux, Φ; the number of neutrons per unit area of beam. As the equation typically represents the average differential cross-section

per atom or molecule, the number of scattering units, N, is included to normalise the answer.

The term Ω represents the solid angle, a 3-D analogue of an angle, describing the area, A of spherical fragment produced when the radius, r, moves in 2-D:

$$\Omega = \frac{A}{r^2} \tag{2.18}$$

The solid angle is a representation that considers all angles and thus equates to the entire surface of a sphere. For a spherical fragment, $d\Omega$, covering a small area unit, dA, the equation is:

$$d\Omega = \frac{dA}{r^2} \tag{2.19}$$

This can be visualised in Figure 2.9. The units for Ω are steradians (sr).

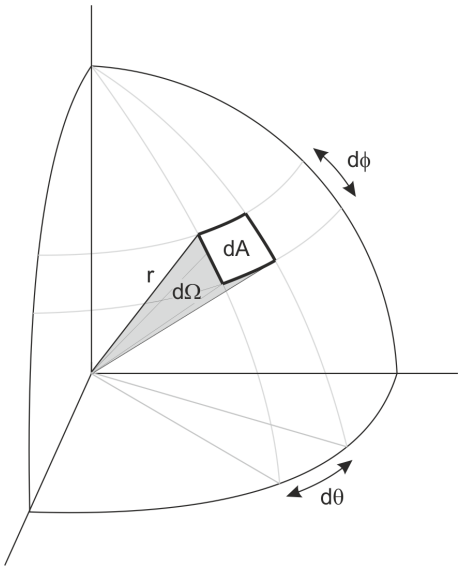


Figure 2.9: Schematic of the solid angle, Ω. An infinitesimal unit of the solid angle, $d\Omega$, is the ratio of the unit area of the sphere covered when the radius moves in 2-D space.

Integrating the differential scattering cross-section with respect to $d\Omega$ gives the total scattering cross-section, σ. Whilst a point nucleus may not have the disadvantage of

scattering over atom-dependent form factor, scattering is still said to happen over a cross-sectional region. Through mathematical analyses specified in neutron scattering texts, it can be determined that $d\sigma/d\Omega$ is the square of the form factor, $f(\lambda, \theta)$:

$$\frac{d\sigma}{d\Omega} = |f(\lambda, \theta)|^2 = b^2 \tag{2.20}$$

As Ω for the sphere is 4π, σ is:

$$\sigma = 4\pi b^2 \tag{2.21}$$

The scattering length, b, was first introduced in Section 2.1.2.3. Both b and σ are characteristic of the atom and isotope. The greater value of σ, the greater the potential sensitivity of the atom to be detected in an experiment.

The total neutrons and X-rays scattering cross-sections for elements typically found in biological molecules are shown in Figure 2.10. These cross-sections can be found in tables by McMaster (185) for X-rays and Sears (186) for neutrons. For X-rays, the cross-section increases proportionally to Z. As discussed earlier in Section 2.1.2.2, this makes X-rays more sensitive to scattering from larger atoms like carbon, oxygen and nitrogen. To extend this point, Figure 2.10 shows that that the scattering for hydrogen and its isotope deuterium, ^{2}H, is larger for neutrons than for X-rays. This makes neutrons a more sensitive probe of these lighter atoms and supports their use be able to resolve the position of hydrogen atoms in a given system.

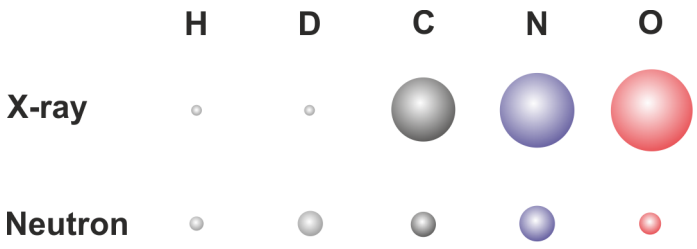


Figure 2.10: Schematic of the neutron and X-ray scattering cross-sections for hydrogen (H), deuterium (D), carbon (C), nitrogen (N) and oxygen (O).

2.1.2.4 Coherency

Section 2.1.2.3 described the relation between the scattering length, b, and the total scattering cross-section, σ. As the total, it is the summation of all scattering contributions, including coherent and incoherent. When the scattering is coherent, there is a phase relationship between the scattered waves that allows them to interfere constructively and destructively. This was described previously in Section 2.1.1.1 with the phase relationship given by Bragg's Law (Equation 2.1), with scattering that does not follow this being incoherent. Continuing on from Section 2.1.2.3 and considering any deviance from the phase relationship, it is possible to consider b to be an average length with a standard deviation:

$$b =< b > \pm \Delta b \tag{2.22}$$

Where the variance is:

$$< b^2 >=< b >^2 +(\Delta b)^2 \tag{2.23}$$

This means that average scattering cross-section can be written as the sum of two contributions:

$$\begin{aligned} < \sigma > &= 4\pi < b^2 > \\ &= 4\pi < b >^2 +4\pi(\Delta b)^2 \\ &= \sigma_{coh} + \sigma_{inc} \end{aligned} \tag{2.24}$$

Here, the two contributions are the coherent, σ_{coh}, and incoherent, σ_{inc}, scattering cross-sections. It is also possible to segregate $d\sigma/d\Omega$ into these components. The detailed mathematics to do this is beyond the scope of this thesis and there are various texts that discuss this, including (74). Nevertheless, from the incident, Ψ_i, and scattered, Ψ_f,

wavefunctions, it is possible to describe $d\sigma/d\Omega$ as:

$$\begin{aligned}\frac{d\sigma}{d\Omega} = b^2 &= \frac{|\psi_f|^2 r^2}{|\psi_i|^2} \\ &= |\sum_n b_n e^{iQ.r_n}|^2 \\ &= \sum_n \sum_m b_n b_m e^{iQ.(r_n - r_m)} \\ &= \overline{b}^2 \sum_n \sum_{m \neq n} e^{iQ.(r_n - r_m)} + \overline{b^2} \\ &= \overline{b}^2 \sum_n \sum_m e^{iQ.(r_n - r_m)} + (\overline{b^2} - \overline{b}^2) \\ &= \left(\frac{d\sigma}{d\Omega}\right)_{coh} + \left(\frac{d\sigma}{d\Omega}\right)_{inc}\end{aligned} \tag{2.25}$$

Here, n and m refer to two uncorrelated nuclei, whilst r is the position of a given nucleus, specified by the subscript. The first part of the equation is the coherent scattering component, $(d\sigma/d\Omega)_{coh}$, which describes correlations between nuclei and provides structural information on a system. Coherent scattering can also give insight into the collective dynamics of nuclei. The second part corresponds to the incoherent scattering, $(d\sigma/d\Omega)_{inc}$, of neutrons that contains no structural information, describing only the dynamics of individual particles. Inelastic scattering is a phenomenon useful for spectroscopy studies, but for elastic scattering, the incoherent contribution manifests itself as a flat background. In Section 2.26, the steps taken to obtain structural information from this coherent scattering will be covered. For a given atom and isotope, it will have specific coherent and incoherent scattering cross-sections.

Table 2.2: Coherent and incoherent scattering cross-sections and coherent scattering lengths for isotopes of hydrogen.

Atom	σ_{coh} (barns)	σ_{inc} (barns)	b_{coh} (fm)
Hydrogen	1.8	80.3	-3.74
Deuterium	5.6	2.1	6.67

The coherent and incoherent scattering cross-sections for isotopes of hydrogen are shown in Table 2.2 and have been taken from the Sears tables (186). These values indicate that hydrogen, despite being more efficient at coherently scattering neutrons than X-rays (see Section 2.1.2.3) is in fact a very strong incoherent scatterer. For a diffraction experiment, it is thus necessary to correct for this (see Section 2.1.4.7), as incoherent scattering increases noise/signal ratio. To overcome this, it is possible to substitute hydrogen atoms for deuterium in the system, which is a more efficient coherent scatterer. Referring to the respective values of b for each isotope in Table 2.2, it can also be seen that the scattering length is not only a different value but also a different direction. The result is that the signal produced for each isotope is different. If multiple experiments are completed, with different isotopic substitutions, it is possible to label species by highlighting the contrast or hide certain groups by matching the contrast between proton and deuterium, therefore locating the position of the hydrogen atoms in a system.

2.1.2.5 Important Considerations when using Neutrons

At the end of Section 2.1.2.4, it was established that hydrogen was a strong incoherent scatterer of neutrons. The issue can be overcome by substituting hydrogen atoms for deuterium, with isotopic experiments contrast experiments helping to highlight interesting features within the group. The caveat of this is that the hydrogen bonding in the samples containing deuterium may differ compared to a fully protiated version, and over multiple contrast experiments the results would be the average of all the differences in the sample.

The other consideration for a neutron experiment is the scattering cross section for atoms. It was shown that the cross-section for X-rays was proportionally related to the atomic number (Section 2.1.2.3). However, for neutrons, the signal from atoms does not alter drastically with increasing atomic number and their cross-sectional value is lower compared to X-rays, as shown in Figure 2.10. Biological molecules typically possess atoms with low mass numbers, with the lower signal intensity meaning it is necessary for higher solution concentrations and larger samples to be used for any neutron experiment.

For this thesis, the amino acid L-glutamine has been studied, which has an inherently low solubility and thus low concentrations were used. Chapter 3 will put forward not only the results achievable for this system, but also the limitations incurred by using a low concentration for a neutron experiment.

2.1.3 Extracting Structural Information from Neutron Diffraction Measurements

In Section 2.1.2.3, it was stated that the quantity measured by a neutron experiment was the total scattering cross-section, $d\sigma/d\Omega$. This value was given by Equation 2.25. For the elastic, coherent component of $d\sigma/d\Omega$, it is possible to divide this into two components; the self and distinct scattering contributions:

$$\begin{aligned}(\frac{d\sigma}{d\Omega}) &= (\frac{d\sigma}{d\Omega})_{self} + (\frac{d\sigma}{d\Omega})_{dist} \\ &= \sum_{\alpha} c_\alpha b_\alpha^2 + \sum_{\alpha \leq \beta} (2 - \delta_{\alpha\beta}) c_\alpha c_\beta b_\alpha b_\beta S_{\alpha\beta}(Q) - 1\end{aligned} \tag{2.26}$$

The detailed mathematics to derive this expression can be found in references (74) and (75). Here, c is the atomic fraction of a given component and b the scattering length. The Kronecker delta function, $\delta_{\alpha\beta}$, is also included to prevent double counting. The 'self' contribution describes the scattering from single nuclei. It contains no structural information and only provides a featureless signal. The 'distinct' part contains information on the scattering between pairs of atoms. The distinct component is proportional to the total structure factor, *F(Q)*, which dictates how radiation interacts with the sample and from Equation 2.26 can thus be described as;

$$F(Q) = \sum_{\alpha \leq \beta} (2 - \delta_{\alpha\beta}) c_\alpha c_\beta b_\alpha b_\beta S_{\alpha\beta}(Q) - 1 \tag{2.27}$$

F(Q) summates all partial structure factors in the system. The partial structure factor for atomic components α and β is described by the equation;

$$S_{\alpha\beta}(Q) = 1 + \frac{4\pi\rho}{Q}\int_0^\infty r[g_{\alpha\beta}(r) - 1]sin(Qr)dr \tag{2.28}$$

where ρ is the atomic number density and r the radial distance. If the Fourier Transform of this equation is taken, it is possible to extract the radial distribution function, $g_{\alpha\beta}(r)$, which denotes how the density of atom type β around atom type α varies with the radial distance. The calculation of g(r) was described previously in Section 1.2.1. For a system with J atomic components, the number of *g(r)*'s and *S(Q)*'s, *N*, equates to:

$$N = \frac{J(J+1)}{2} \tag{2.29}$$

If a calculated $g_{\alpha\beta}(r)$ is integrated, an associated coordination number is derived. The coordination number determines the number of β atoms around an α atom, within a distance range of r_{min} to r_{max}:

$$n_{\alpha\beta}(r) = 4\pi\rho c_\beta \int_{r_{min}}^{r_{max}} r^2 g_{\alpha\beta}(r)dr \tag{2.30}$$

The value thus reports the average number of bonds that are being formed between a given pair, in the system.

The relations presented in this section assume that attenuation, as well as multiple and inelastic scattering effects are not present. For a neutron diffraction experiment, these effects have to be corrected for, with the corrections needed dependent on the instrument being utilised. The corrections completed for this thesis will be described in more detail in Section 2.1.4.

Whilst there is the potential to extract *N*, *S(Q)*'s and *g(r)*'s from the data alone; this is an idealistic situation. In the case of the biomolecular solutions studied here,

much of the signal will be from water molecules, as these molecules are more prevalent in the solutions. Consequently, this makes it extremely difficult to extract meaningful information about all pairwise interactions from the data alone. To overcome the issue, a computational model of the system that fits the experimental data, which structural information can be extracted from. The method to produce such a model here is presented in Section 2.2.

2.1.4 The Neutron Diffraction Experiment

The experiment completed by James Chadwick (Section 2.1.2.1), though successfully producing neutrons, was not a self-sustaining reaction that could be used to produce a consistent supply of neutrons (74). Such a supply of neutrons is necessary to be able to complete any experiment and develop a suite of instruments for neutron scattering. This section details the ways by which neutrons are produced and the instruments available to study biological matter. It will cover the practical details of the neutron diffraction experiment.

2.1.4.1 Producing Neutrons via Fission

Following Chadwick's experiments, the first step to producing a self-sustaining source of neutrons was the development of the Chicago-Pile 1 reactor (74). This was built as part of the Manhattan Project in the 1940's. The reactor utilised Uranium (^{235}U) to initiate a self-sustaining reaction, capable of producing neutrons.

The process being used for Chicago-Pile 1 and many reactors for producing neutrons is fission. This process is shown schematically in Figure 2.11. The fission process requires neutrons to be propelled towards a larger nucleus, such as Uranium or Lead. The larger nucleus, upon collision, is broken up into smaller fission fragments and a continuous stream of neutrons. For the case of Uranium, for every one neutron that collides with the nucleus an average of 2.4 neutrons are produced (74). These neutrons are able to interact with other nuclei in the target, thus producing more neutrons. The neutrons produced emerge for an experiment as a continuous white beam. Facilities that utilise

this form of neutron production include the Institut-Laue Langevin (ILL), France, and OPAL, Australia.

Whilst neutron diffraction experiments are completed at fission reactor sources, there is more difficulty in the instrumental design. The fission process results in a white beam of neutrons which have a variety of wavelengths and travel over a range of angles. As stated in Section 2.1.2.4, a diffraction experiment requires a coherent source of radiation in order to obtain a diffraction pattern. Therefore, the beam for a diffractometer at a fission neutron source has to be monochromated to ensure coherency. Monochromation allows only neutrons with a certain wavelength to interact with the sample. This is an ineffective use of the neutrons, with most of the neutrons lost as they meet the monochromator and any signal being vastly reduced with only a limited number reaching the sample and detector.

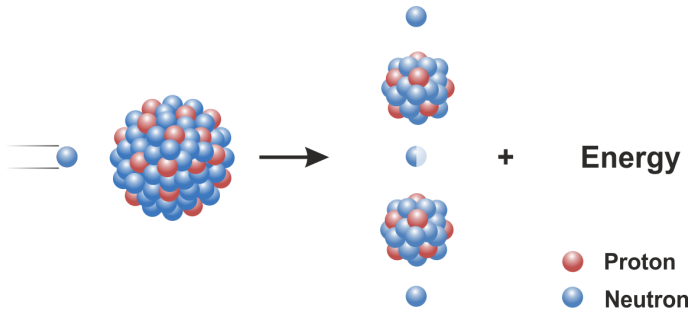


Figure 2.11: A schematic of the fission process. When a neutron collides with a larger nucleus, such as Uranium, this causes the nucleus to fragment, releasing neutrons and other fission products. The process also causes a loss of energy. Fission is used to produce neutrons at facilities such as the Institut-Laue Langevin, France.

Despite the the beam being monochromated, the neutrons produced also have high energy levels. The fission reaction produces a high flux and 192.9 MeV of energy. This excessive energy is a downfall as it inhibits the ability of the reaction to be self-sustaining. With high energy comes high velocity, and neutrons at increased velocity are less likely to interact with another nucleus; the overall cross-sectional area is reduced for these fission neutrons. As the energy is related to the momentum, this will influence the wavelength of the neutrons, as depicted in the de Broglie relation (Equation 2.3). The wavelengths of such neutrons are too small to be efficiently used as a probe of atomic

structure in a diffraction experiment. Therefore, it is necessary to reduce the energy of these neutrons.

A way to decrease the energy of these neutrons is via the use of moderator. A moderator is material that causes the radiation to undergo multiple inelastic events. In inelastic scattering, the interaction between the probe and sample causes a loss of energy. This process can be optimised if the neutrons collide with nuclei of a similar mass to their own. This means neutrons can be passed through a low atomic mass moderator to lose energy. Moderators are common feature at neutron sources and are used to finely tune the energy and thus wavelength to suit the experiment (see Section 2.1.4.5). For a reactor source, moderators have to be used extensively to drastically lower the energy of the neutrons that emerge.

2.1.4.2 Producing Neutrons via Spallation

A more recent development in the production of neutrons is the use of spallation sources. Spallation consists of pulses of charged particles being accelerated to high energies and interacting with a target nucleus. This interaction leads to the excitation of the nucleus and starts an intranuclear cascade, whereby the subatomic particles of the nucleus collide with one another. This process causes high energy particles to be ejected. Some of these particles are neutrons, but they possess too high an energy to be used for an experiment. Theoretically, they do have the capability to interact with other nuclei, via an internuclear cascade, though the low cross-sectional area means the probability of this is low, as described previously (Section 2.1.4.1). The process of losing high energy particles leaves the target nucleus in an excited state. In order to de-excite the nucleus, it undergoes evaporation, whereby lower energy neutrons and other particles are released. It is these neutrons that can be used for scattering experiments. The spallation process is shown in Figure 2.12.

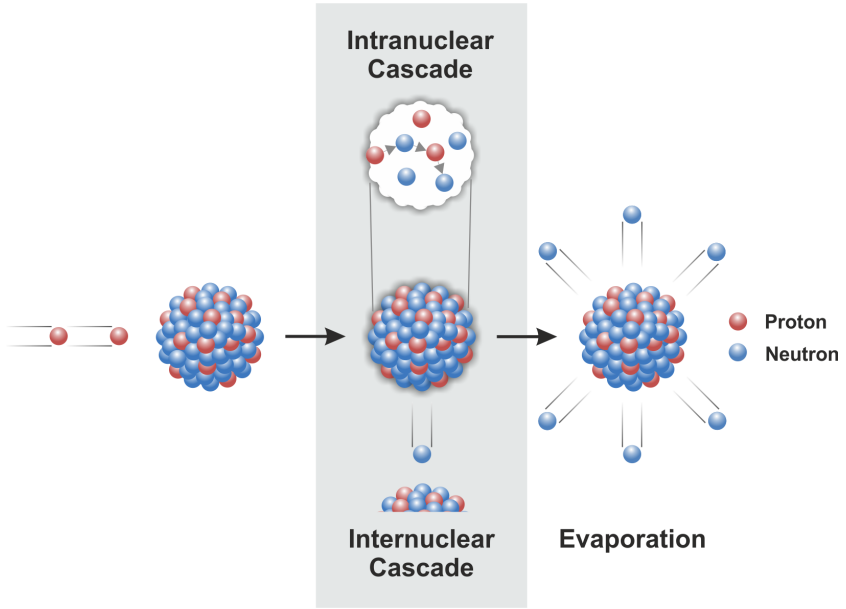


Figure 2.12: A schematic of the spallation process. Only the loss of neutrons is shown, though other particles are lost during the process, including protons, muons and gamma rays. Upon protons interacting with the nucleus, neutrons are lost as subatomic particles collide in the nucleus (intranuclear cascade), with these neutrons being able to collide with other nuclei (internuclear cascade). This puts the nucleus in a high energy state. To lower the energy, neutrons evaporate from the nucleus. Spallation is used to produce neutrons at facilities such as ISIS, UK.

Spallation sources can either be continuous or pulsed. An example of a continuous source is SINQ in Switzerland. With such a source, the same requirement of monochromation is necessary for any diffractometers. For a pulsed source, the neutrons arrive in packets due the protons in the first stage of the spallation process arriving in pulses. This means the neutrons are in time-of-flight mode, making it possible to calculate the precise wavelength of the neutrons through kinematics:

$$\lambda = \frac{h(t - t_0)}{m(L - L_0)} \tag{2.31}$$

Here, h is Planck's constant, m the mass of a neutron, t is the time between consecutive proton pulses hitting the target and L the detection distance. The values t_0 and L_0 are pre-calibrated values and are specific to a given instrument. This means both the angle and wavelength can be calculated for an experiment, eliminating the

need for a monochromator. Whilst present at pulsed sources, it is possible for certain continuous-source instruments to also work in time-of-flight mode. Pulsed spallation sources include the Spallation Neutron Source (SNS) at Oak Ridge, USA and JPARC, Japan. The experiments for this work have been completed at the UK ISIS facility, which is also a pulsed spallation source.

2.1.4.3 The ISIS Facility

The ISIS facility is the UK's national source for neutron science and is situated within the Rutherford Appleton Laboratory. The first beam was produced in 1984 and now neutrons are supplied to 30 instruments for diffraction, spectroscopy, reflectometry and other scattering experiments, with 4 more in development. As mentioned in Section 2.1.4.2, the process of spallation is also able to produce other subatomic particles, including muons. Consequently, ISIS also incorporates a muon facility, comprised of 5 operational instruments that are situated within Target Station 1.

A schematic of the ISIS facility is shown in Figure 2.13. Production of neutrons at ISIS first begins with the formation of H^- ions at an ion source. The ions drift towards a linear accelerator (LINAC), where they are bunched together within the LINAC with copper electrodes and accelerated to energies between 650 keV and 70 MeV, in bursts lasting 500 μs. With each burst occurring every 20 ms, the frequency of the pulses is 50 Hz. The ions enter the synchrotron, with their electrons removed by passing them through a sheet of alumina, leaving protons, H^+ ions. The incoming protons are accumulated and accelerated to energies in the range of 70 to 800 MeV. The protons are collected into 2 bunches in the synchrotron, with more added over a period of $\sim$10000 iterations. These bunches are then guided by magnets to a target station, where they interact with a tungsten target. Colliding with the tungsten target initiates the process of spallation (see Section 2.1.4.2) that will produce neutrons. Whilst most bunches go to Target Station 1, every 5th proton bunch is sent to Target Station 2. Thus, the average frequency for a pulse at Target Station 1 is 40 Hz and 10 Hz at Target Station 2.

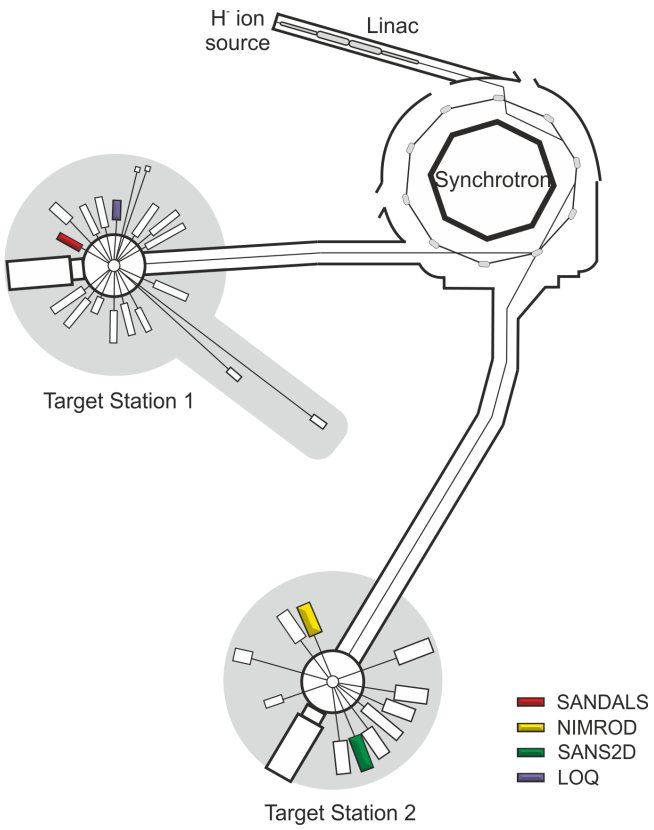


Figure 2.13: The layout of the ISIS facility. Protons from the source are accelerated in a LINAC and collect in the synchrotron. Two pulses are then emitted to supply target stations 1 and 2. Target station 1 is the location of the SANDALS instrument (red); the instrument which all neutron diffraction experiments for this thesis have been completed on

2.1.4.4 Using ISIS for the Study of Biological Matter

For the fields of soft and biological matter, the ISIS facility has a selection of instruments available to complete neutron-based experiments. Neutron reflectometry can measure the properties of a medium by analysing the reflection of neutrons from it. Reflectometry is specialised to the study of thin films, layers and interfaces and instruments at ISIS that cater for such experiments include Surf and Polref. Neutron spectroscopy is used to study the dynamics of atoms and molecules. The information attained from a given spectroscopy experiment depends on the elastic/inelastic and coherent/incoherent nature of the scattering and can be completed on instruments like Mari, OSIRIS and TOSCA.

Small-angle scattering instruments (SANS) and neutron diffractometers are specialised to the study of structure. At ISIS, there are four operational instruments designed to study the structure of materials in solution, including biological matter; two small-angle scattering instruments (LOQ and SANS2D), and two diffractometers (SANDALS and NIMROD). Figure 2.13 shows the location of these four instruments within the ISIS facility.

Figure 2.14 shows the distance range probed by each of these techniques. Small-angle instruments are specialised to the study of bulk structure and can be exploited for studying biomolecular clustering. For an amino acid, which has an average backbone length of ∼4 Å, the small angle scattering instruments, from their minimum distance, would be able to detect the presence of small dimers, trimers or higher if they were a dominant or consistent species in the solution. For the work completed in this thesis, the main interest is in the Ångstrom-scale interactions between molecules. To probe intermolecular hydrogen bonding interactions between biological molecules for instance, which are no longer than 3.5 Å (see Figure 1.6), only a neutron diffractometer could detect such length-scales. Consequently, this is the reason behind using neutron diffraction for this thesis.

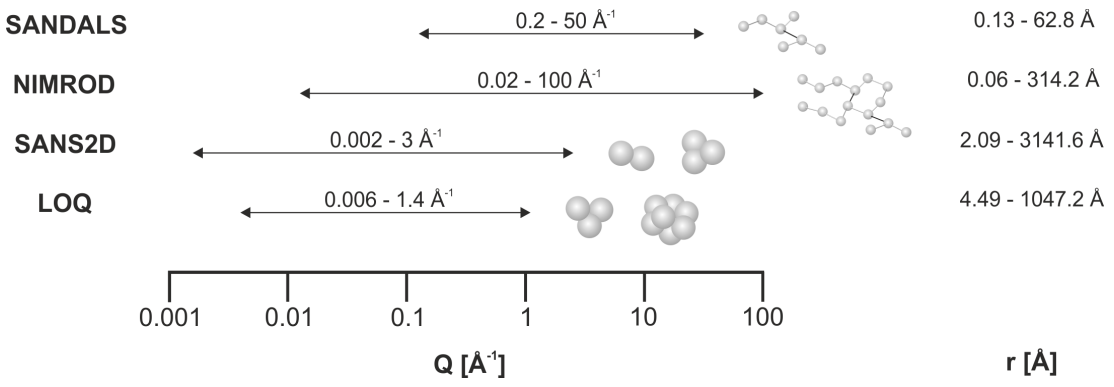


Figure 2.14: A comparison of the Q-range covered by the instruments at ISIS that can probe the structure of biological matter in solution. Whilst SANS instruments such as SANS2D and LOQ can determine bulk structure and the ability of small molecules to associate, diffractometers such as SANDALS and NIMROD aim to study the atomic level interaction of small and larger biomolecules respectively in solution.

The information obtained from each structure-based method is complementary. In this thesis, there will be an example of where the LOQ instrument has been used to support the information obtained from neutron diffraction on the association of amino acid molecules in solution. The theory and brief experimental procedure for the SANS measurements is described in Section 2.3.

2.1.4.5 The SANDALS Instrument

Neutron diffraction experiments have been completed using the SANDALS instrument. SANDALS is an acronyn for 'Small Angle Neutron Diffractometer for Amorphous and Liquid Samples'. This instrument has been used to study the structure of myriad organic (101–118), ionic (119–125) and biological molecules (127–135; 137–142) in solution. It is tailored to probe the structure and bond distances within systems containing small molecules, making it appropriate for the systems in this thesis.

The SANDALS instrument is shown in Figure 2.15. The instrument possesses 663 detectors to allow for the detection of neutrons in an angular range of 3 - 40°. These detectors are made from Zinc Sulphide (ZnS) and have been doped with other elements to allow for the neutrons to be detected (see Section 2.1.4.6). The raw data is recorded as the number of counts measured by each detector, with one count equating to one neutron. As these are time-of-flight measurements, the data is in the form counts μs^{-1}. The wavelength of the neutrons is fine-tuned with a methane moderator, before interacting with a sample. The wavelength range of the incident neutrons is 0.05 - 4.95 Å. It can provide information in a Q-range of 0.1 - 50 $Å^{-1}$, corresponding to distances of 0.1 - 62.8 Å.

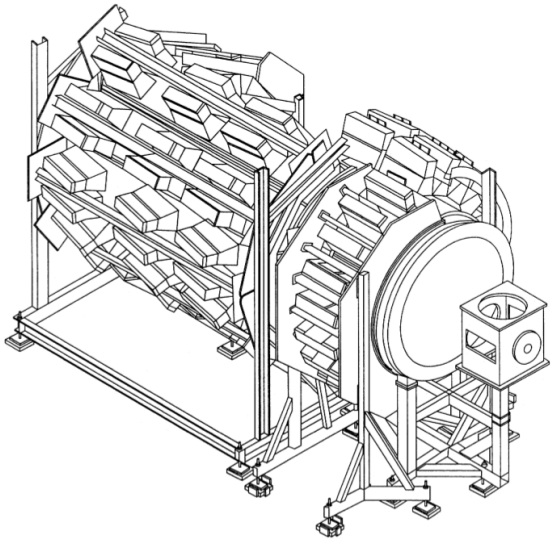


Figure 2.15: Diagram of the SANDALS instrument, used for all neutron diffraction experiments discussed in this thesis. This instrument has been used as it is sensitive to the location and interactions of hydrogen atoms. SANDALS is an acronym for Small Angle Neutron Diffractometer for Amorphous and Liquid Samples. Picture taken from (187).

Samples to be run on SANDALS are encased in Titanium-Zirconium (TiZr) cans. A diagram of the TiZr cans used is shown in Figure 2.16. The cans are flatplates, meaning that the thickness of the sample is the same across the whole neutron beam. They can hold a volume of approximately 1.3 cm^3. The cans are made out of TiZr as these two elements have high incoherent signals and low coherent signals, meaning the coherent scattering present will be predominantly from the sample and will be easier to extract. As well as recording data for the sample, it is necessary to take data for the empty can, the empty instrument and a Vanadium standard. Vanadium is a very strong incoherent scatterer and thus serves as a good standard for calibration (see Section 2.1.4.7). These three extra samples allow for corrections to be completed on the data for attenuation effects. The data correction process is detailed in Section 2.1.4.7.

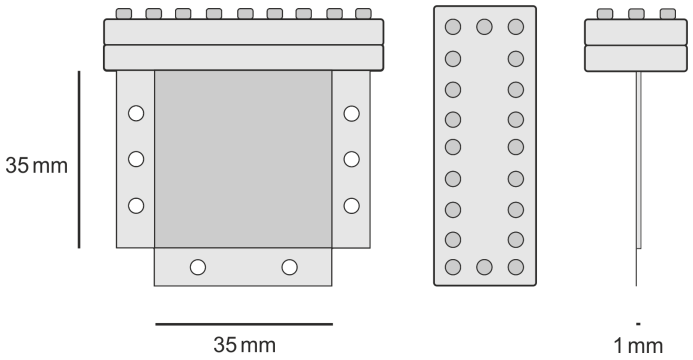


Figure 2.16: Schematic of the Titanium-Zirconium cans used in most SANDALS experiments. These are used as the elements are good incoherent scatterers. The TiZr cans have dimensions of 35 x 35 x 1 mm, making the volume of the can approximately 1.3 cm^3.

2.1.4.6 Neutron Detection

Section 2.1.4.5 referred to the many detectors present to intercept scattered neutrons. As neutrons have no charge for reasons described in Section 2.1.2.1, this means they cannot be directly detected. To overcome this issue, a charged particle has to be produced from the incoming neutrons, which can then be detected. This charged particle is produced through one of the following reactions occurring within the detector:

The scintillation detectors used by the SANDALS instrument are doped with lithium ions (^{6}Li) and silver (Ag) to allow the following reaction to take place:

$$^6Li + n \rightarrow {}^4He + {}^3H + 4.8\,\text{MeV} \tag{2.32}$$

This interaction produces Helium (^{4}He), tritium (^{3}H) and MeV of energy. When this reaction takes place, the energy produced is able to excite an electron within the ZnS lattice (188). This electron combines with a hole and forms an exciton; a type of quasi-particle. The exciton will travel along the lattice until reaching an Ag site. Ag is an impurity activator and upon reaching one of these atoms, the electron and hole recombine, which releases a photon. The photon will pass through the lattice to a

photomultiplier tube (PMT), which multiplies the signal from the photon and produces a signal. The photon can pass through the lattice as the energy is less than that required to excite the ZnS atoms.

This process allows a data signal to be produced for the experiment. This data contains not only the results from elastic scattering, but also inelastic scattering and attenuation effects. Therefore, it is necessary to pre-treat the data to remove these unwanted effects. The process by which this is done is explained in the next section.

2.1.4.7 Using Gudrun to Process Raw Neutron Diffraction Data

As discussed in Section 2.1.3, it is necessary to extract the distinct component of the cross-section to be able to obtain structural information on a system. This requires studying the elastic, coherent scattering of neutrons that takes place. There are also multiple scattering effects that will also appear in the raw data of a diffraction experiment, which requires correcting before analysis can take place.

The datasets are corrected using the software Gudrun (189). The Gudrun routines are based upon the algorithms within the ATLAS suite of programs (190). Gudrun utilises the information from the three background datasets (see Section 2.1.4.5) to remove unwanted scattering effects in the data from the can and plots the data on an absolute scale.

The first stage of the process is to remove data from faulty detectors. Within the Gudrun package, there is a routine called 'purge' that can complete this correction. Purge works by grouping the detectors in 18 groups and comparing their standard deviations. Those detectors with extremely high and low standard deviations are removed. A high standard deviation, defined as a standard deviation more than 10 times the average value for the group, can be caused by a noisy dataset. A low standard deviation, equating less than one tenth of the average, would be the result of a weakly counting detector.

The next step is to correct for attenuation, any effects that cause a decrease in the signal. Such effects[1] include multiple scattering (when a neutron interacts with the sample more than once), inelastic scattering (when energy is transferred upon interaction) and absorption effects (when a neutron is absorbed upon approaching a nucleus). The total cross-section from these effects can be defined as:

$$\sigma^{tot}(\lambda) = \sigma^{s}(\lambda) + \sigma^{a}(\lambda) \tag{2.33}$$

Here, $\sigma^{s}(\lambda)$ is the scattering cross-section and $\sigma^{a}(\lambda)$ is the absorption cross-section. To correct for these effects, it is necessary to find $\sigma^{tot}(\lambda)$. In many cases, particularly where the elements have a high atomic mass, these cross-sections do not depend on wavelengths and can be found in Sears tables (186). For systems like those in this thesis, which contain hydrogen and deuterium, the cross-sections do have wavelength dependency. Therefore, it is necessary to measure the transmission cross-section of these elements (191):

$$\sigma^{trans}(\lambda) = e^{-\rho\sigma^{tot}(\lambda)L} \tag{2.34}$$

where ρ is the sample density and L is the sample thickness. The sample thickness is 1 mm, equivalent to the thickness of the TiZr can. Measurements of $\sigma^{trans}(\lambda)$ occur simultaneously with the other datasets. By measuring this, as well as determining the density of the sample, $\sigma^{tot}(\lambda)$ can be calculated.

All datasets are also normalised to remove the effects of the incident flux, or change in levels, of the incident neutrons. The flux is monitored throughout the experiment by having a monitor in the line of the incident beam. To ensure the neutrons still reach the sample, a low efficiency counter is used.

[1]The scattering effects specified can also be termed as recoil effects in the literature (1). This term is more often used when describing a process whereby momentum is transferred, causing recoil of the nucleus, and is an effect observed in neutron scattering.

Finally, the data is converted to an absolute scale. This requires a vanadium standard and this conversion uses the equation:

$$\text{Corrected Result} = \frac{(\text{Sample Data} - \text{Empty Container})}{(\text{Vanadium Standard} - \text{Empty Instrument})} \tag{2.35}$$

As referred to in Section 2.1.4.5, Vanadium is an incoherent scatterer and due to its high atomic mass, it is easy to calculate $\sigma^{tot}(\lambda)$ for this sample. The data is put onto a scale so that the final result is in barns per atom per steradian (barns atoms^{-1} sr^{-1}).

During the completion of the work in this thesis, the corrections for attenuation, as well as absolute scale conversion, became an automated process. The format of the routine is shown in the flowchart in Figure 2.17. The steps that consider the wavelength serve to determine $\sigma^{tot}(\lambda)$. For the Q-dependence, this is the step by which the data is put on an absolute scale and it is possible to view the structure factors as a function of Q. This process is iterative and the routine follows the steps of the flowchart repetitively until all corrections have been optimised. Upon completing all corrections, it is possible to analyse the data, to obtain structural information.

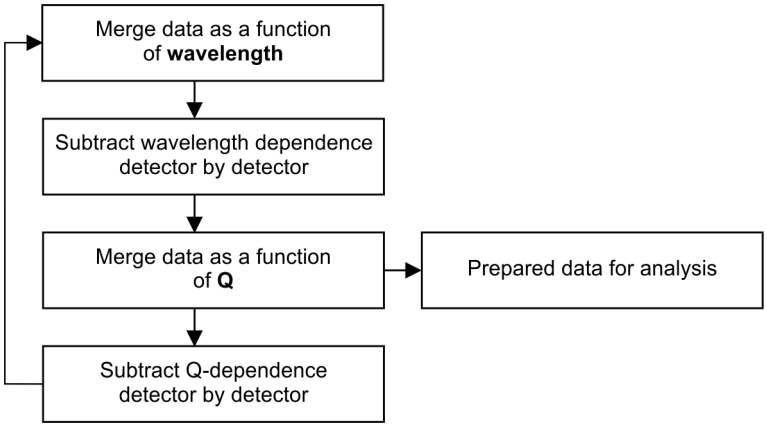


Figure 2.17: The Gudrun routine to correct for attenuation, multiple and inelastic scattering effects. This procedure is iterative and cycles through corrections for scattering and attenuation, by considering the wavelength, and putting the data on an absolute scale for analysis in Q-space. The process runs through the iterations until these corrections have been optimised.

2.2 EPSR Modelling

This section discusses the computational modelling completed to obtain atomic information on the systems covered in this thesis. Section 2.1 described the theory and experimental methodology completed for the neutron diffraction experiment. Upon completing a neutron diffraction experiment, the data requires decomposition to extract atomic-level information on the system. As described in Section 2.1.3, it is not possible to obtain all partial structure factors for a given system using only the experimental data. In order to obtain these, the solution is to use computational modelling, which can produce a simulated structure that mimics that shown by the data.

EPSR, which stands for Empirical Potential Structure Refinement, is a software package specially designed to produce simulations for the analysis of neutron and X-ray diffraction data. The first version was developed in 1996 (192) and was developed as an alternative to completing Reverse Monte Carlo (RMC) simulations on molecular systems; RMC being a method of obtaining computational parameters from data. Whilst this section will cover the key theory of the EPSR process, it will not serve to give a guide to computational modelling in general. A plethora of literature exists that delves into the details of computational modelling, including (97; 193).

Figure 2.18 shows a more extensive flowchart of the integration of neutron diffraction into the EPSR process. Stage 1 is the neutron diffraction process, which was described in Section 2.1. This section will now aim to provide details for Stages 2 and 3. The first stage of the EPSR method, Stage 2, requires the creation and equilibration of a model of the system, consisting of the appropriate molar ratio and relevant potentials. The result is a box with a reference potential $U_{REF}\,(\mathrm{r})$. A detailed explanation can be found in Section 2.2.1. In Stage 3, the data and reference simulation are compared, allowing for the calculation of coefficients and an empirical potential, U_{EP}, that can be introduced into the simulation iteratively. This drives the simulation to fit the experimental data by deriving the missing potential between the simulation and the data. The sought energy

of the system shown in the data can be classified as:

$$U_{TOT} = U_{REF} + U_{EP} \tag{2.36}$$

The structure derived from EPSR may not be the only result that is consistent with the experimental data. However, EPSR will produce a physically plausible model that fits the experimental data. Stage 3 is described in Section 2.2.2.

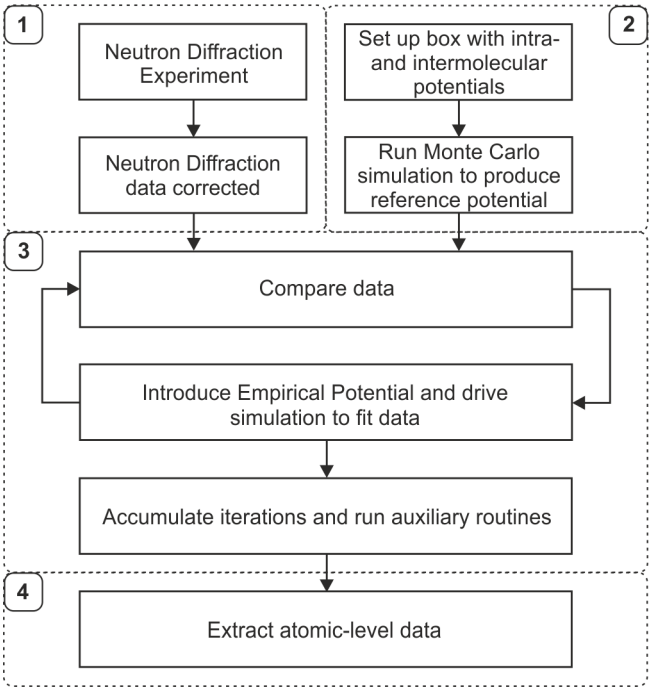


Figure 2.18: Flowchart of the core components of obtaining atomic-level information (4) from neutron diffraction data and EPSR modelling. Upon completing the neutron diffraction experiment (1), a reference model with computationally-derived parameters is made up of the same molar ratio as the system (2). The data and simulation are then compared, allowing an empirical potential to be introduced that drives the simulations to fits the data (3).

2.2.1 The Reference Potential

The first stage of EPSR is to create a simulation box of the system. Such a box will be created to have the correct molecular ratio and density, and will consist of suitable constraints on the molecules to guide their conformations and interactions. This

creates an energy term for the box; the reference potential. The reference potential uses literature-based or computationally-derived parameters to assert these constraints. These are intra- and inter-molecular forces, concerning the structure of the molecules, and forces governing interactions between them. Extracted from the literature will be values for the Lennard-Jones potentials and Coloumbic potentials for the atoms, which were introduced in Section 1.1.2, and geometry of the molecules. This helps to create a box similar to that described by the data and prevent occurrences of atomic overlap and repulsive terms. Upon creating the box, the simulation is equilibrated to produce an initial fit to the data. The reference potential parameters used for the simulations in this thesis are presented in Chapters 3, 4 and 5.

2.2.1.1 The Intramolecular Potential

The intramolecular potential describes the energy between atoms within a given molecule, based on its geometry. This intramolecular potential between atoms α and β is defined by the equation:

$$U_{intra} = C \sum_{i} \sum_{\alpha\beta \neq \alpha} \frac{(r_{\alpha_i\beta_i} - d_{\alpha\beta})^2}{2w_{\alpha\beta}^2} \tag{2.37}$$

In this definition, the actual distance, $r_{\alpha\beta}$ between atoms α and β in a molecule is compared to the average distance, $d_{\alpha\beta}$. Also present is a width function, $w_{\alpha\beta}$, a broadening function used to account for fluctuations in the intramolecular distance. It is used in place of the Debye-Waller factor, typically used to account for thermally-sourced fluctuations, but which is difficult to incorporate for atomic pairs. The square of $w_{\alpha\beta}$ is defined as:

$$w_{\alpha\beta}^2 = \frac{d_{\alpha\beta}}{\sqrt{\mu_{\alpha\beta}}} \tag{2.38}$$

where $\mu_{\alpha\beta}$ is:

$$\mu_{\alpha\beta} = \frac{M_\alpha M_\beta}{M_\alpha + M_\beta} \tag{2.39}$$

Here, $\mu_{\alpha\beta}$ represents the reduced mass between atoms α and β, with M_α being the mass of atom α.

Equation 2.37 shows that geometrical parameters in the simulation are constrained by calculating distances. EPSR does not explicitly define parameters like bond angles. To impose a bond angle, $\overline{ABC}$, EPSR does this by calculating a pseudo bond between the first and last atoms, $\overline{CA}$. The $\overline{CA}$ pseudo bond is not a physical bond between the two atoms but is able to constrain the angle due to the law of cosines:

$$\overline{CA}^2 = \overline{AB}^2 + \overline{BC}^2 - 2\overline{AB}\,\overline{BC}cos(\overline{ABC}) \tag{2.40}$$

The pseudo bond calculated by determining the difference between the coordinates of A and C. The bond angle though will fluctuate around the specified value, due to the smearing function that will alter the bond lengths, as described by Equation 2.38.

2.2.1.2 The Intermolecular Potential

The intermolecular potential describes the energy between atoms in different molecules. This is governed by Lennard-Jones (LJ) potentials and Coulombic charges, described in more detail in Section 1.1.2. The total intermolecular potential between atoms α and β, in different molecules, is therefore:

$$U_{inter}(r_{ij}) = 4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{6}\right] + \frac{q_\alpha q_\beta}{4\pi\epsilon_0 r_{ij}} \tag{2.41}$$

Here, $\epsilon_{\alpha\beta}$ and $\sigma_{\alpha\beta}$ are defined by the Lorentz-Berthelot mixing rules (97):

$$\epsilon_{\alpha\beta} = (\epsilon_\alpha \epsilon_\beta)^{\frac{1}{2}} \tag{2.42}$$

$$\sigma_{\alpha\beta} = \frac{1}{2}(\sigma_\alpha + \sigma_\beta) \tag{2.43}$$

2.2.1.3 Accounting for Long-Range Effects

When the simulation runs, it is not possible to refine the model to account for long-range effects. It is not feasible to calculate the empirical potential at distances larger than r_{max}, the largest radial distance possible in the box, and to do so would be computationally and prohibitively time-consuming. Long-range effects are eliminated using a function to truncate the potential. Eliminating long-range effects does not have any significant impact on the position and movement of atoms on a short-range basis, which is of interest in this work. Therefore, the potential is truncated over a distance range, to allow for the truncation to be smooth. For the non-Coulombic part, this is defined by the following set of rules:

$$T_{LJ}(r) = \begin{cases} 1 & r < r_{minpt} \\ 0.5\left(1 + cos\pi\left(\frac{r - r_{minpt}}{r_{maxpt} - r_{minpt}}\right)\right) & r_{minpt} < r < r_{maxpt} \\ 0 & r > r_{maxpt} \end{cases} \tag{2.44}$$

r_{minpt} is the minimum distance to start truncating for a given reference potential, and r_{maxpt} the maximum. The Coulombic part is truncated using the 'Reaction Field Method', from the work of Hummer et al (194):

$$T_c(r) = \left(1 - \frac{r}{r_{maxpt}}\right)^4 \left(1 + \frac{8r}{5r_{maxpt}} + \frac{2r^2}{5r_{maxpt}^2}\right) \Theta(r_{maxpt} - r) \tag{2.45}$$

$\Theta(r_{maxpt} - r)$ in Equation 2.45 is the Heaviside function; a step function that equals 0 when $r_{maxpt} < r$ and 1 when $r_{maxpt} > r$.

2.2.1.4 Periodic Boundary Conditions

In a neutron diffraction experiment, the solutions will contain myriad water and biomolecules. To create a simulation box with so many molecules far breaches the computational limits of the EPSR method; most versions of the software have the constraint that no more than 30,000 atoms can be present in a box. Realistically, it is not possible to create a box that matches the dimensions of the system studied.

This issue can be helped with the use of periodic boundary conditions. Instead of constraining the movement of molecules to the limits of the box, they are able to pass through the walls and appear on the other side. A schematic is shown in figure 2.19. This helps to mimic the idealistic situation of having an 'infinite' number of boxes, and ensures that upon passing through the walls, no molecules are lost.

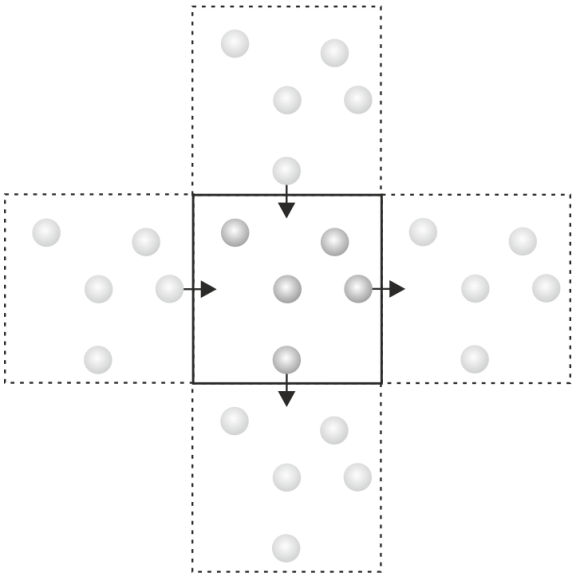


Figure 2.19: Schematic of the periodic box set up used in EPSR. The periodic box permits a molecule that leaves the box to enter at the other side. This allows the molecular composition of the box to be maintained.

2.2.1.5 Running the Monte Carlo Simulation

Upon setting up the reference model with the relevant potentials and parameters, a pure Monte Carlo simulation is run on the box. The Monte Carlo simulation statistically samples the simulation box to produce a simulated structure factor for each dataset. In order to obtain the lowest energy system, atomic and molecular translations are permitted. The four movements that can take place within EPSR are:

- Atomic translations
- Molecular translations
- Molecular rotations
- Functional group rotations

For any move to be accepted, the movement has to satisfy the Metropolis Condition. The Metropolis Condition states that a movement can be accepted if it causes a decrease in the potential energy of the system:

$$\Delta U(r) = U_{before} - U_{after} \begin{cases} < 0 & \text{Move accepted} \\ > 0 & \text{Move rejected} \end{cases} \tag{2.46}$$

The probability of accepting any move is dependent on the movement-type. For an atomic translation, only the change in the intramolecular potential is of importance, with the consequential probability defined as:

$$\text{Acceptance Probability} = e^{-\Delta U_{intra}(r)} \tag{2.47}$$

The molecular movements are influenced by groups within and surrounding a given molecule. Thus, the probability of acceptance, for the remaining three movements, is defined as:

$$\text{Acceptance Probability} = e^{-(\Delta U_{intra}(r) + \frac{1}{k_B T}(\Delta U_{REF}(r) + \Delta U_{EP}(r)))} \tag{2.48}$$

Here, U_{EP} only becomes relevant when the empirical potential is added. The Monte Carlo simulation will run until the total and partial structure factors have stopped changing. Upon equilibrating the simulation containing the reference potential, the empirical potential can be introduced. The definition of the empirical potential and how it is introduced is described in the next section.

2.2.2 The Empirical Potential

The next stage is to introduce the empirical potential, U_{EP}. This drives the simulation to reach the potential energy from the experimental data, U_{TOT}, as denoted in Equation 2.36. Compared to the reference potential, the empirical potential does not possess strict guidelines and upon increasing its value, it guides the molecular movements in a manner that fits the experimental data. The following sections detail how the empirical potential is defined and calculated for a given system.

2.2.2.1 Defining the Empirical Potential

When the empirical potential is introduced it should only portray realistic differences between the data and simulation, and should not possess any artefacts due to over-structuring, statistics and truncation. After many trials of EPSR, the most successful way found to attempt to overcome these potential errors is to express the empirical potential as a series of Poisson functions;

$$U_{EP}(r) = kT \sum_{k} C_k p_{n_k}(r, \sigma_r) \tag{2.49}$$

where:

$$p_n(r, \sigma_r) = \frac{1}{4\pi\rho\sigma^3(n+2)!} \left(\frac{r}{\sigma}\right)^n e^{-\frac{r}{\sigma}} \tag{2.50}$$

In Equation 2.49, C_k is a coefficient that takes on a positive or negative, real value; σ is a width function and ρ the atomic number density of the system. A Fourier Transform converts the function p_n into Q-space:

$$\begin{aligned} P_n(Q,\sigma) &= 4\pi\rho \int p_n(r) e^{i\mathbf{Q}.\mathbf{r}} d\mathbf{r} \\ &= \frac{1}{(n+2)(\sqrt{1+Q^2\sigma^2})^{(n+4)}} [2cos(n\alpha) + \frac{(1+Q^2\sigma^2)}{Q\sigma} sin(n\alpha)] \end{aligned} \tag{2.51}$$

Here, $\alpha = arctan(Q,\sigma)$.

Therefore, C_k can be estimated from the diffraction data by fitting a series of the form:

$$U_{EP}(Q) = \sum_k C_k P_{n_k}(Q,\sigma_Q) \tag{2.52}$$

In the EPSR process, the empirical potential is introduced by comparing the data and simulation, with the coefficient recalculated with each iteration. This process is described in more detail in the next section.

2.2.2.2 Refining the Empirical Potential

As described in the previous section, it is necessary to calculate an empirical potential iteratively to drive the simulation to fit the data. Therefore, EPSR is set up to recalculate the coefficients and refine the empirical potential values, to drive the simulation to fit the experimental data.

Upon completing an EPSR simulation, the aim will be to attain the relation between pairs of atoms in the form of partial structure factors. Up to the point of equilibrating the reference potential, the simulation has not considered the confidence that should be placed on the data and how accurate it is in describing the relation between pairs of atoms. The refinement process thus takes into account the reliability of the data when recalculating the coefficients and empirical potential. For this stage, the total structure

factor needs to be considered in the following form to the difference function:

$$F_i(Q) = \sum_{j=1,N} w_{ij} S_j(Q) \tag{2.53}$$

Where i is the dataset number and j a given atomic pair, $\alpha\beta$. The equation includes the weights matrix, w_{ij}, which for un-normalised data takes on the form:

$$w_{ij} = (2 - \delta_{\alpha\beta}) c_\alpha c_\beta b_\alpha b_\beta \tag{2.54}$$

This can be put into the format of an (M x N) matrix, where M is the total number of datasets. If there are a total of J atomic components, then there are $N = J(J + 1)/2$ partial structure factors and RDFs. To obtain the partial structure factors, it is necessary to invert w_{ij}. The common scenario though is that there are too few datasets compared to partial structure factors, which makes this matrix difficult to invert. To allow the matrix to be more easily inverted and account for the reliability of the data, a confidence factor or probability, f, can be assigned to each dataset. The remaining probability is given to the data. Thus, w_{ij} is equal to:

$$w'_{ij} = \begin{cases} f w_{ij} & 1 \leq i \leq M \\ (1 - f)\delta_{(i-M),j} & M < i \leq (M + N) \end{cases} \tag{2.55}$$

The contribution from the simulation takes the form of an (N x N) matrix. This matrix can be merged with the original one to form a ((M + N) x N) matrix. This matrix, when applied to $F_i(Q)$, can be visualised in Figure 2.20. The matrix for the contribution of the simulation contains values only when i = j due to presence of Kronecker delta, δ_{ij}, which states:

$$\delta_i(x) = \begin{cases} 1 & \text{if } x = i \\ 0 & \text{otherwise} \end{cases} \tag{2.56}$$

Using the rules of pseudo-inversion of matrices, it is then possible to invert this matrix to obtain the appropriate number of partial structure factors. The inverse of w_{ij} can be used to find the coefficient $C_k^{(i)}$.

$$C_{k,m+1}^{(j)} = C_{k,m}^{(j)} + \sum_{i=1,M} w_{ji}^{-1} C_k^{(i)} \tag{2.57}$$

For the first iteration, $C_{k,m}^{(j)}$ is zero, and $C_{k,m+1}^{(j)}$ is calculated by finding the difference between the data and the simulation, *D(Q) - F(Q)*. $C_{k,m+1}^{(j)}$ will become the new $C_{k,m}^{(j)}$ in the following iteration. The value calculated for $C_k^{(i)}$ in Equation 2.55 can be used in Equation 2.49 to determine a value for the empirical potential.

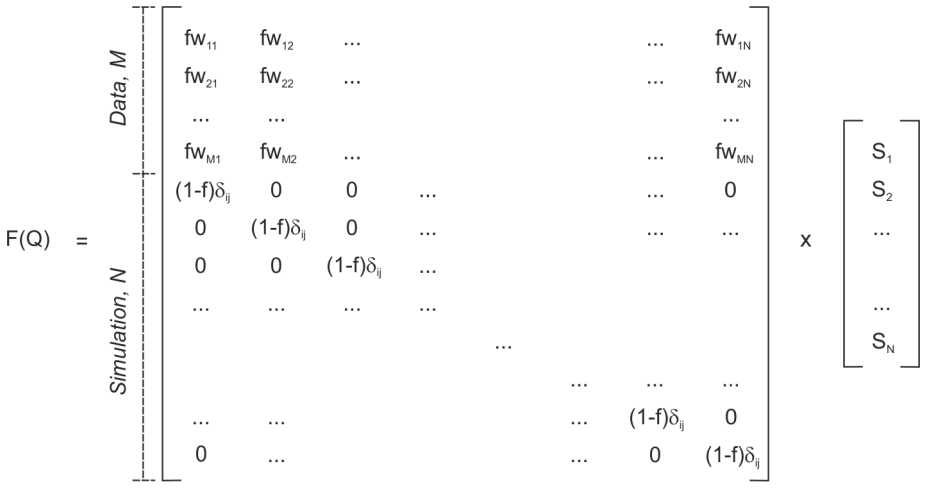


Figure 2.20: The equation for F(Q) in matrix form. The matrix for w_{ij} includes the contributions from the data and simulation, with their respective confidence factors. It is necessary to invert this matrix in order to obtains all S(Q)'s.

The process is repeated and the empirical potential increased until either i) the coefficients reach zero or ii) a good fit to the data is attained, without the introduction of any spurious information. In this thesis, all the simulations satisfy condition ii). When this condition is reached, the iterations are accumulated to gather statistics and obtain data that has sampled the full length of the simulation.

2.2.3 Auxiliary Routines

When the simulation is accumulating, it is gathering statistics on parameters such as F(Q), S(Q) for pairwise interactions and the pairwise RDFs. It is also possible to obtain information on other quantities using auxiliary routines that run alongside the simulation; analysing and accumulating the information from each iteration. Table 2.3 elicits the routines used for this thesis that will provide supplementary results in Chapters 3 - 5.

Table 2.3: Auxiliary routines used to provide atomic-level information on amino and imino acids in solution.

Routine	Description	Where routine used
COORD	Calculates the coordination number	Throughout 3, 4 and 5
CLUSTERS	Calculates the proportion of clusters of a given size	3.2, 3.5, 4.5.3, 4.6 and 5.5.1
PLOT3D	Calculates the spatial density function	3.2
SHARM	Calculates the spherical harmonic coefficients	3.2
TRIANGLES	Calculates the bond angle distribution	5.4.4 and 5.6.2
TORANGLES	Calculates the dihedral angle distribution	4.5.1 and 5.3

2.2.3.1 COORD

The COORD routine allows for the coordination number to be determined between a given pair of atoms. The coordination number is defined in Equation 2.30. The routine gives the average number of particles counted over a distance range specified by the user. For each atomic pair in the system included in this calculation, the derived value is histogrammed to produce a distribution of coordination number values. The final value is the mean of this distribution, $\overline{x}$, and the error is the root mean squared

deviation (RMSD) or 'standard deviation' of the population, given by the equation:

$$Standard\ Deviation = \sqrt{\frac{\sum (x - \overline{x}^2)}{N}} \tag{2.58}$$

Here, x is a given value calculated for the coordination number and N the number of coordination number values calculated.

2.2.3.2 CLUSTERS

The CLUSTERS routine is derived from the COORD routine and determines whether any molecules interact to form clusters. For two molecules interacting, they are defined at forming a cluster if the interaction of interest exists within a given distance range; the interaction and distance of which are specified by the user. The distance range for a clustering interaction would be equal to the minimum and maximum distance encompassed by the first coordination shell.

In this routine, the probability or proportion of clusters of a certain size through an interaction is obtained. The number of clusters containing i molecules, $M(i)$, as a fraction of the total number of clusters, M, to be plotted against the cluster size:

$$ClusterProbability = \frac{M(i)}{\sum M(i)} = \frac{M(i)}{M} \tag{2.59}$$

When detailing the interactions involved in clustering, it is possible to look at single or multiple criteria. By looking at the interactions separately, e.g. A and B, it is possible to look at the possibility of clustering via A or B. If the bond length criterion are listed together, it would determine the clustering via both interactions A and B.

2.2.3.3 SHARM and PLOT3D

The SHARM and PLOT3D routines are used to produce spatial density function (SDF) plots. An SDF portrays a 3-D distribution of the location of a particle β around α, with

respect to radial and angular position. It is also possible produce orientation coefficient function (OCF) plots using these routines, which describe the orientation β (or the molecule is within) takes when surrounding α, but no examples of these are shown within this thesis.

Figure 3.4 shows a schematic of the coordinate system for two particles in a system. The distribution and orientation of a molecule with respect to a set of coordinates is defined by Euler angles, $\omega = (\phi, \theta, \chi)$. Euler angles are described in more detail by Gray and Gubbins (195).

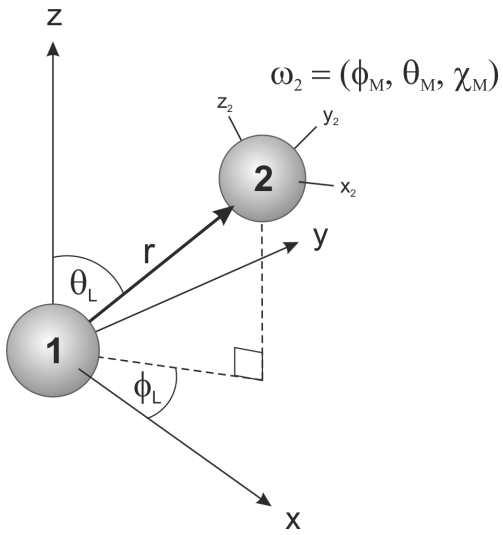


Figure 2.21: The coordinate system for a particle at a given orientation, represented by Euler Angles, (θ, ϕ, χ). The relation between particles 1 and 2 is here represented as the function of 6 variables; $r, \theta_L, \phi_L, \theta_M, \phi_M$ and χ_M.

The function for the plots produced using these routines is:

$$g(r_L, \omega_M) = 1 + \sum_{l_1 l_2 l} \sum_{m_2 m} \sum_{n_1 n_2} h(l_1 l_2 l; n_1 n_2; r) C(l_1 l_2 l; n_1 m_2 m) D^{l_2}_{m_2 n_2}(\omega_M) * D^{l}_{m_0}(\omega_L) \tag{2.60}$$

This function is a partially simplified form of the expression that comes from assuming that particle 1 is at the origin and particle 2 a distance r_L away from 1. The equation, instead of separate Euler angles for 1 and 2, incorporates the Euler angles ω_L that describes the relative position of 2 with respect to 1, and ω_M that describes the

relative orientation of 2 to 1. A more detailed explanation of the mathematical procedure used to obtain the spatial density function plot can be found in text such as (196) and (197). In Equation 2.60, $C(l_1 l_2 l; n_1 m_2 m)$ are Clebsch-Gordan coefficients, values arising in angular momentum coupling in quantum mechanics and D^l_{mn} generalised rotation matrices, describing transformations in the reference frame, which take the form of spherical harmonics, spatial functions that describe the surface of a sphere. Finally, $h(l_1 l_2 l; n_1 n_2; r)$ are coefficients, related to the structure of the system. These coefficients are calculated by SHARM, by inverting Equation 2.60, which can be used to derive the spatial functions for a specific pair.

In the input file for the routine, atoms α and β, the degree, l, the order, m, and the symmetry of the molecule or group the atoms are in, n, are specified. The coordinate axes are also defined for the atoms for which rotations will take place around.

PLOT3D analyses the results of SHARM to produce the final SDF. The input file for this routine includes the user-specified radial distance the function will present, as well as the angle to view the plot at. It is also possible to alter the percentage of atoms considered in the calculation, so the regions of highest probability can be probed.

Equation 2.60 is in fact a function of 6 variables $(r, \theta_L, \phi_L, \theta_L, \phi_M, \chi_m)$, which are all shown in Figure 3.4. To plot the function, it is necessary to reduce the number of variables. The variables that are excluded or fixed in the calculation determine the type of plot that will be produced. For an SDF, the aim is to produce a distribution of 2 around a central 1 particle that defines the coordinate reference frame. This requires varying the angles θ_L and ϕ_L, but averaging over all orientations of ω_M. Therefore, it is necessary to sum over the terms where $l_2 = m_2 = n_2 = 0$. Whilst in the SHARM files, values for particle 2 may be inputted as there are certain symmetry rules that apply for example when the molecules are identical, it can be specified which parameters are desired in the PLOT3D input file to obtain specifically the SDF.

2.2.3.4 TRIANGLES

The TRIANGLES routine calculates bond angles for three atoms, ABC. In the input file of the routine, the user specifies the miminium and maximum distances permitted for the bond lengths $\overline{AB}$ and $\overline{BC}$. The lengths for $\overline{AB}$ and $\overline{BC}$ are extracted from the RDFs, where the maximum distance (r_{max}) corresponds to the trough after the peak of interest, whilst the minimum distance (r_{min}) is based on a radial distance before the peak. These lengths may be intra- or intermolecular. For all triplets that satisfy the distance ranges specified in the input file, the routine calculates $\overline{BC}$ and calculates the bond angle using the law of cosines (Equation 2.40). The value is also divided by the sine of the angle to account for the uniform angle distribution. As for the CLUSTERS routine, by taking the ratio of the number of bond angles of a specific value versus the total number of angles, a distribution can be plotted of the probability or proportion versus the bond angle value, θ.

2.2.3.5 TORANGLES

For atoms A, B, C and D, all attached together as a chain, the dihedral angle is the angular deviance between these four atoms. If it is assumed that ABC and BCD sit on two separate planes, the dihedral angle equates to the angle between the two planes. This is shown in Figure 2.22.

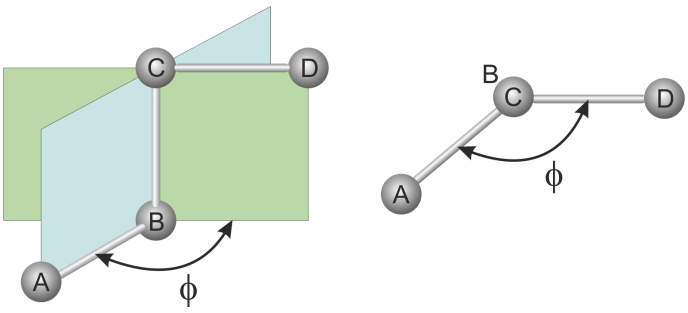


Figure 2.22: A schematic showing the definition of a dihedral angle. For a set of atoms $ABCD$, where ABC and BCD lie on two separate planes, the dihedral is the angle between these two planes, ϕ.

In the routine, by specifying $ABCD$, the angle is calculated by forming three line vectors, $\overrightarrow{AB}$, $\overrightarrow{BC}$ and $\overrightarrow{CD}$. The line vectors are transformed then into two plane vectors, $\overrightarrow{ABC}$ and $\overrightarrow{BCD}$ using the vector product. The vector product calculation for plane $\overrightarrow{ABC}$ is:

$$\overrightarrow{ABC} = \overrightarrow{AB} \times \overrightarrow{BC} = |\overrightarrow{AB}||\overrightarrow{BC}|sin\theta \tag{2.61}$$

Here, θ is the angle between the two vectors. The scalar product is used to obtain the cosine of the dihedral angle between the planes:

$$\overrightarrow{ABC}.\overrightarrow{BCD} = |\overrightarrow{ABC}||\overrightarrow{BCD}|cos\phi \tag{2.62}$$

With the cosine of the dihedral angle, ϕ, being:

$$cos(\phi) = \frac{\overrightarrow{ABC}.\overrightarrow{BCD}}{|\overrightarrow{ABC}||\overrightarrow{BCD}|} \tag{2.63}$$

The scalar product produces a scalar quantity and does not consider the direction of the dihedral angle. The vector product, as defined in Equation 2.61 can be used to determine the direction instead. The direction will depend on the order which the atoms were listed in the TORANGLES input file. Like the other routines, the ratio of the number of dihedral angles of a given value over the the number of dihedral angles recorded gives the proportion that can be plotted against ϕ.

2.3 Small Angle Neutron Scattering

To supplement information on the solution of L-glutamine, additional SANS data was taken for understanding amino acid association. The results of this experiment are presented in Chapter 3. SANS experiments are optimised to study structures in solution in the range of tens to thousands of Ångstroms and thus is sensitive to bulk structure, rather than the atomic components of the system.

In Section 2.3.1, the key SANS theory for the experiments completed is covered. This builds upon the information provided in Section 2.1.3 on obtaining structural information from scattering data. In Section 2.3.2, the instrument used and the parameters utilised for this experiment will be presented.

2.3.1 The Theory of SANS

Referring back to Bragg's Law (Equation 2.1), it is possible to probe large distances if either the wavelength is increased or the angle is decreased. As extremely long-wavelength neutrons or copious fluxes are not currently attainable, it is necessary in such an experiment to probe the scattering at small angles. This is achieved by increasing the separation between the sample and detector (198).

Like a neutron diffraction experiment, the aim of a SANS experiment is to determine the differential cross-scattering section of the sample. For a SANS experiment, this can be defined by the following equation (199):

$$\frac{d\sigma}{d\Omega} = N_p V_p^2 (\Delta\delta)^2 P(Q) S(Q) \tag{2.64}$$

Here, N_p is number concentration of particles, each of volume V_p. The terms, $(\Delta\delta)^2$ and $P(Q)$ are described in more detail below and represent the contrast and form factor respectively. $S(Q)$ is the structure factor, the same as described in section 2.1.3. For dilute solutions though, $S(Q)$ tends to 1. Equation 2.64, the absolute differential cross-section, is also defined as the intensity of the incoming signal, $I(Q)$.

The contrast term, $(\Delta\delta)^2$, is derived from the scattering length density of a substance, δ. The scattering length density describes the strength of the interaction between a neutron and the nuclei of the sample. This value can be determined for a given sample using the expression:

$$\delta = \sum_i b_i \frac{D N_A}{M_W} \tag{2.65}$$

In this equation, the sum of all coherent scattering lengths, $\sum_i b_i$, is calculated. The concept of the scattering length, b, was presented previously in Section 2.1.2.4, with example values for hydrogen and deuterium provided in Table 2.2. D is the density of the sample being studied, N_A is Avogadro's constant[1] and M_W the molecular weight of the scattering body. From this, $(\Delta\delta)^2$ equals the difference between the scattering length densities of the sample, δ_p, and the solvent, δ_m, squared:

$$(\Delta\delta)^2 = (\delta_p - \delta_m)^2 \tag{2.66}$$

The form factor, $P(Q)$ is comparable to that specified in Section 2.1.2.2 and describes how the cross section varies according to the shape of the sample. The form factor can be described more generally by the van de Hulst equation (200):

$$P(Q) = \frac{1}{V_p^2}\left|\int_0^{V_0} e^{if(Q\alpha)}dV_p\right| \tag{2.67}$$

With α denoting a shape parameter indicative of the radius of gyration or length of the sample. For neutron diffraction, the form factor is redundant as the atomic components, the nuclei, are of a similar size to the incoming neutrons. A form factor is of relevance for a diffraction experiment if X-rays are being used, with the relationship for this shown in Equation 2.15. As a SANS experiment detects the sample in bulk, the form factor is important again to assess the shape of the sample (201–206). The form factor can be adapted to suit specific sample geometries, which are derived from mathematical models. The form and structure factor can be determined by fitting the data to various quantitative models until one is reached that best fits the data. There are many software packages available that support such analyses (205; 207; 208). For the SANS study presented in this thesis, on aqueous L-glutamine, it was only necessary to look at the raw $I(Q)$ data to examine molecular association and the degree of clustering in the solution.

[1]The value for Avogadro's constant can be found in the *Abbreviations* section at the start of this thesis

2.3.2 The LOQ Instrument and the SANS Experiment

SANS measurements have been completed using the LOQ instrument at ISIS. A schematic of the LOQ instrument is shown in Figure 2.23. This is a fixed-geometry, white beam, time-of-flight instrument that utilises neutrons with wavelengths between 0.2 and 1 nm. Data is simultaneously recorded on two, two-dimensional, position-sensitive, neutron detectors. At a frequency of 25 Hz, the instrument's main area detector is able to obtain readings in a Q range of 0.006 - 0.24 $Å^{-1}$, corresponding to length-scales of 26.2 - 1047.2 Å. To achieve the small angles required, the area detector is located 4.05 m away from the sample. LOQ also has a high-angle detector, located 0.5 m away from the sample, which records scattering at angles up to $\sim 35°$. This allows a Q range of 0.15 - 1.4 $Å^{-1}$ to be studied, which overlaps with that studied by the SANDALS instrument. The r range covered by the high-angle detector is 4.5 - 41.9 Å.

The main detector is a ^{3}He-CF_4 filled proportional detector. Incoming thermal neutrons will react with ^{3}He to allow the following interaction to take place:

$$^3He + n \rightarrow {}^3H + p + 765\,\text{keV} \tag{2.68}$$

This produces protons (p), that can interact further with ^{3}He cumulatively to produce pulses (209). The energy produced is dissipated by the CF_4 quench gas. The electric field in the detector directs the charged particles towards the negative cathode for detection. The detector components are made from Aluminium, which has a small neutron cross-section (186). The high-Q bank possesses scintillation detectors that function in the same way as those of the SANDALS instrument, as described in Section 2.1.4.6.

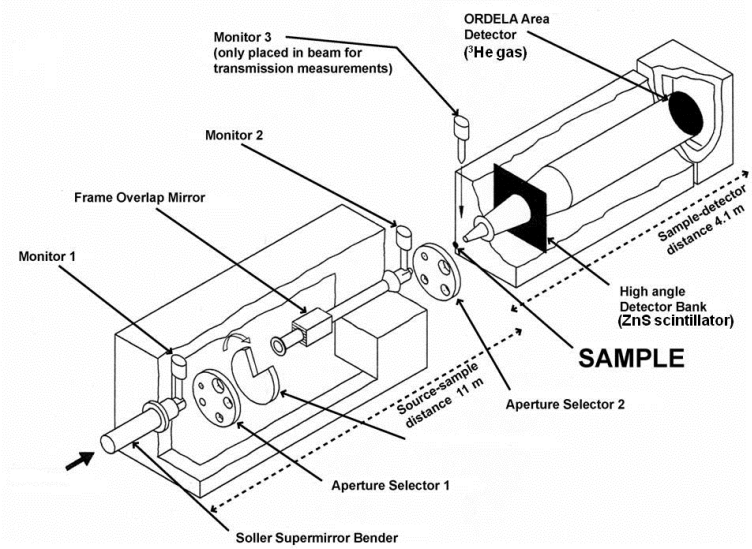


Figure 2.23: A schematic of the LOQ instrument used to complete SANS measurements. LOQ can be used to probe structures spanning a Q-range of 0.006 Å to 1.4 Å, if both main area and high-angle detectors are utilised. Picture taken from (210) and (211).

The sample was placed in a 1 mm path length quartz cuvette and was measured for a total of 1.5 hours under a given set of conditions in order to gather data with statistical precision. As a SANS Xpress measurement, the experiment run time is shorter and thus the noise in the data was not completely eliminated. Each raw scattering dataset was then corrected for the detector efficiencies, sample transmission and background scattering and converted to the absolute scattering cross-section using the instrument-specific software, COLETTE (212; 213). Each dataset was then converted to an absolute scale (cm^{-1}) using the scattering from a standard sample in accordance with established procedures (214). The standard sample is a solid blend of hydrogenous and perdeuterated polystyrene (214).

2.4 Methods Summary

This chapter covered the theory and motivation for using neutron diffraction and computational modelling for the work in this thesis. Section 2.1.1 described the theory behind neutron diffraction. Section 2.1.2 explained the reasons for selecting neutrons over other diffraction probes. Section 2.1.3 detailed the information that is possible to obtain from

a diffraction experiment. Section 2.1.4 described how neutrons could be produced and detected, as well as discussing the reasons for using the SANDALS instrument for experiments. Section 2.2 covered the theory behind the EPSR computational modelling completed in order to obtain atomic level data. Section 2.3 presented the theoretical and instrument details for the SANS data completed to supplement select diffraction results. Chapters 3 - 5 will report the results obtained on the aqueous properties of amino and imino acids from exploiting the methods detailed in this chapter.

Chapter 3

Initial Studies on Dilute Aqueous L-Glutamine

The aim of this thesis is to take a bottom-up approach to understanding the properties of peptoids by first obtaining structural information about their individual imino acids. The molecule that will be studied is the glutamine imino acid (Ngln), first introduced in Section 1.3.2. This is an original approach, as no comprehensive study on this imino acid has previously been undertaken. To allow a detailed understanding of the imino acid it is imperative to compare it to its counterpart amino acid, L-glutamine. Given the interest in studying the imino acid, it is important to obtain quantitative information also on both the amino and imino acids. This chapter will detail the results of neutron studies completed on the amino acid L-glutamine.

As introduced in Section 1.3.1, glutamine is thought to play a role in metabolism, as well as in neurodegenerative disease when in homopolypeptide form. Its role is attributable to the interactions its functional groups form, as despite being considered as hydrophilic on many hydrophobicity scales (215–218), it has a propensity to aggregate, potentially via interactions involving the polar groups it contains. Research into polyglutamine aggregation has suggested that glutamine residues will bond via a range of interactions. Spectroscopy on Ataxin-3 has suggested that side chain-side chain interactions are a major driving force in polyglutamine aggregation (219), whilst backbone-side chain interactions have been observed in molecular dynamics simulations for these se-

quences, where the side chain nitrogen atoms interacts with the lone pairs of electrons on backbone oxygen atoms (220). Whilst a study of single, zwitterionic L-glutamine cannot comprehensively gauge the most important interactions relevant to disease, with only the amide side chain common to both L-glutamine and glutamine-homopolypeptides, it can be used as a first step to highlight the important kinds of interactions in glutamine-based systems, namely the balance of hydrophilic versus hydrophobic interactions. As the solubility limit for L-glutamine is low compared to many other standard amino acids (48), the propensity of glutamine to associate is still present in single amino acid form. As neutron diffraction aims to obtain a time-averaged view of a system, it is not possible to monitor the aggregation of such molecules using this method. What can be monitored instead are the interactions that are initially involved in association, prior to aggregation of biomolecules.

This chapter in Section 4.1 will first describe the neutron diffraction experiments completed on L-glutamine. Given the low concentration of L-glutamine solution studied, this work probes the attainable information attained with neutron diffraction on this system. After presenting initial results for the molecule in Section 3.2, Section 3.3 will examine limitations with the methodology utilised. It will then present modifications which were employed to improve the method. This new method is then applied to re-examine the L-glutamine data in Section 3.4 and compare the impact of the modifications, as well as to study the effects of temperature in Section 3.5. Finally, Section 3.6 will present small-angle neutron scattering data to support the results obtained in this chapter.

3.1 Experimental Procedure

Solutions of L-glutamine were studied using neutron diffraction at approximately 30 mg ml^{-1}, equating to a molar concentration of 0.205 M. Whilst the concentration is low for neutron studies, it is approaching the solubility limit of the amino acid (48).

For L-glutamine, four isotopic substitution experiments were completed, as shown in Table 3.1. These experiments study both protiated (Sigma Aldrich, purity $\leq 99\,\%$) and partially deuterated (Cambridge Isotope Laboratories, purity 98 %) L-glutamine. For the partially deuterated L-glutamine, the 5 hydrocarbon hydrogens have been substituted for deuterium. These L-glutamine samples were studied in milli-Q water and deuterium oxide (Sigma Aldrich, 98 %) at a temperature of 24 °C and standard pressure of 1 bar.

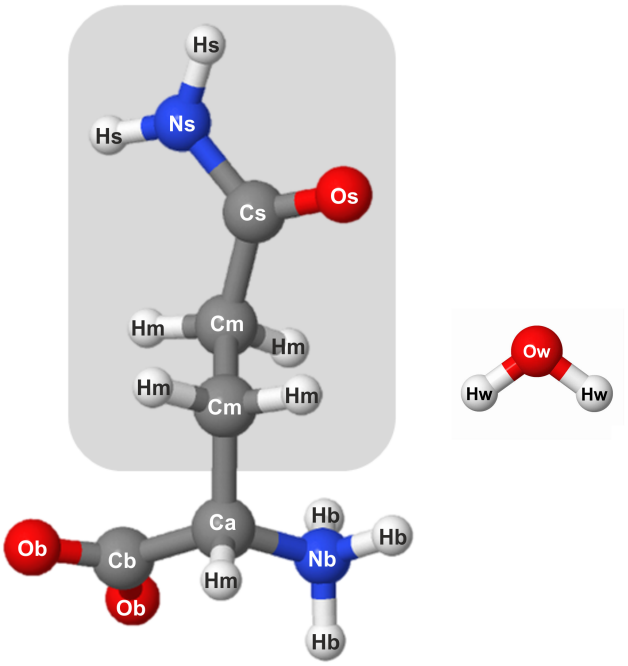
Table 3.1: Isotopic substitution experiments completed for L-Glutamine (Gln).

Sample No.	Sample Name	Description
i	Gln-D_2O	Fully protiated Gln in D_2O
ii	Gln-H_2O	Fully protiated Gln in Milli-Q water
iii	Gln-D5-D_2O	Gln with deuterated hydrocarbon groups in D_2O
iv	Gln-D5-H_2O	Gln with deuterated hydrocarbon groups in Milli-Q water

3.2 Initial Studies of L-glutamine at 24 °C

In this initial structural study (1), a box of 25 L-glutamine molecules and 6725 waters was constructed in EPSR. The molar ratio of this system is thus 1 L-glutamine to 269 water molecules. The structural parameters for L-glutamine in this simulation; bond lengths and angles, as well any dihedral angles, were defined according to those found in its crystal structure (221). Crystal structures have been used in earlier neutron diffraction and EPSR studies of amino acids in solution (127; 128). OPLS potentials for peptides were used for the LJ potentials and Coulombic charges (91), the same as for EPSR investigations on L-glutamic acid (127) and glutathione (141). For water, LJ potentials and Coulombic charges were defined according to SPC/E rules (93).

In the simulation, there are 11 distinct atomic components defined for L-glutamine. These components are labelled in Figure 3.1. The hydrogen atoms in both simulations are labelled as Hb, Hm and Hs, referring to hydrogens on the backbone ammonium, hydrocarbon groups and the side chain amide respectively. The oxygen atoms are Ob, for the two found on the backbone caboxylate anion and Os for the one on the amide group. Nb refers to the ammonium nitrogen and Ns the amide nitrogen. For the carbon atoms, Cb is that found on the carboxylate, Ca the α-carbon and Cm for those on the methylene groups. There are also two components for water in each simulation, Ow and Hw for the water oxygen and hydrogen atoms respectively.



(a) (b)

Figure 3.1: Diagram of the (a) L-glutamine amino acid, as well as (b) water. The side chain of the molecule is highlighted in grey. C=O double bonds are not shown in this diagram. The labels are those used to identify each of the atomic components in the EPSR simulations. There are 11 distinct atomic components in L-glutamine and 2 for water.

The total structure factors, $F(Q)$, for this simulation are shown in Figure 3.2. The simulation was accumulated for $\sim$ 10000 iterations. The quality of the fit of the simulation to the data attained can be assessed using the 'R-factor', defined by the equation:

$$R = \frac{1}{M}\sum_{i}\frac{1}{n_Q(i)}\sum_{Q}[D_i(Q) - F_i(Q)]^2 \tag{3.1}$$

For the case where R is the R-factor, M is the number of datasets, D_i the experimental structure factor for a given dataset i, $F_i(Q)$ the simulated structure factor and $n_Q(i)$ the number of Q values for the i^{th} dataset. The closer the R-factor is to zero, the better the fit. For the simulations, L-glutamine has an R-factor of 2.3 $\times$ 10^{-3}. These values suggest a good fit to the data has been achieved for the simulation.

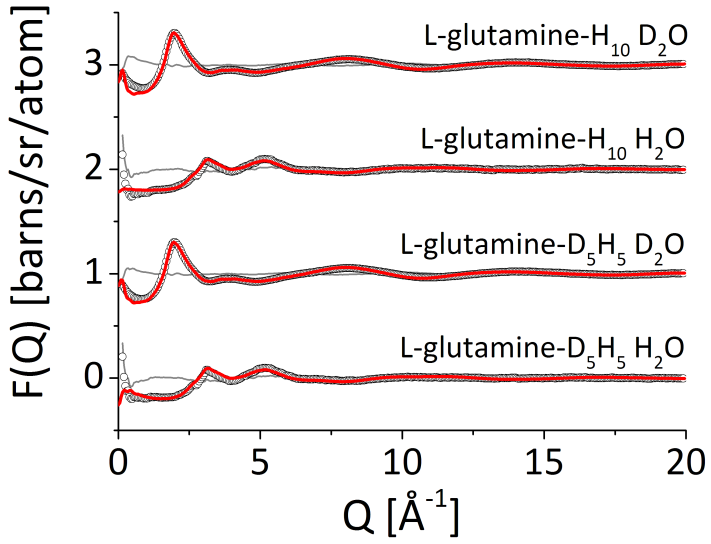


Figure 3.2: The fit of the total simulated structure factor, $F(Q)$, to the experimental data, $D(Q)$, for initial studies of L-glutamine. The data is shown by the black circles, whilst the fit is shown by the red lines. The difference between data and simulation for each dataset, $D(Q) - F(Q)$ is shown by the grey line. Each dataset has been shifted by 1 for clarity.

In the process of fitting the simulation to the data, it is expected that that deuterated samples will fit better. For protiated samples, as is evident in the data, the fit at low Q is poorer due to residual inelastic effects. For H_2O for instance, the scattering length is close to zero and this inhibits the ability to observe density fluctuations in the sample

when fitting, as the scattering length contribution is negligible. This means it is difficult for Gudrun to remove this information in the low Q region as there is nothing meaningful to fit to.

Any signal from clustering or bulk structure would be expected to appear in the low-Q region, due to the larger size of the structures. For biomolecular clustering also, the information is coming from the simulation as opposed to the data, as these interactions are less discernible in the data compared to biomolecule-water or water-water interactions. The information on rarer events or low-Q region is instead constrained by the appropriate fit to high-Q regions, where the information is more discernible. This also highlights the importance of having a suitable reference potential to describe the biomolecular interactions, as it is this which is determining the nature of the interactions in the system.

When presenting radial distribution function (RDF) coordination shell peak and trough positions for the pairwise interactions, for this study and throughout this thesis, these are specified to 0.01 Å or femtometre accuracy. This level of accuracy is comparable to that used in previous studies of biological and organic molecules (101–105; 127; 128; 130), as well as distances specified in higher-resolution crystallographic studies (221; 222). Section 1.2.1 explained that an RDF is calculated by forming concentric spheres around a particle and counting those in each shell of length δr. The Δr separation in EPSR is to femtometre level, making the specified positions comparable. It is important to iterate, as first stated in Section 2.2, is that the results given may not be the only answer, but one that fit the experimental neutron diffraction data. There is the possibility for fluctuation, though to accurately determine the error for this would require repeating the whole EPSR simulation many times, which would be time-consuming and unlikely to improve the results. Even given the accuracy of the distances, the distances specified are indicative, not absolute and serve as a guide to possible trends seen in the system.

The relevant RDFs for hydrophilic functional groups with water are shown in Figure 3.3, with first coordination shell numbers and distances presented in Tables 3.2 and 3.3. The results obtained from this study highlight that L-glutamine plentifully hydrogen

bonds with its surrounding aqueous environment. A carboxylate oxygen for instance is able to interact with 2.9 water molecules at a given time, as elicited by the Ob-Ow coordination, a value similar to that observed for other amino acids such as L-proline (128). On the terminal ammonium, each hydrogen atom, Hb, can coordinate 1 water molecule. The terminal groups possess a partial charge, coming from an imbalance between the partial charges of the separate atoms. What is also suggested in the results is that the carbon atom of the carboxylate may also coordinate water molecules too. This pattern has also been seen for backbone groups of L-glutamic acid (127) and L-proline (128) by neutron diffraction, as well as for nitrogen groups on neurotransmitter dopamine (138).

These hydrogen bonding patterns can also be visualised using spatial density function (SDF) plots (Section 2.2.3.3). For the carboxylate group (Figure 3.4(a)), the SDF represents 25% of oxygen water atoms in the Cb-Ow pairwise function, taken up to the peak position of the first coordination shell (2–3.56 Å). By probing only 25%, this plot gives information on the most probable location of water molecules around the group. Whilst density can be found over the oxygen atoms, it is also possible to find it in between, indicating the carbon may also coordinate water molecule, due to the partial charge on the carbon atom. Such distributions are also present for other amino acids, including L-glutamic acid and L-proline, which possess this same carbonyl backbone. For the backbone ammonium group (Figure 3.4(b)), which portrays the most likely position (25%) of water molecules within the entire first coordination shell (2–3.42Å), the function suggests hydrogen bonding predominantly occur through the ammonium hydrogen atoms.

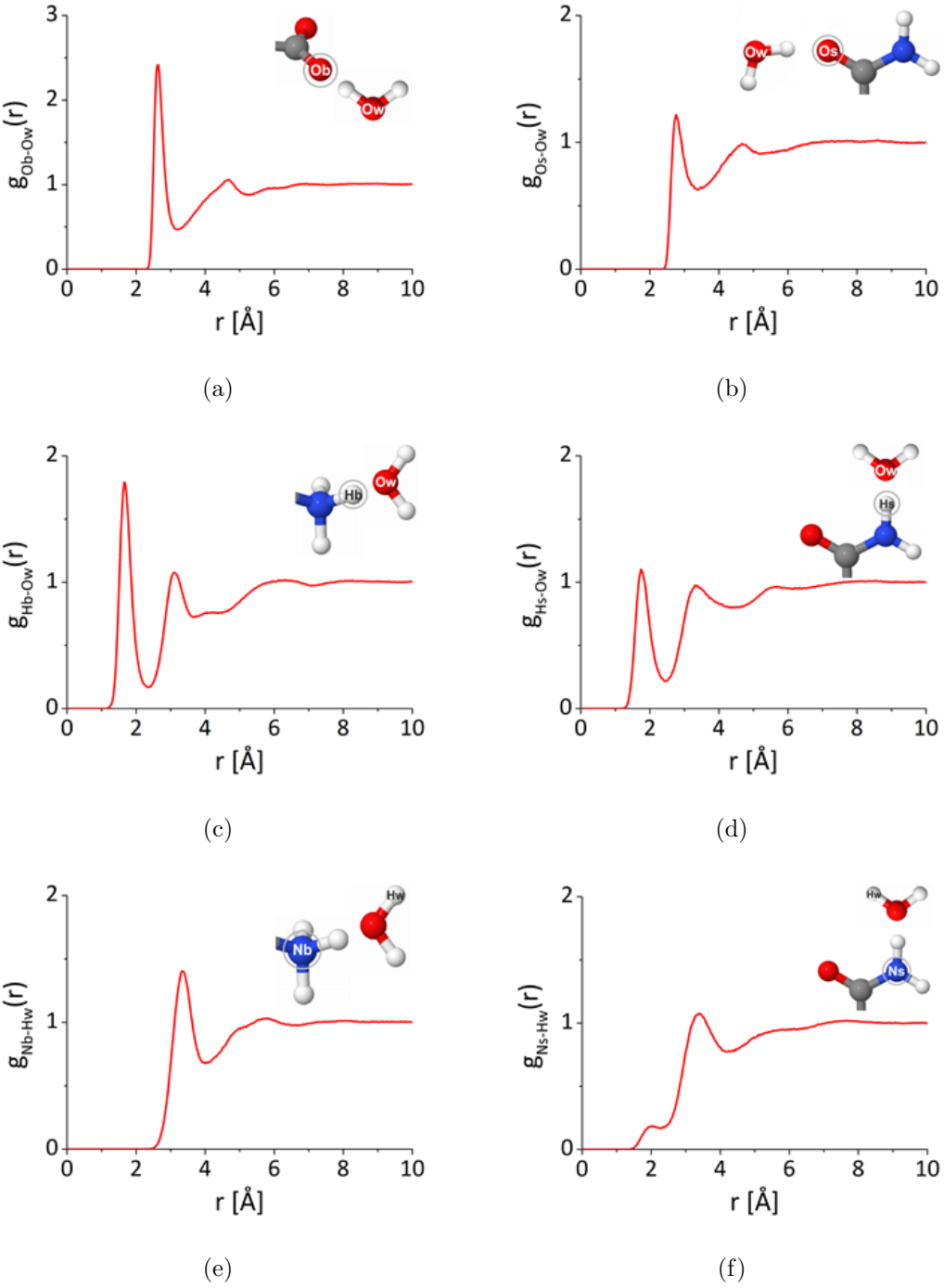


Figure 3.3: The pairwise RDFs for glutamine-water interactions for initial studies of aqueous L-glutamine. The functions shown are for the (a) backbone oxygen - water oxygen (Ob-Ow), (b) side chain oxygen - water oxygen (Os-Ow), (c) backbone hydrogen - water oxygen (Hb-Ow), (d) side chain hydrogen - water oxygen (Hs-Ow), (e) backbone nitrogen - water hydrogen (Nb-Hw) and (a) side chain nitrogen - water hydrogen (Ns-Hw) interactions.

Table 3.2: Coordination numbers for interactions of hydrophilic-based oxygen and hydrogen atoms with water oxygen atoms.

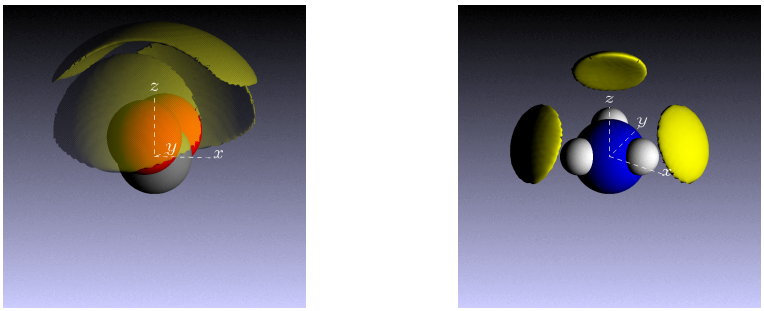
Atom Label	r_{min} (Å)	r_{max} (Å)	Peak Position ($\mathring{A}$)	Coordination Number	Standard Deviation
Ob-Ow	1	3.21	2.63	2.9	0.7
Os-Ow	1	3.36	2.76	2.5	0.9
Hb-Ow	1	2.35	1.66	1.0	0.3
Hs-Ow	1	2.45	1.74	0.9	0.4

Table 3.3: First coordination numbers for interactions of hydrophilic-based nitrogen atoms with water hydrogens.

Atom Label	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Nb-Hw	1	3.99	3.35	10.7	1.8
Ns-Hw	1	2.23	2.03	0.3	0.5
Ns-Hw	2.23	4.20	3.39	13.4	2.2

The side chain amide can coordinated water molecules through its carbonyl oxygen and its amine hydrogen atoms. A lone pair present on the amide group allows for extra coordination with the amide group. This is derived from the lone pairs and due to sp^3-hybridisation, the electrons are conjugated over the amide group (6). More information is given in Section 3.3. Its influence is apparent in RDFs such as that for Ns-Hw (Figure 3.3(f)), where a hydrogen bonding peak is present at 2.03 Å. However, with a coordination number of only 0.3, as shown in Table 3.2, it contributes very little to the overall hydration state of the amide group. Coordination numbers for the backbone and side chain groups suggest that backbone groups coordinate more water molecules than side chain groups. Though both the backbone and the side chain are involved in hydrogen bonding, the partial charge present on backbone groups facilitates

interactions between the solute and the solvent. Whilst L-glutamine does form multiple hydrogen bonds with water, this has little influence on water structure, due to the low concentration. Other systems known to induce very little impact on water structure include certain dipeptides containing hydrophobic moieties (130), where coordination numbers do not vary substantially compared to bulk water.



(a) (b)

Figure 3.4: Spatial density function plots for L-glutamine, showing the most likely position of water molecules around these functional groups. The functions shown are for (a) 25% of water molecules up to the peak position of the carboxylate carbon - water first coordination shell (Cb-Ow, 2–3.56Å) and (b) 25% of water molecules in the ammonium nitrogen - water first coordination shell (Nb-Ow, 2–3.42Å).

In relation to its role in disease, L-glutamine molecules are thought to aggregate in aqueous solution. This can be visualised with a snapshot of the box where the biomolecules can be seen to associate in solution, though the dominant species for the molecule to exist in is as a monomer.

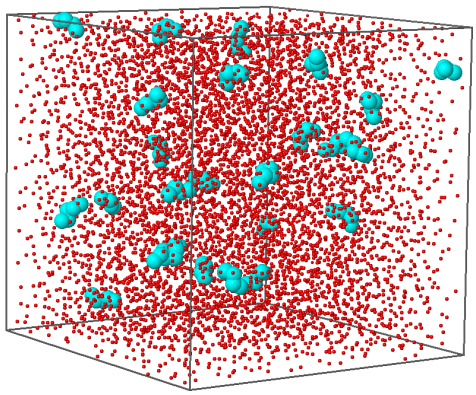


Figure 3.5: Snapshot of the simulation box for initial studies of L-glutamine, with water molecules represented by their oxygen atoms (red) and L-glutamine molecules by their carbon and nitrogen atoms (blue). The snapshot suggest L-glutamine can self-associate in solution.

Glutamine association has been studied through each of its polar groups. To be consistent with previous EPSR work on biomolecular association (131), the hydrogen bond interaction between an amide or carboxylate oxygen and an amide or ammonium hydrogen was studied. This allows for four types of interaction to be studied, backbone - backbone (Ob-Hb), backbone - side chain (Ob-Hs), side chain - backbone (Os-Hb) and side chain - side chain (Os-Hs). Two glutamines are defined as interacting if they are within the first coordination shell of the interaction. The RDFs for these are presented in Figure 3.6, which suggest that L-glutamine can interact with like molecules via all four interactions. In the neutron diffraction crystal structure of L-glutamine, all four types of interaction were also observed (221), with a complex hydrogen-bonding network being observed from oxygen atoms interacting with all five hydrophilic hydrogen atoms in the one unit cell.

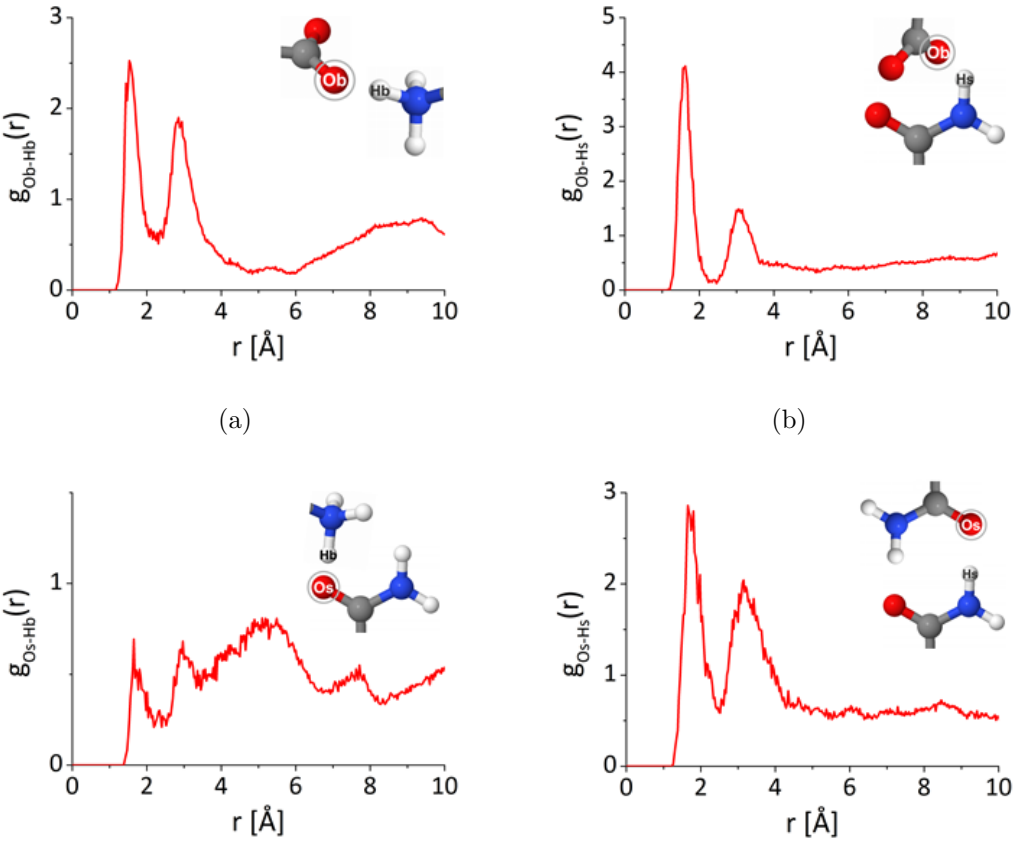


Figure 3.6: The pairwise RDFs for hydrophilic glutamine-glutamine interactions for initial simulations of L-glutamine. The four interactions are (a) backbone - backbone (Ob-Hb), (b) backbone - side chain (Ob-Hs), (c) side chain - backbone (Os-Hb) and (d) side chain - side chain (Os-Hs).

Glutamine-glutamine interactions can be quantified in terms of glutamine clusters, where separate L-glutamine molecules have associated to form aggregates containing a certain number of molecules. Clustering studies have been completed to determine the size of clusters formed by each hydrophilic interaction. Details of the CLUSTERS routine can be found in Section 2.2.3.2 and the resulting cluster distribution is shown in Figure 3.7. In this solution, L-glutamine typically exists in monomeric form. However, it is possible for a small population of dimers to form via Ob-Hb, Ob-Hs and Os-Hs interactions. Whilst many of the interactions are not applicable to disease, it does suggest the presence of side chain - side chain interactions, an interaction possibly thought to stabilise the structure of polyglutamine aggregates (219). However, the limitation

with this study is that it is dependent on sampling only 25 L-glutamine molecules. In Section 3.5, these interactions will be studied in greater detail using a larger simulation box to ascertain the scale of structures being produced from hydrophilic interactions with varying temperature.

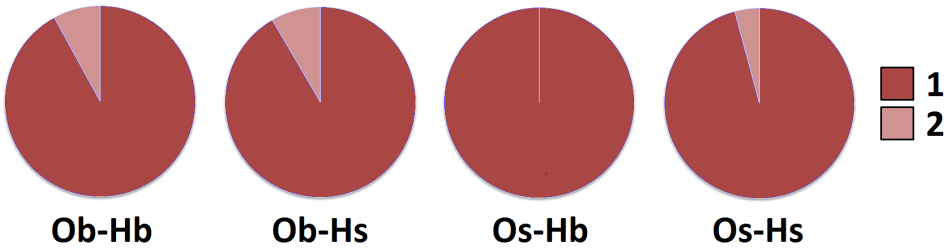


Figure 3.7: Percentage of L-glutamine molecules existing as monomers (1) and in hydrophilic intermolecular dimers (2).

3.3 Modifications to the EPSR Methodology

Section 3.2 summarises the results from initial EPSR simulations of aqueous L-glutamine. Over the course of this project, a number of important improvements have been made to the neutron diffraction software, Gudrun, and to the initial models used to describe the molecules. The following subsections will discuss these improvements.

3.3.1 Gudrun Updates

Updates have been made to Gudrun, with newer versions of the software including an automated correction routine that iteratively refines data until inelastic and multiple scattering effects have been removed (223). More details of the new version of Gudrun are described in Section 2.1.4.7. This improvement helps to reduce the bias possible from completing corrections manually. The data utilised in subsequent chapters is corrected using this new iterative method.

3.3.2 Dihedral Angle Constraints

The concept of a dihedral angle was introduced in Section 2.2.3.4. Within the simulation, it is possible to define dihedral angles, though fixing all of these angles may not always provide the most reliable picture of the system. Many EPSR simulations of biomolecules do not incorporate dihedral angles (127; 128), with the exception of groups that need to keep a specific conformation, including the trans-isomer backbone in peptides (139–141) and planar aromatic groups (118; 138). Investigations into the most probable conformation was completed for studies of glycerol (101; 103). For pure glycerol, the $\alpha\beta$ dihedral conformation was found to produce the best fit to the experimental data, eliciting the necessity to incorporate these parameters into simulations. The necessity for dihedral angles appears to be dependent on the species in question and the concentration of the molecule in solution. Another group that maintains planarity is the backbone carboxylate, due to sp^2 hybridisation of electrons.

In the initial studies, L-glutamine was modelled using a semi-flexible structure, with dihedrals being constrained along the carbon and nitrogen atoms of the molecule. Due to sp^3 hybridisation of the amide group, it is expected that this group should also be constrained to be planar, so as all atoms are positioned along a single plane (6), though NMR studies show that some level of rotation of this group with respect to the rest of the molecule is present in proteins (224).

The studies in Section 3.2 suggested that L-glutamine was very flexible. An example dimer formed between side chain groups is shown in Figure 3.8, where the nitrogen and hydrogen atoms have rotated out of plane. Therefore, the approach has been modified to constrain all dihedral angles in the structure to allow the molecule to maintain a more plausible structure.

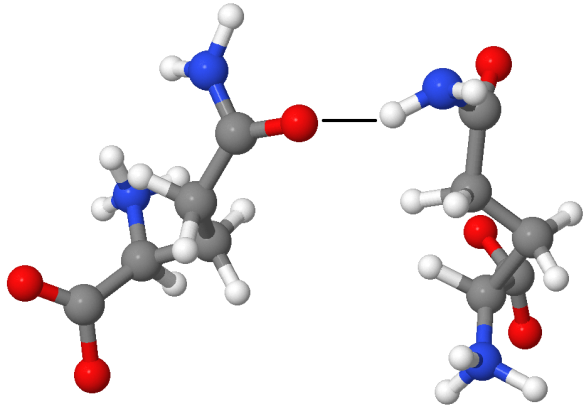


Figure 3.8: Side chain - side chain clustering observed in initial simulations of L-glutamine. It is evident that the $-NH_2$ portion of the amide for the right-hand L-glutamine is not planar to the amide group.

The reason to constrain more dihedral angles comes from the way they are defined in the simulation. In Section 2.2.1.1, the intramolecular potential that defines geometrical parameters was presented, with Equation 2.37 being used to determine parameters within EPSR. Like bond angles, where a pseudo bond is used to define the angle instead of it being calculated explicitly as described in Section 2.2.1.1, the same applies for dihedral angles. This is shown schematically in Figure 3.9, where EPSR determines the distance between atoms A and D and creates a pseudo bond to represent the dihedral angle. As the simulation classifies the dihedral as being satisfied only if this bond length is met, it is possible that more than one dihedral angle could orient the molecule in such a way to fit the $\overrightarrow{DA}$ bond length. The more flexible the molecule, the more possible angles that could be probed, which could satisfy the pseudo bond constraint. By fixing more dihedral angles in the molecules, fewer conformations can be probed, which in turn lessens the impact on the final dihedral angle.

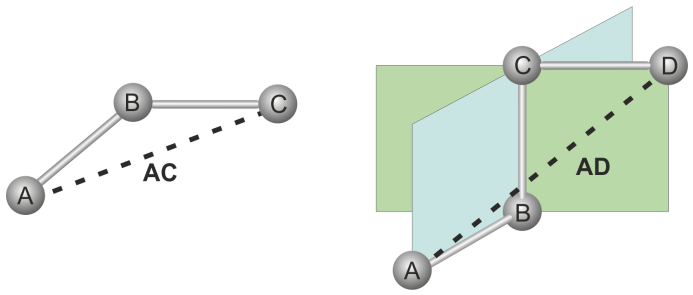


Figure 3.9: Angle and dihedral angle constraint in EPSR, where a pseudo bond is used to constrain the molecule. The distance between atoms A and C is calculated for angles and the distance between atoms A and D is calculated for dihedral angles.

If there is conformational information in the data, a flexible model may produce a better fit. EPSR uses information from the data to inform the simulation on the structure, given those sampled in the simulation are present in the data. Due to the low concentration, it is clear that information on the molecular conformation is not present in the data. Whilst for higher concentration simulations it may be possible to test the effects of fixing dihedral angles, all dilute, $30\,\mathrm{mg\,ml^{-1}}$ simulations have all dihedral angles fixed.

The simulations presented in this thesis use the stable EPSR18 version of the software. Very recent developments have taken place to improve the software, though due to time constraints it has not been possible to thoroughly analyse all of the systems studied in this thesis using newer versions of the software. More information on the most current version, EPSR24, is presented in Chapter 6. One of the updates has been to improve the way dihedral angles are constrained by directly calculating the angle between planes. In Section 3.5, the newer EPSR24 has been utilised to probe hydrophilic interactions at higher temperatures.

3.3.3 Model Modifications for Molecules

In the initial studies, an SPC/E model was used to characterise the molecular geometries of water, as well as the Lennard-Jones and Coulombic potentials (93). This is similar to work on aqueous glycerol, which uses a SPC model to define all potentials and geometries (102–105). The geometries for SPC and SPC/E are the same (93).

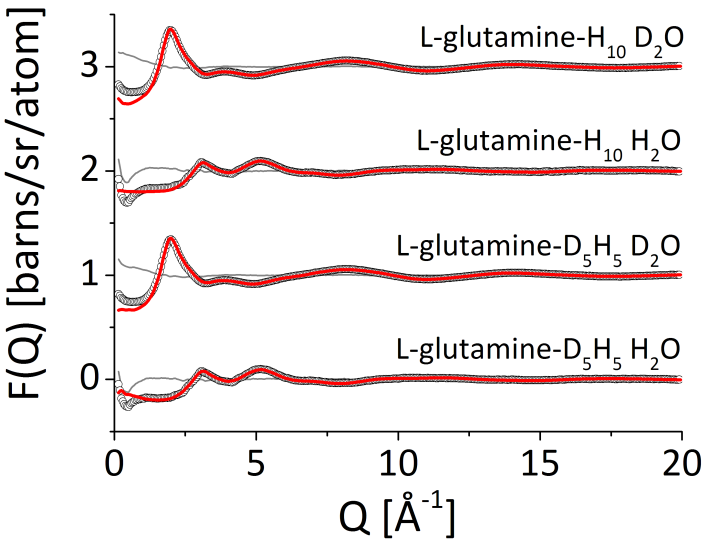
We have modified this approach by using an adapted geometrical version of water, which is tailored to better fit and represent the neutron diffraction data. The bond angle has been changed to 104.5 ° and the bond length to 0.976 Å, also used in a recent study of pure water published by Soper in 2013 (223).

For subsequent L-glutamine simulations and for the other biomolecules presented in this thesis, the structures were refined using the semi-empirical quantum chemistry software MOPAC-7 (225), using the AM1 Hamiltonian (87). The structure attained form this model was used to define the intramolecular bond angles, lengths and dihedrals. MOPAC-7 automatically lists all these parameters, reducing systematic error of extracting parameters from literature coordinates.

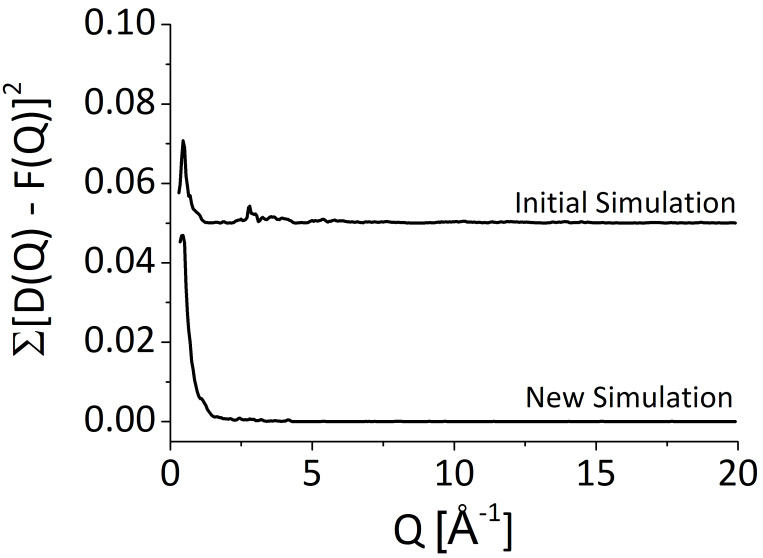
3.4 Repeat Studies of L-glutamine at 24 °C

A repeat simulation has been completed for the L-glutamine system that incorporates the modifications described in Section 3.3. The simulation box contains 25 L-glutamine molecules and 6725 waters, like that used in initial studies of L-glutamine (Section 3.2) and was run for $\sim$ 10000 iterations. This simulation will be compared to that for Ngln in Chapter 4. These parameters are tabulated in Appendix B.

The fits to the experimental data are shown in Figure 3.10(a), with the R-factor being 3.8 $\times$ 10^{-4}, lower than the 2.3 $\times$ 10^{-3} attained in initial studies. This suggests that the improvements made have improved the fit to the data, though most is due to changes to the data. Within Equation 3.1, the value of $\sum_{Q}[D_i(Q) - F_i(Q)]^2$ is known as the sum of residual squares and can be plotted to give a measure of the deviance between data and fit over the Q-range studied. $\sum_{Q}[D_i(Q) - F_i(Q)]^2$ has been plotted in Figure 3.10(b) for the initial and repeat simulations. This plot shows that a better fit is obtained also in the Q-range of 2 - 5 Å.



(a)



(b)

Figure 3.10: (a) The fit of the total simulated structure factor, $F(Q)$, to the experimental data, $D(Q)$, for repeat studies of L-glutamine. The data is shown by the black circles, whilst the fit is shown by the red line. The difference between data and simulation for each dataset, $D(Q) - F(Q)$ is shown by the grey line. Each dataset has been shifted by 1 for clarity. (b) A comparison of the sum of residuals squared, $\sum[D_i(Q) - F_i(Q)]^2$, for initial and repeat simulations, with the former offset by 0.05 for clarity.

Figure 3.11 presents a comparison of selected RDFs for the initial and repeat simulations for L-glutamine. What is clear from both sets of functions is that key hydrogen bonding interactions are still present in both simulations. Any changes present arise from the modifications to the data and model described in Section 3.3. The variations in the peak height of the coordination shells could be linked to the introduction of dihedral angles, where keeping to a strict conformation varies how the molecule may form intermolecular hydrogen bonds with the surrounding hydration shell. For functions such as Os-Ow, shown in Figure 3.11(b), there is more variation between the two simulations in the outer coordination shells, which is possibly due to changing the intramolecular geometry of water.

Table 3.4 compares the first coordination shell numbers and peak positions for the initial and repeat simulations. The peak positions of the first shell in each remain fairly consistent, the largest difference between peaks occurs for functions such as Hb-Ow, where the hydrogen bonding distance has increased from 1.66 Å to 1.71 Å. For interactions involving water oxygen atoms, there is very little change in coordination number, with all values being within error. The results still show that a carboxylate oxygen will coordinate approximately 3 Ow atoms, Os can coordinate between 2 and 3 Ow atoms and the hydrogen atoms will form a a hydrogen bond with 1 water molecule.

For interactions involving water hydrogens, the coordination shell numbers do not match within error. For interactions involving a direct hydrogen bond, such as the amide lone pair with Hw, the coordination number is identical. For nitrogen-water hydrogen functions where the atoms are indirectly correlated, with the hydrogen bond being made via a glutamine hydrogen - water oxygen interaction, the coordination numbers are within error of each other and the repeat simulation suggests that more water hydrogens surround nitrogen atoms. This is attributable in part to the changes to the model. Furthermore, utilising the iterative version of Gudrun may be revealing features in the data that were not previously evident in the manually-corrected data. In computational simulations of simple systems such as water, there is much fluctuation in the produced radial distribution functions (223). Work by Busch and co-workers on peptides and L-proline (139; 140; 146), which include both EPSR and MD simulations, suggest that the functions produced vary according to the methods and parameters utilised.

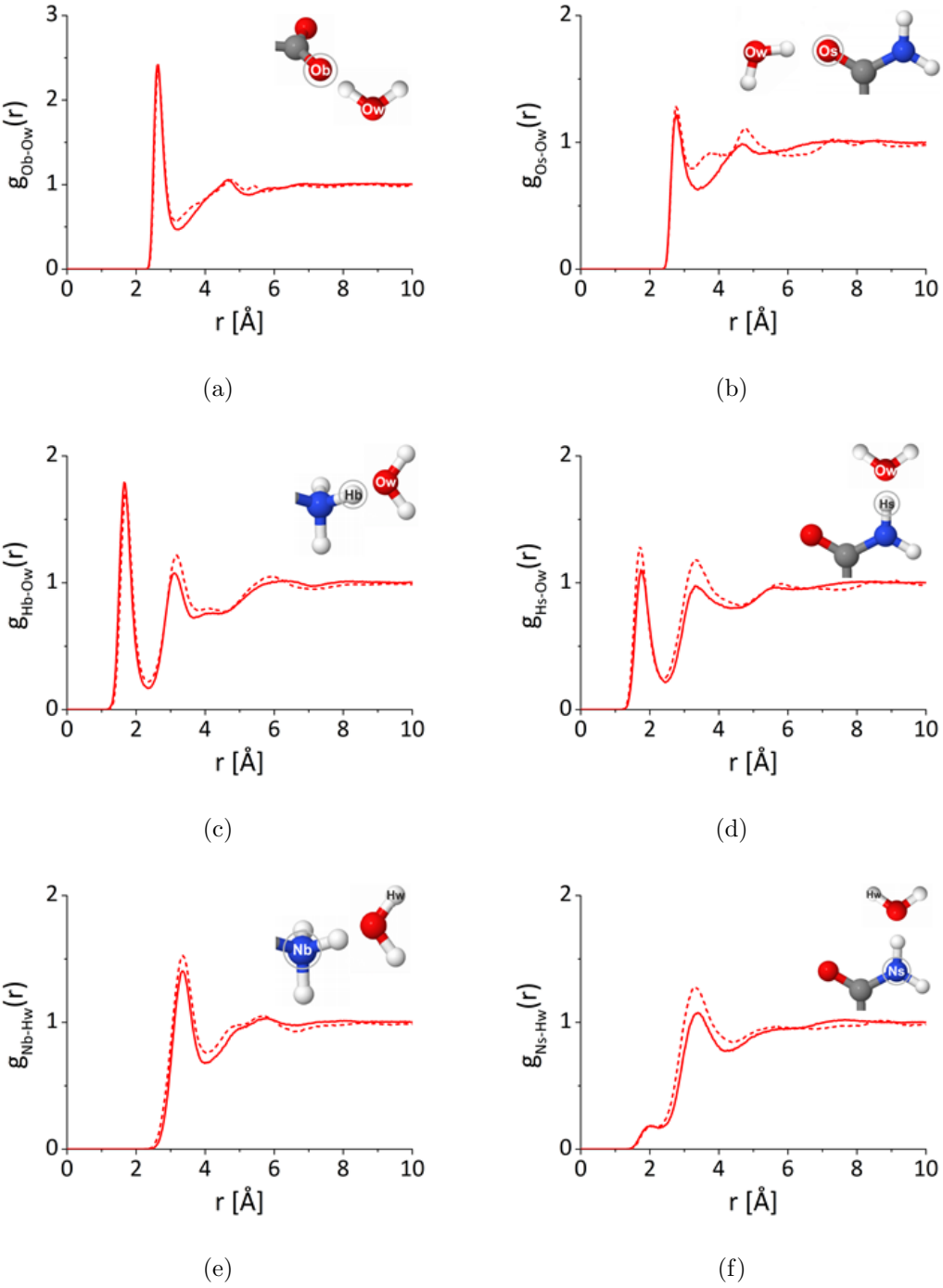


Figure 3.11: The pairwise RDFs for glutamine-water interactions for initial (solid) and repeat (dashed) studies of L-glutamine. The functions shown are for the (a) backbone oxygen - water oxygen (Ob-Ow), (b) side chain oxygen - water oxygen (Os-Ow), (c) backbone hydrogen - water oxygen (Hb-Ow), (d) side chain hydrogen - water oxygen (Hs-Ow), (e) backbone nitrogen - water hydrogen (Nb-Hw) and (f) side chain nitrogen - water hydrogen (Ns-Hw) interactions.

Table 3.4: Coordination numbers for interactions of hydrophilic-based oxygen, nitrogen and hydrogen atoms with water, comparing the results of the initial (1) and repeat simulations (2).

Atom Label	Simulation	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Ob-Ow	1	1	3.21	2.63	2.9	0.7
	2	1	3.15	2.64	2.8	0.7
Os-Ow	1	1	3.36	2.76	2.5	0.9
	2	1	3.21	2.76	2.3	0.8
Hb-Ow	1	1	2.35	1.66	1.0	0.3
	2	1	2.34	1.71	1.0	0.3
Hs-Ow	1	1	2.45	1.74	0.9	0.4
	2	1	2.40	1.71	0.9	0.4
Nb-Hw	1	1	3.99	3.35	10.7	1.8
	2	1	4.02	3.36	12.6	1.7
Ns-Hw	1	1	2.23	2.03	0.3	0.5
	2	1	2.22	2.08	0.3	0.4
NsHw	1	1	4.20	3.36	13.4	2.2
	2	1	4.35	3.33	17.8	2.2

The results suggest that the conclusions of initial studies on L-glutamine have been maintained in the repeat simulation. Where there is variance, due to improvements made to the data and model, these changes have provided new insight into the extended hydration of L-glutamine.

In this section, the association of L-glutamine has not been covered. Due to the low concentration of the solution, it was concluded previously in Section 3.2 that sampling with only 25 L-glutamine molecules was likely insufficient to produce conclusive quantitative information on the clustering of L-glutamine in solution. Hydrophilic interactions exist and the molecules can form instantaneous dimers. To probe the association of L-glutamine, a simulation has been completed using a newer version of EPSR on a larger simulation box. This is presented in the next section.

3.5 The Persistence of Hydrophilic Clusters at Higher Temperatures

To obtain more insight into the clustering behaviour of L-glutamine a simulation has been carried out using a larger simulation box (3). This simulation was completed on a newer version of EPSR, EPSR24, which can complete simulations on larger simulation boxes due improvements to the computational running of the software, thus allowing for greater sampling of the clustering in the system. EPSR24 will be discussed in more detail in Chapter 6. To run the basic simulation and one routine, CLUSTERS, takes many weeks. Therefore, it has not been possible in the time-frame allowed to fully analyse details regarding the hydration. This was, instead, only attained using a smaller simulation in EPSR18, as described in Section 3.4. Consequently, the simulations presented in this section are the only ones in this thesis run using EPSR24.

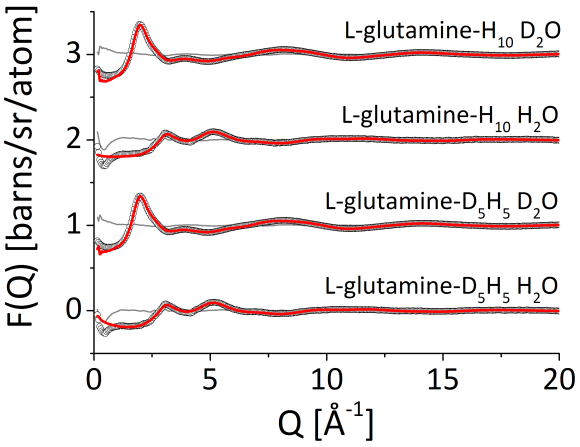
To extend this study, the clustering of L-glutamine has also been monitored at two higher temperatures, 37 °C and 60 °C. It is of interest to study biomolecules at higher temperatures than ambient conditions so as to understand the interactions of these molecules under more life-like conditions. 37 °C is basal body temperature and thus is a temperature of relevance to many living systems and the mechanisms behind disease. At 60 °C, molecules adapted to cope with extreme temperature exist, such as those found in extremophilic organisms (226), and a point where certain proteins may begin undergoing thermal denaturation (227). Glutamine is of importance to this latter process as it is the deamidation of the glutamine or asparagine side chain that constitutes the first stage of the thermal denaturation process (227).

In addition, temperature studies allow for the examination of the temperature dependence of important non-covalent interactions. The hydrophobic effect has been long thought to dominate the folding of proteins (22). However, recent research is now elucidating the role of hydrophilic groups in association. McLain *et al* (131) completed neutron diffraction studies on aqueous dipeptides of varying hydrophobicity. This study showed that charged-based interactions dominated association, with the least hydrophobic dipeptide forming the largest clusters. These dipeptides though did not contain polar side chains, with association only being driven by interactions of the backbone.

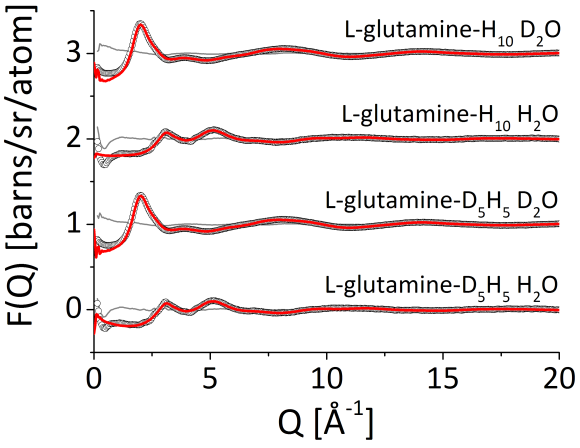
However, experiments on the amphiphilic molecule tertiary-butanol (106), which contains 3 non-polar methyl groups and a polar hydroxyl group, suggested solute-solute interactions via polar groups are more dominant at higher temperatures, whilst non-polar interactions prevail at room temperature. The tertiary-butanol work suggested that there is a balance between non-polar and polar interactions, as well as thermodynamics, that govern association (106). Comparatively, L-glutamine is less hydrophobic than tertiary-butanol, with only two methylene and the α-carbon methanetriyl groups forming a hydrophobic region. Therefore, it is interesting to complete studies on more hydrophilic molecules with varying temperature so as to understand the role of polar interactions in molecular association.

Simulations containing 200 L-glutamine molecules and 53800 waters were constructed for all three temperatures, equating to the same ratio of 1 L-glutamine to 269 water molecules as that used in previous simulations. These simulations were run for $\sim$ 16000 iterations. The fits to the data are presented in Figure 3.12, as well as the respective residual squared measurements. All three simulations show good fits to the data. The R-factors for the simulations are 1.6×10^{-4} for 24 °C, 1.8×10^{-4} for 37 °C and 1.5×10^{-4} for 60 °C. Compared to the simulation presented in Section 3.2, the R-factor values have decreased by a factor of 2×10^{-4}. Compared to the repeat simulation in Section 3.4 (R-factor 3.8×10^{-4}), the improvement to the R-factor is attributable to the increased size of the box and the use of a newer version of the software. The sum of residuals squared plots for each temperature are shown in Figure 3.13.

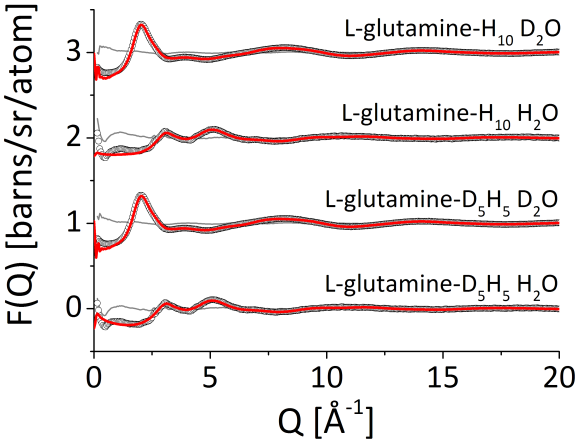
In initial studies on L-glutamine (Section 3.2), the first coordination shell of the O-H bond between two L-glutamine molecules was interrogated to determine whether an ammonium or amide hydrogen were coming into short-range distance of an amide or carboxylate oxygen atom. The infrequency of the interactions, inadequate sampling and noise make determining a precise value difficult for the minimum of the first coordination shell. For this study, an average distance was used, based around an average of the previous O-H distances. Here, the average was 2.37 Å and thus the distance was taken to be 2.4 Å. This value is comparable to definitions given in the literature for intermolecular interactions between organic molecules (228).



(a) 24 °C



(b) 37 °C



(c) 60 °C

Figure 3.12: The fit of the total simulated structure factor, $F(Q)$, to the experimental data, $D(Q)$, for EPSR24 studies of L-glutamine at (a) 24 °C, (b) 37 °C and (c) 60 °C. The data is shown by the black circles, whilst the fit is shown by the red line. The difference between data and simulation for each dataset, $D(Q) - F(Q)$ is shown by the grey line. Each dataset has been shifted by 1 for clarity.

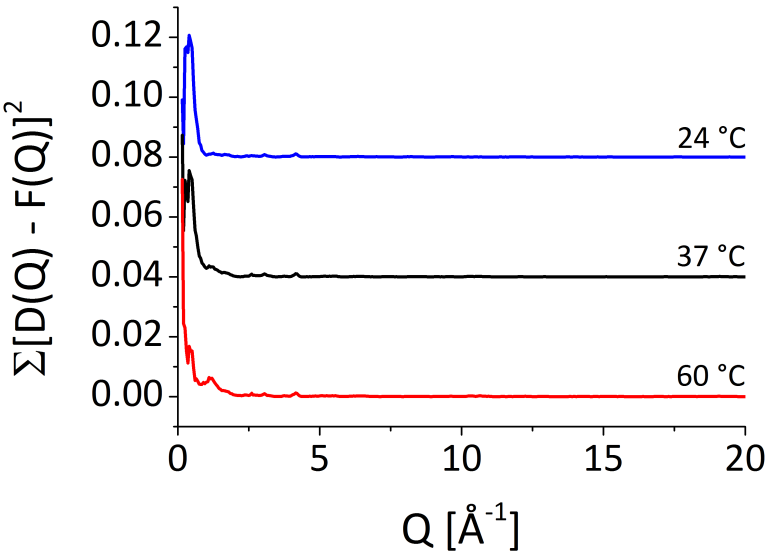


Figure 3.13: A comparison of the sum of residuals squared, $\sum[D_i(Q) - F_i(Q)]^2$, EPSR24 studies of L-glutamine at 24 °C (blue), 37 °C (black) and 60 °C (red). Each dataset has been shifted by 0.04 for clarity.

The RDFs for the four hydrophilic association interactions studied are shown in Figure 3.14. Some noise is still present in the RDFs due to sampling, though the consistency of each is more stable, given the larger simulation size, larger number of iterations and computational constraints present. Whilst previously the 24 °C results were shown in red, this result is now shown in blue, the 37 °C functions in black and 60 °C functions in red.

Figure 3.15 shows pie charts for the percentage of L-glutamine molecules found as monomers and in a cluster of 2 or more at any given time with increasing temperature. The percentages given are for both clusters produced via all possible hydrophilic interactions (All) and through each separate interaction. Cluster statistics have been collected for all interactions separately. The percentages suggest that hydrophilic interactions of all four kinds are present in the solution. The most predominant of these is Ob-Hb. This is attributable to backbone atoms being on functional groups with partial charges; a form of salt bridge is being produced to strengthen the interaction.

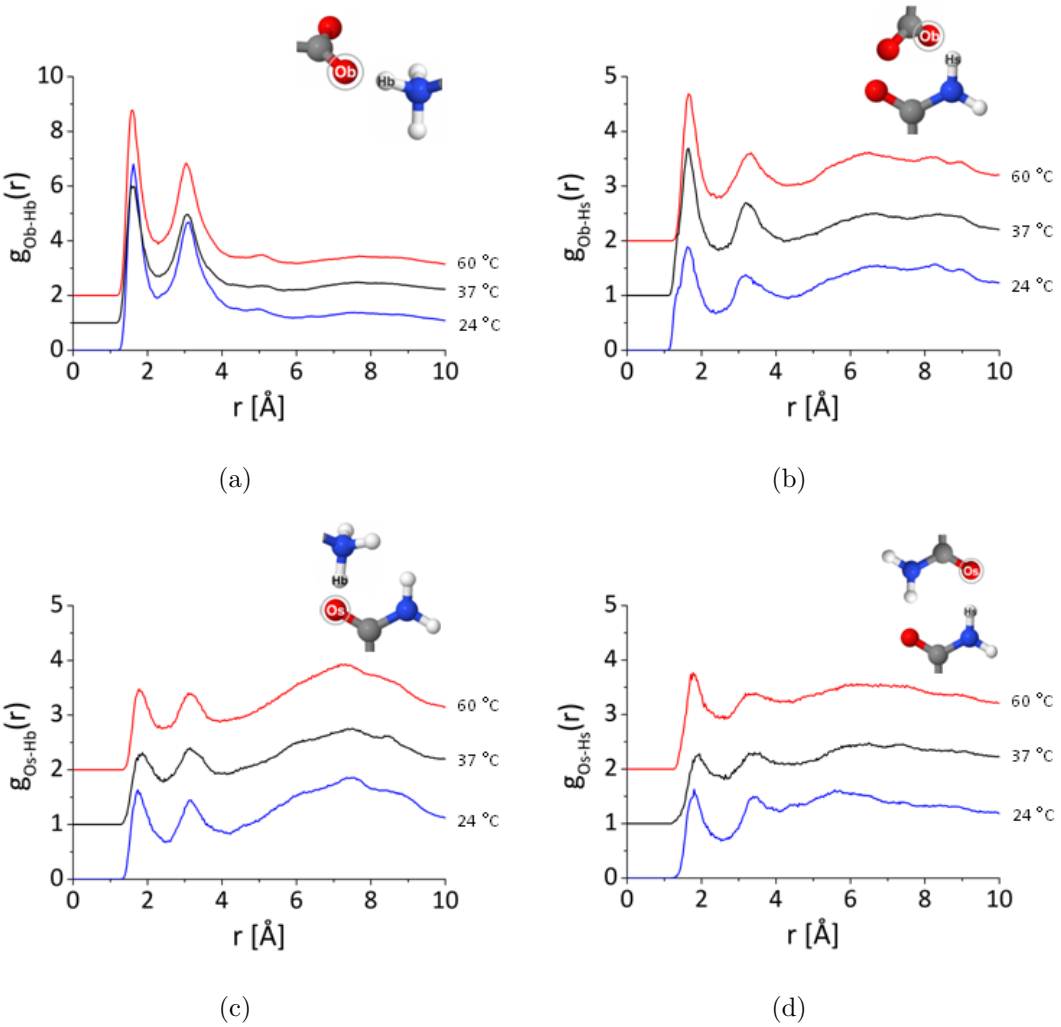
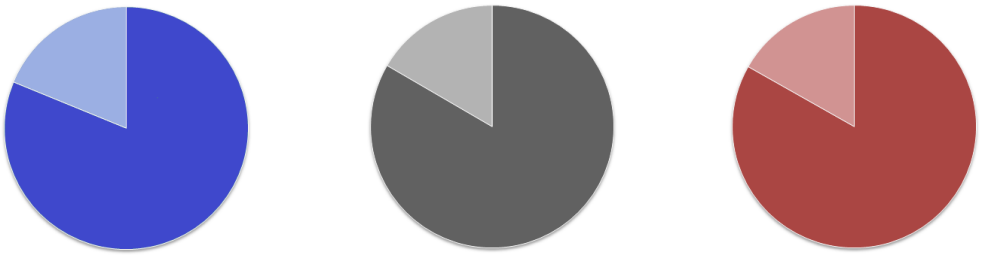


Figure 3.14: The pairwise RDFs for hydrophilic glutamine-glutamine interactions for EPSR24 studies of L-glutamine at 24 °C (blue), 37 °C (black) and 60 °C (red). The four interactions are (a) backbone - backbone (Ob-Hb), (b) backbone - side chain (Ob-Hs), (c) side chain - backbone (Os-Hb) and (d) side chain - side chain (Os-Hs). The RDFs have been shifted by 1 each for clarity.



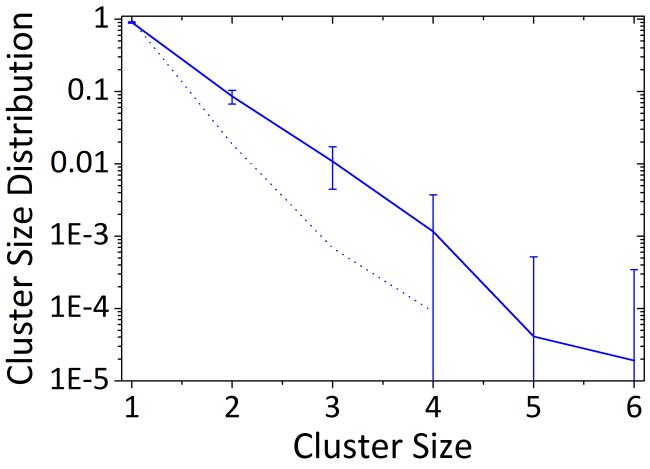
(a) 24 °C (b) 37 °C (c) 60 °C

Figure 3.15: Percentage of L-glutamine molecules existing as monomers (dark colour) and as non-monomers in hydrophilic intermolecular clusters of 2 or more (light colour), at (a) 24 °C (blue), (b) 37 °C (grey) and (c) 60 °C (red).

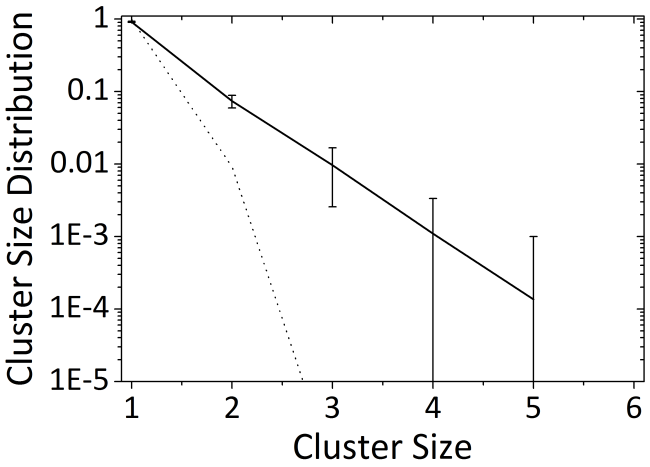
From Figure 3.15, 15–20% of the molecules are found to be involved in a cluster formed by hydrophilic interactions at any given time. The degree of fluctuation in the clustering suggests there is no appreciable trend with increasing temperature. Studying each interaction separately, there is consistency in the clustering observed for the three temperatures. However, it is clear that monomers are the dominant species in the solution. The clusters present are not stable globules, instead instantaneously forming and breaking apart. Had they been stable and long-lasting, there would have been evidence in the low-Q region of the $F(Q)$ data, particularly for D_2O, which is less influenced by incoherent and multiple scattering effects. Another means to support this is using SANS data that can detect structures in the range of tens to thousands of Ångstroms. Complimentary SANS data is presented in Section 3.6 of this thesis.

The CLUSTERS routine (Section 2.2.3.2) also allows the size of the clusters to be monitored in the EPSR simulation. At 24 °C, it is possible to observe up to 6-mers in the solution, as shown in Figure 3.16(a). With increased temperature, no appreciable trend is present, with clusters of up to size 5 seen for 37 °C (Figure 3.16(b)) and size 6 for 60 °C (Figure (c)). Due to low statistics, these maximum values are likely not stable and as observed in 3.16, the majority of clusters are in the form of dimers. As hydrogen bonding is still prevalent in this system, this defies the notion of hydrogen bonds becoming less stable with increasing temperature (9; 229). However, an increase of hydrophilic interactions at higher temperatures has been suggested in tertiary-butanol solutions (106), with the conclusion of this work indicating that there is a complex balance between enthalpic and entropic contributions governing association. Such a trend has also been observed in Raman spectroscopy and electron microscopy studies of polyglutamine chains, where increased aggregation was seen at 60 °C compared to room temperature (230). Whilst the thermodynamics is not presented in this thesis, there are many factors that could be contributing to the effect observed, including the ordering of water around hydrophilic groups and enthalpy changes from breaking bonds and forming new ones. This investigation covers only a small temperature range, making this a first step to understanding the complex regime that covers the hydrogen-bonding patterns for this system.

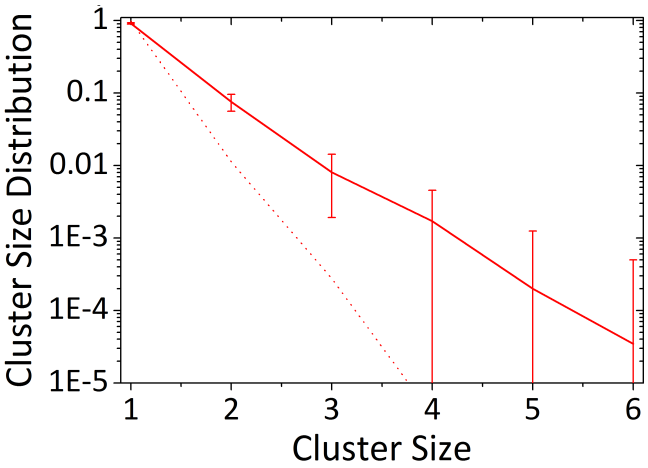
To check whether the interactions observed are the result of hydrogen bonding and not the random packing of biomolecules, a simulation was completed for each temperature with the hydrogen bonding switched off. As hydrogen bonding is dependent on the atomic electronegativity and partial charges, setting the partial charges to zero turns off the hydrogen bonding. For these simulations, the clustering has decreased to a mere 3-4% of molecules being involved in clusters, indicating that hydrogen bonding drives the formation of clusters. The percentage of molecules involved in a cluster of size 2 or more for each temperature is presented in Figure 3.17. The clusters present in these random packing simulations are never larger than dimers.



(a) 24 °C

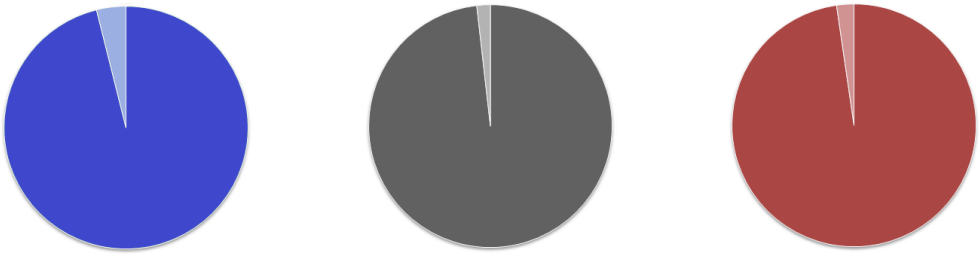


(b) 37 °C



(c) 60 °C

Figure 3.16: Probability of hydrophilic L-glutamine clusters forming versus the size of the clusters observed, at (a) 24 °C, (b) 37 °C and (c) 60 °C. The probability scale is logarithmic.



(a) 24 °C (b) 37 °C (c) 60 °C

Figure 3.17: Percentage of L-glutamine molecules existing as monomers (dark colour) and as non-monomers in hydrophilic intermolecular clusters of 2 or more (light colour) when hydrogen bonding is switched off, at (a) 24 °C (blue), (b) 37 °C (grey) and (c) 60 °C (red).

Another interaction to consider is the hydrophobic effect. Though the surface area is small, the presence of three alkyl groups in a chain means some hydrophobic surface is presence that could permit such interactions to take place. In studies of N-acetyl-L-glutamine-methylamide by Plumley and Danneberg (231), the molecules are thought to align so as hydrophobic groups are in short-range distance of each other. This arrangement lowers the free energy of the system.

In simulation box snapshots, it is difficult to find evidence of any hydrophobic interactions. An RDF shows that Cm groups can come into short-range distance, under 5 Å, of each other (Figure 3.18), though no prominent peak is present to indicate a notable interaction taking place. Any hydrophobic interactions are negligible compared to their hydrophilic counterparts and any that do occur happen *in situ* with a hydrogen bonding interaction. Hydrophilic dominance is also seen in studies of dipeptides (131) and L-proline (146). It is likely that result seen by Plumley and Dannenberg is not present here due to the molecule being in zwitterionic form, with hydrophilic interactions dominating over hydrophobic ones. This provides evidence that hydrophilic interactions drive association between polar amino acids. It is possible that hydrophilic interactions initiate association and at certain concentrations hydrophobic interactions dominate to drive the salting out of the amino acid from the solution.

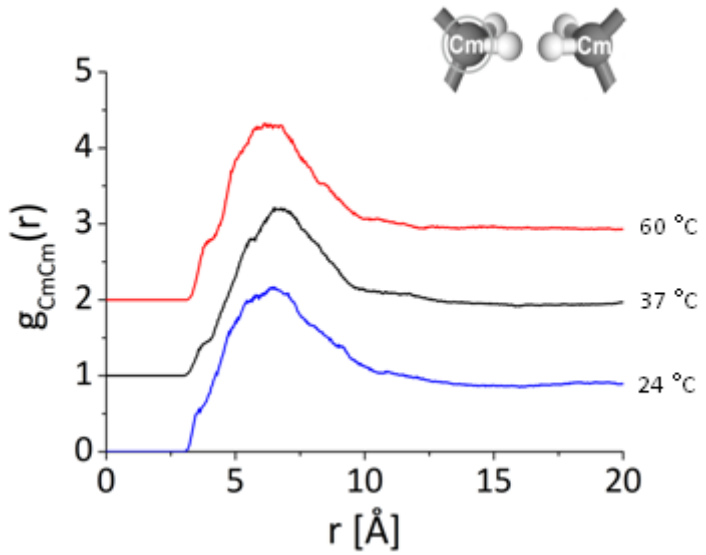


Figure 3.18: The methylene carbon - methylene carbon (Cm-Cm) RDF for L-glutamine at 24 °C (blue), 37 °C (black) and 60 °C (red). The RDFs have been shifted by 1 each for clarity.

What should be appreciated in this study is that the results for L-glutamine association are the result of EPSR simulations that include characterised force fields from defined references, which have been refined against total cross-sectional neutron diffraction data. As a result of the low concentration of the solution, the weighting of the glutamine-glutamine interactions in the diffraction data is small. For instance, whilst the intermolecular correlation for the Ow-Hw interactions is 1.6×10^{-1} and 1.2×10^{-3} for the Ob-Hw, it is only a mere 6.7×10^{-6} for Ob-Hb, the most prevalent glutamine-glutamine interaction. In more dilute simulations using the EPSR method, it has been noted that being able to fit the simulation to the diffraction data based on solvent-solvent and solute-solvent interactions can aid in directing the finer solute-solute correlations appropriately (138). Here, it should be highlighted that the correlations observed are due to the simulation and are reliant on the potentials introduced into the model. Overall, there are limitations in completing neutron diffraction studies on dilute studies, with conformational information and low-weighting molecular correlations potentially not being present in the data. For such cases, information can only be obtained using an appropriate computational model.

3.6 Small Angle Neutron Scattering on Aqueous L-Glutamine

In the previous section, it was determined that L-glutamine in aqueous solution had the propensity to form aggregates of 6-8 molecules. The dominant species was determined to be monomers, with any clusters formed being transitory, forming and breaking almost instantaneously. It was speculated that any formation of large-scale, long-standing globules would have been evident in the low-Q region of the data, particularly for D_2O that is a low incoherent scatterer.

To further support this finding; SANS experiments have been completed on L-glutamine, at varying temperatures, to determine the presence of any larger clusters in the solution. Details of the SANS-specific theory and the instrument used are given in Section 2.3 of Chapter 2.

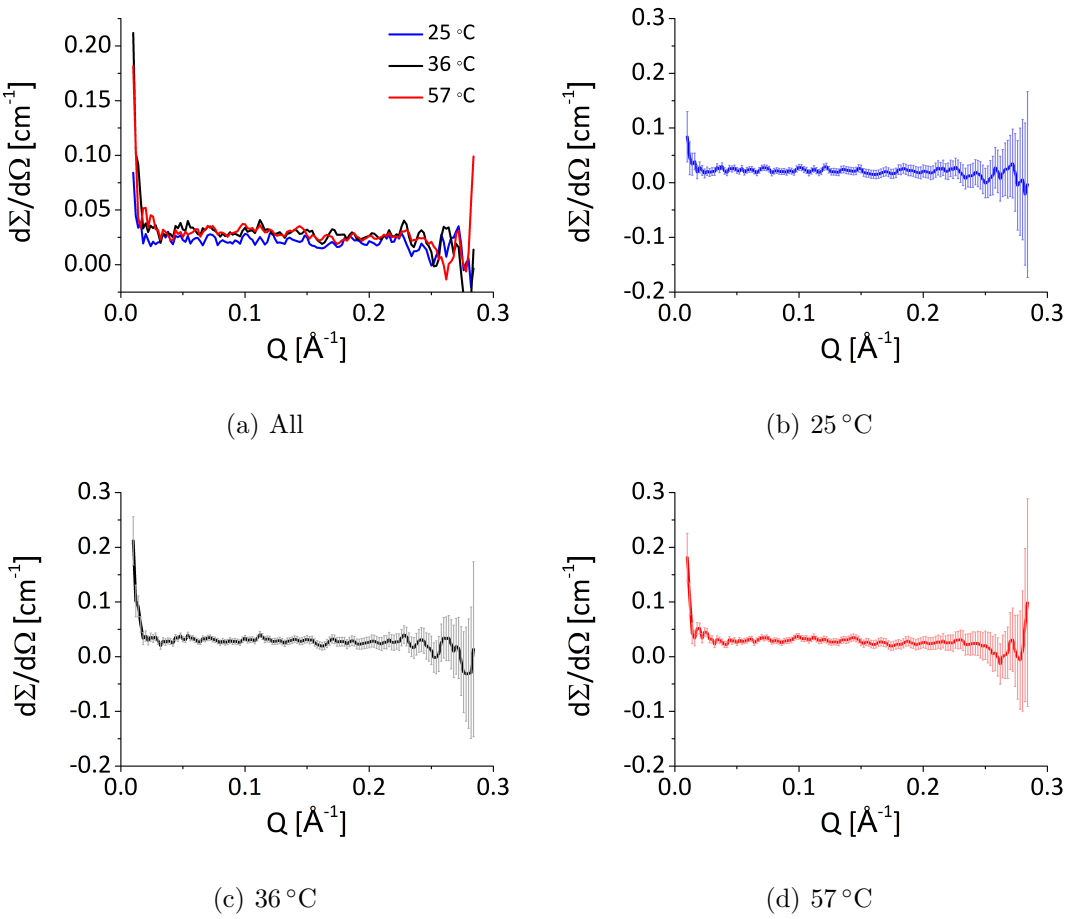
Experiments have been completed on a $30\,\mathrm{mg\,ml^{-1}}$ sample of L-glutamine in D_2O. D_2O was used instead of H_2O due to hydrogen being a strong incoherent scatterer leading to an unfavourable noise/signal ratio. Moreover, D_2O also increases the contrast between the solvent and the biomolecule's non-labile protons. The solution was studied at three temperatures similar to and in the range of those studied for the neutron diffraction experiments. These temperatures were 25 °C, 36 °C and 57 °C. For this experiment, it was not possible to equilibrate the temperature to the values used for the diffraction data in the Xpress experiment time allowed (Section 2.3.2). Both main and high-angle detectors have been used for this study, allowing for the detection of small stable clusters and larger aggregates. Data was collected for each temperature measurement over 1.5 hours. It should be noted that whilst data was collected at all three temperatures using the main detector, the high-angle detector was only used to attain data on L-glutamine at 25 °C. Whilst the concentration and experiment time are not ideal, the aim of this experiment to not to give an extensive analysis into glutamine association. Whilst the expected noise/signal ratio would be unfavourable for an extensive quantitative analysis, if long-lived clusters are present, a scattering signal would be observed nonetheless, therefore, providing the confirmation of their presence. Alternatively, an absence of signal would suggest there is no long-lived clustering in the solution.

The average backbone of an amino acid can be approximated to at least 4 Å in length, if the average bond length between atoms on the backbone is taken to be 1 Å. Referring to the minimum distances being probed by each detector, this would correlate to the high-angle detector being able to detect a minimum of dimers in the solution. For the main area detector, if the glutamine molecules were to form a chain in solution, then the main area detector could distinguish the presence of 6-mers in the solution, comparable to the size of the very largest clusters being detected in the neutron diffraction data.

Presented in Figure 3.19 is the data obtained at all three temperatures using the main area detector. Figure 3.19(a) shows the differential cross-section for all temperatures without errors and the data with their respective errors are presented separately in (b), (c) and (d). The data for all three temperatures is still subject to notable error bars, particularly at high Q values due to the data acquisition time, as well as the low concentration of the solution.

Comparing the results at each temperature, the cross-section of the sample initially appears to be correlated to the temperature, with results at higher temperature resulting in higher cross-sectional values. However, this is thought to be attributable though to a small increase in the incoherent background from the thermalisation of neutrons, as well as the effect of increasing the temperature on the density of D_2O.

The plots in Figure 3.19 suggest that no long-lived clusters are present in the solution, with the data taking the form of a flat background. The oscillations that are present can be attributed to noise instead of biomolecular association. If there were any stable clusters in the solution, there would be signal at high Q values, the form of this signal being dependent on the size, type and homogeneity of the clusters present. The cross-sectional values are also notably smaller than even those attained from water, which has a typical cross-section of $1\,\mathrm{cm}^{-1}$.



(a) All

(b) 25 °C

(c) 36 °C

(d) 57 °C

Figure 3.19: SANS data for L-glutamine in D_2O at 25 °C, 36 °C and 57 °C, using the main area detector. Presented here is the raw data for (a) all three temperatures with no errors, as well as the individual datasets for (b) 25 °C, (c) 36 °C and (d) 57 °C, with error bars.

The data obtained using the high-angle detector at 25 °C is presented in Figure 3.20. This plot also takes the form of a flat background, indicating a lack of stable, long-lived clusters in the solution, even down to simple dimers or trimers. Whilst the neutron diffraction data suggested that 15–20% of L-glutamine molecules would be in non-monomer state at a given time, their speculated transitory nature could provide an explanation as to why their presence is not detected in the SANS data.

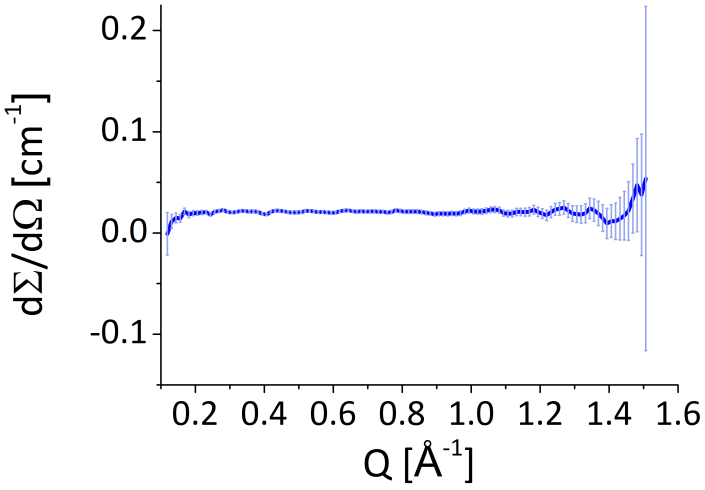


Figure 3.20: SANS raw data for L-glutamine at 25 °C, using the high-Q bank detector.

The only point where a rise in the absolute cross-section is seen at $Q < 0.02\,\text{Å}^{-1}$. A similar rise was seen in concentration studies of L-proline by McLain and co-workers (128). This work aimed to determine whether any aggregation was present in aqueous L-proline solutions to determine if this could be a potential mechanism for osmoprotection. This rise at very low Q values was explained to be the results of the instrument cut-off, instead of meaningful clustering data. Comparing to this work, it can be assumed that the low-Q rise for L-glutamine is due to the same reason. Had this been evidence of significant clustering in either case, then clusters would be present with dimensions of $\geq 314\,\mathring{A}$. The L-proline work also shares the same low cross-sectional values and lack of oscillations, with this study suggesting that this signified a lack of aggregation in the system. Overall, the SANS study completed here supports the results discussed in section 3.5, whereby L-glutamine is not thought to form stable clusters or larger globules in solution.

3.7 Conclusions

This chapter aimed to cover the foundations of completing studies on dilute solutions with neutron diffraction and EPSR. It focused on the study of L-glutamine and de-

scribed the general conclusions of the hydration and association of this molecule, as well as the limitations with the methodology. It has also highlighted the importance of hydrophilic interactions in the association of amino acids, over the contribution from the hydrophobic effect. Section 4.1 described the initial neutron diffraction experiment completed on L-glutamine. Section 3.2 summarised the key results attained from initial simulations of aqueous L-glutamine, which showed that the molecule readily hydrogen bonds with its aqueous environment and associates through its hydrophilic functional groups. Section 3.3 presented updates that have been introduced to the software and model, so as to improve the fit and account for conformational information not being present in the data, with repeat simulations in Section 3.4 showing that the changes had not significantly impacted the conclusions previously observed. The results on L-glutamine association from a larger simulation were presented in Section 3.5, where hydrophilic association dominates over those through hydrophobic contacts at all temperatures studied. Supporting SANS data discussed in Section 3.6 suggested that no stable globules were formed in the L-glutamine solutions. The temperature simulations trialled a new version of the EPSR software, though it has not been possible to analyse all systems for this thesis using this software due to project time constraints. This chapter has served as a foundation to determine the way to define biomolecular constraints for dilute systems and concentration limitations, which have been exploited in subsequent chapters.

Chapter 4

Determining the Properties of the Glutamine Imino Acid

4.1 Introduction

In Chapter 1, the concept, structure and properties of a peptoid were introduced. By moving the side chain onto the backbone nitrogen, the properties of these biomolecules compared their natural peptide counterparts are altered. This structural change results in the properties of these molecules, such as their solubility and resistance to degradation, being more tunable and enhanced relative to their peptide equivalents. Some of these property differences are not only seen for peptoid chains but also for imino acid building blocks. One of the key features noted in imino acids is their increased solubility and propensity to interact with water, their hydrophilicity. Chromatography studies have suggested that imino acids may be more hydrophilic than even charged amino acids. Such studies showed it was possible to produce peptide-peptoid hybrids that are both more cell-permeable and more hydrophilic, by replacing select amino acids with imino acid versions (47; 232). Additionally, the increased hydrophilicity is thought to be key for imino acids and short peptoid chains to act as osmoprotectants (62; 67). Consequently, it is vital to understand the hydration properties of these monomer units if they are to be exploited effectively for a variety of applications.

This chapter will focus on the glutamine imino acid (Ngln). This synthetic building block is more soluble, to at least ∼10x that of L-glutamine, making it possible to complete more detailed analyses of the behaviour of this novel molecule. As described in Chapter 1, the use of the Ngln residue has been limited in peptoid-based research. Its side chain has been utilised in small neuropeptides (171). However, it has been shown not to be useful in forming complex biological structures such as super-helices (44), despite the tendency of peptoids to form helical structures as described earlier in Section 1.1.4.3. Research such as this has led to the hypothesis that hydrogen bonding is not important in the assembly of more complex peptoid structures (34; 36). Analysing the important structural interactions for this building block will thus help to determine the process of self-assembly of these molecules, as well as potentially highlighting the key interactions relevant in glutamine-based diseases, which share a common side chain.

This chapter will present structural properties of the Ngln building block. It will compare the hydration of Ngln to that of its amino acid equivalent, L-glutamine, to determine the source of its increased hydrophilicity. Glutamine has been selected due to its interesting hydrogen bonding properties, role in disease and notable difference in solubility between amino and imino acid versions. These results will be presented in Section 4.4. In Section 4.5, results will be presented for the impact of concentration on these hydration interactions, as well as details on the intermolecular association of imino acid molecules that could not be probed for L-glutamine. Finally, Section 4.6 will determine the impact of temperature on the association of Ngln. Initially, the chapter will first present background literature relevant to the present study in Section 4.2 and Section 4.3 will describe the isotopic substitution experiments and simulations completed for this comparison. Analysing the important interactions for this building block will thus help determine the source of the properties important for self-assembly, as well as potentially highlight the key interactions involved in glutamine-based diseases.

4.2 Background Structural Studies on Imino Acids versus Amino Acids

This chapter will include a comparison of the hydration of model amino acid and imino acid (Section 4.4). By observing how each interacts with its aqueous environment and associates in solution, it is possible to being to relate their structural properties to their function.

To date, only one structural study has been completed that compares amino and imino acid. This research was completed on L-alanine and its monopeptoid-equivalent sarcosine and used Near-Edge X-ray Absorption Fine Structure (NEXAFS) spectroscopy to study these molecules in a hydrated and dehydrated state (82), by Uejio and co-workers. Absorption spectroscopy and neutron diffraction are complementary techniques, and it is possible to use EPSR to analyse both types of data simulataneously (122; 124; 125). As detailed in Section 1.2.2.1, NEXAFS spectroscopy is sensitive to the electronic structure of a molecule, where it is possible to produce Lowest Unoccupied Molecular Orbitals (LUMO) for a given system. LUMO plots provide information on the distribution of orbitals over a molecule capable of being involved in an interaction (6). NEXAFS spectroscopy can also provide information on the conformation and flexibility of a molecule.

In work by Uejio *et al* (82), NEXAFS data was compared to simulations of the molecules in a hydrated and dehydrated state. For work on the hydrated form of the sarcosine, the simulation that produced a spectrum most like the experimental data was achieved when no dihedral angle constraints were in place. For L-alanine, the flexibility of the molecule did not affect the simulated spectra produced, likely due to the size and chirality of the molecule.

However, this paper also highlighted the importance of using constraints in simulations to prevent unphysical structures from being produced. Simulations were completed to observe the spectra produced if the dihedral angles fixed within the simulation were varied. Whilst trialling different dihedral angles in L-alanine had little impact on the simulated spectra, notable changes were observed in the simulated spectra when dihedral

angles for sarcosine were varied. It was determined that different fixed conformations produced different simulated NEXAFS spectra, with only certain structures producing good fits to the data. This indicates that sarcosine is a molecule sensitive to conformation; though it is more flexible compared to L-alanine, it will only take on certain conformations.

LUMO plots for L-alanine and sarcosine were also compared in both the hydrated and dehydrated states. The available orbitals varied notably for L-alanine between the two states as opposed to sarcosine that showed orbitals in a similar location. Whilst sarcosine is more influenced by its conformation, L-alanine is affected by its hydration state. For sarcosine, orbitals were not only present over the ammonium and carboxylate groups but also the backbone methylene and the side chain methyl groups. For L-alanine, orbitals were only present over the backbone zwitterionic groups. This suggests that water molecules interact with more parts of sarcosine, compared to L-alanine, which could account for sarcosine's increased solubility (Section 1.3.3). However, the results of this study were qualitative and no information on the number of molecules coordinated was ascertained. This chapter will provide quantitative information on how the number of molecules coordinated varies between an amino and imino acid.

4.3 Experimental Procedure

Solutions of Ngln have been studied using neutron diffraction and EPSR modelling. Ngln was synthesised using the methods of McKinney *et al* (233) and Stewart (234). The synthesis procedure is detailed in Appendix A.

Ngln has been studied at two concentrations; $30\,\mathrm{mg\,ml^{-1}}$, to be comparable with work on L-glutamine, and $300\,\mathrm{mg\,ml^{-1}}$, to allow the conformation and association of the molecule to be studied. This equates to a molecular ratio of 1 imino acid to 270 water molecules for the dilute sample and 1 imino acid to 27 water molecules for the concentrated sample. The latter ratio is comparable to those used to study other amino acids using neutron diffraction and EPSR (127; 128; 130; 131). The protiated molecule was studied with neutron diffraction in deuterium oxide, milli-Q water and 'null water'

(Table 4.1). Null water is 64 % H_2O and 36 % D_2O; proportions of the scattering length are such that the water hydrogen scattering is virtually eliminated, leaving only signal from heavier atoms in the system. The three isotopic samples for each concentration were studied under ambient conditions of 25 °C and 1 bar. For the 300 mg ml^{-1} sample, data has also been obtained at 37 °C and 60 °C.

Table 4.1: Isotopic substitution experiments completed for Ngln.

Sample No.	Sample Name	Description
i	Ngln-D_2O	Fully protiated Ngln in D_2O
ii	Ngln-H_2O	Fully protiated Ngln in Milli-Q water (H_2O)
iii	Ngln-Null Water	Full protiated Ngln in a mixture of 36% D_2O and 64% Milli-Q water (H_2O)

The simulation parameters for the imino acid; bond lengths and angles, as well as dihedral angles, were defined by a a semi-empirical AM1 model (87) produced using MOPAC-7 software (225), as was used previously for L-glutamine. OPLS potentials were used to define the Lennard-Jones and Coulombic potentials for both biomolecules (91). Lennard-Jones and Coulombic potentials for water were defined according to SPC/E rules (93), with the geometry the same as the updated values specified in Section 3.3.3. These parameters are tabulated in Appendix B.

The labels used for L-glutamine and water are the same as those shown in Figure 4.2. There are 12 distinct atomic components defined in EPSR simulations for the imino acid. These components are labelled in Figure 4.1. The hydrogen atoms are labelled as Hb, Hm and Hs, referring to hydrogens on the backbone ammonium, hydrocarbon groups and the side chain amide respectively. The oxygen atoms are Ob, for the two found on the backbone carboxylate anion and Os for the one on the amide group. Nb refers to the ammonium nitrogen and Ns the amide nitrogen. For the carbon atoms, Cb is that found on the carboxylate and Ca the α-carbon. Whilst the methylene carbons are both

labelled Cm for L-glutamine, the two for the imino acid have been classified separately in EPSR. The Cm label only counts for the methylene nearest the amide group in the imino acid. The side chain methylene attached to Nb is Cn, labelled distinctly due to its connection to the backbone nitrogen and differing LJ potentials (91; 92; 235).

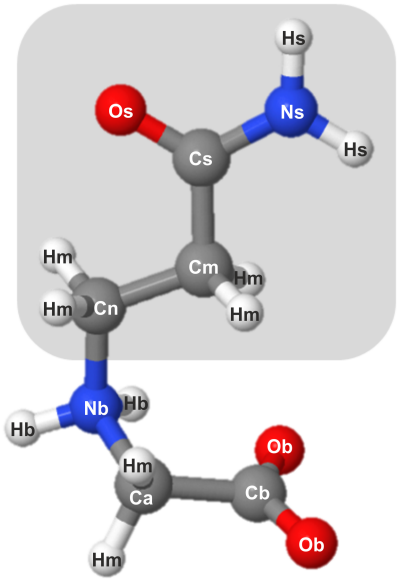


Figure 4.1: Diagram of the glutamine imino acid (Ngln), with the location of the backbone and side chain groups detailed. The side chain of the molecule is highlighted in grey. C=O double bonds are not shown in this diagram. The labels are those used to identify each of the atomic components in the EPSR simulations. There are 12 distinct atomic components for Ngln.

4.4 Comparing the Hydration of Amino Acid and Imino Acid

An EPSR simulation box containing 25 Ngln molecules and 6750 waters was set up for the 30 mg ml^{-1} sample. The fit to the neutron diffraction data is shown in Figure 4.2. This simulation has an R-factor of 1.9×10^{-3}, suggesting a good fit to the data has been achieved.

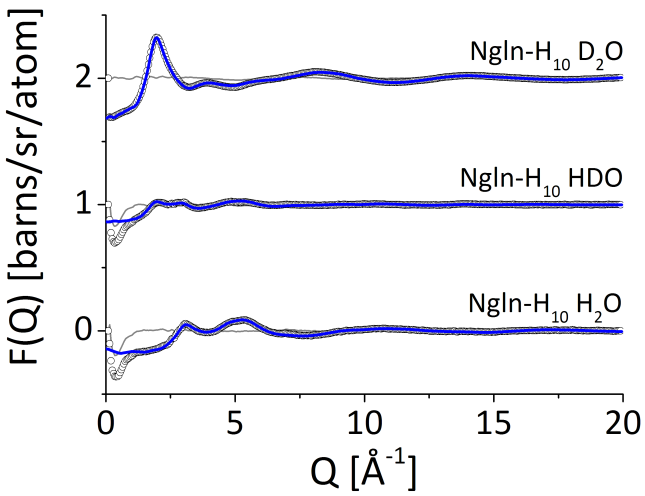


Figure 4.2: The fit of the total simulated structure factor, $F(Q)$, to the experimental data for Ngln at 30 mg ml^{-1}. The data is shown by the black circles, whilst the fit is shown by the blue line. The difference between data and simulation for each dataset, $D(Q) - F(Q)$ is shown by the grey line.

With the isomer difference being on the backbone, Sections 4.4.1 and 4.4.2 will present the effect of the structural change to the backbone nitrogen and adjacent methylene groups respectively. Section 4.4.4 covers how changes to the backbone impacts hydration of the side chain. For this, comparisons will be made to the repeat simulation for L-glutamine, introduced in Section 3.4.

4.4.1 Changes to the Backbone: Ammonium Group

The main structural difference between an amino and an imino acid is that the primary ammonium cation present (R - NH_3^+) has been converted into a secondary ammonium cation (R_2 - NH_2^+), by moving the side chain onto the nitrogen of this functional group, Nb. The presence of ammonium groups is derived from the molecule being zwitterionic; it possesses charged terminals, an ammonium cation and a carboxylate anion, but still maintains a net charge of 0. Converting to a secondary ammonium removes one hydrogen atom, reducing the number of water molecules, or specifically water oxygen atoms coordinated by this nitrogen by approaching 2. Table 4.2 shows the coordination numbers for the number of water oxygen atoms around Nb and its attached hydrocarbon groups and shows that whilst the coordination number for the first coordination shell of

Nb-Ow is 3.9 for L-glutamine, it is only 2.2 for Ngln. Studies completed on other amino acids, including L-glutamic acid (127) and L-proline (128), as well as neurotransmitters like dopamine (138) suggest that the presence of an extra hydrogen on the primary ammonium cation allows for the coordination of an additional water oxygen atom. Figure 4.3 shows the radial distribution functions for the Nb-Ow interaction and suggests that the water molecules hydrogen-bonded to this group are of a similar strength, with first coordination shell peaks for both molecules positioned at 2.73 Å.

Table 4.2: First coordination shell numbers of backbone - water interactions for L-glutamine (Gln) and Ngln.

Atom Label	Molecule	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Nb-Ow	Gln	1	3.46	2.73	3.9	0.8
	Ngln	1	3.21	2.73	2.2	0.5
Ca-Ow	Gln	1	4.02	3.57	5.4	1.1
	Ngln	1	4.20	3.39	7.6	1.3
Cn-Ow	Gln	N/A	N/A	N/A	N/A	N/A
	Ngln	1	4.23	3.42	7.4	1.2

Whilst the number of water molecules present has decreased for the first coordination shell, greater coordination can be found in outer shells for Ngln. Figure 4.3 suggests there is a more prominent coordination shell in the region of 3.21 Å to 5.76 Å, as well as evident subsequent shells after, shown by the succession of peaks found in the function. This structuring of water at longer length-scales is not seen to the same extent for L-glutamine. A coordination number analysis (Table 4.3) of the second shell shows 27.5 water molecules can be found in this second shell compared to 25.4 for L-glutamine. This increased density suggests that water molecules may still be situated more closely to Ngln. The position of the second shell has also shifted to shorter length-scales for Ngln, from 5.31 Å to 4.65 Å , suggesting water molecules are forming stronger hydrogen bonds to other regions of the imino acid. Therefore, whilst less water is found in the first coordination shell of the ammonium groups, there is more water in subsequent shells that may correlate to other parts of the biomolecule.

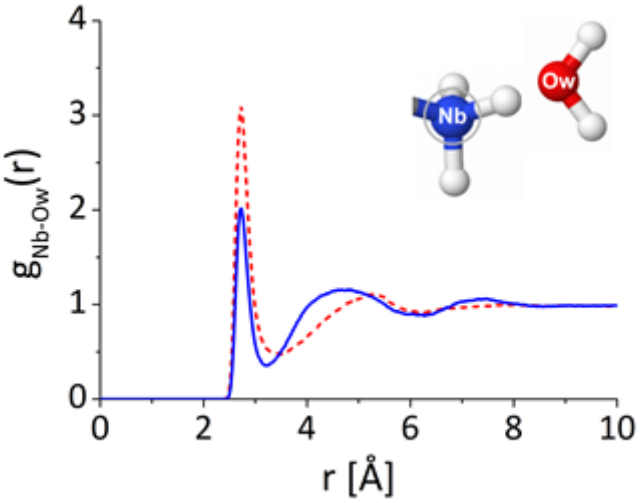


Figure 4.3: The pairwise RDF for the backbone nitrogen - water (Nb-Ow) interaction for L-glutamine (red, dashed) and Ngln (blue, solid).

Table 4.3: Second coordination shell numbers of ammonium - water interactions for L-glutamine (Gln) and Ngln.

Atom Label	Molecule	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Nb-Ow	Gln	3.46	6.23	5.31	25.4	1.8
	Ngln	3.21	6.23	4.65	27.5	1.8

4.4.2 Changes to the Backbone: Methylene Groups

Attached to the secondary ammonium cation are two methylene groups. One methylene ($CaHm_2$) is connected to the carboxylate anion on the backbone, whilst the other ($CnHm_2$) is adjoined to the rest of the glutamine side chain. Generally, hydrocarbon groups such as these methylene groups are considered hydrophobic; they repel water molecules. Nevertheless, even hydrocarbon groups possess small amounts of partial charges, even if smaller than their hydrophilic counterparts. Taking into account the structure of Ngln compared to the amino acid, it is evident that a region of hydrophobicity has been broken up in the case of the imino acid; whilst there are three hydrocarbon groups in a row for L-glutamine (NH_3^+ - CH - CH_2 - CH_2 -), the alpha-carbon (Ca) has

been segregated from the other two methylene groups (- CH_2 - NH_2^+ - CH_2 - CH_2 -). It should be noted that Ca is converted from a methanetriyl (> CH -) to a methylene group (CH_2) upon changing the location of the side chain to the nitrogen. This structural change makes this region less hydrophobic. This is evident in radial distribution functions for the number of water oxygen atoms around the adjacent methylene carbons, Ca and Cn (Figure 4.4). Both sets of functions show an inward movement of the position of the first coordination shell and a more prominent peak which correlates to a coordination number approximately two atoms higher than for L-glutamine. The coordination numbers for these groups are found in Table 4.2. Additionally, for secondary ammonium groups, there is more partial charge situated on the adjacent methylene groups by Nb. For instance, whilst the β-methylene for L-glutamine ($CmHm_2$) has an overall net charge of 0, the equivalent group for Ngln ($CnHm_2$) has a net charge of $+0.29$. Incorporating this factor as well into the EPSR simulation produces two methylene groups which have at least 7 water oxygen atoms situated around them (Table 4.6). Some contribution to these coordination numbers will come from nearby functional groups, but the value suggests that the presence of methylene groups does not repel water molecules. Additionally, an RDF for the interactions between alkyl hydrogen atoms, Hm, with water (Figure 4.4(c)), though a collated function considering interactions through all Hm atoms, indicates that there are more water molecules surrounding the hydrocarbon groups of Ngln. Ngln appears to want to be surrounded by water, with even the methylene groups forming more van der Waals bonds with water molecules compared to L-glutamine. This redistribution of functional groups and charge could account for the increase in aqueous solubility seen for imino acids.

The interaction of water with alkyl groups has been observed in other neutron diffraction studies of biomolecules. In studies of phosphatidylcholine (134; 135), the hydrogen of a methylene group coordinated up to 1 water hydrogen atom. The methylene groups were attached to the nitrogen of a quaternary ammonium salt (- N^+ - $(CH_3)_3$). For studies on indole systems (118), bonds were being formed between the pyrrole ring and water molecules, as well as hydrophilic portions of methanol.

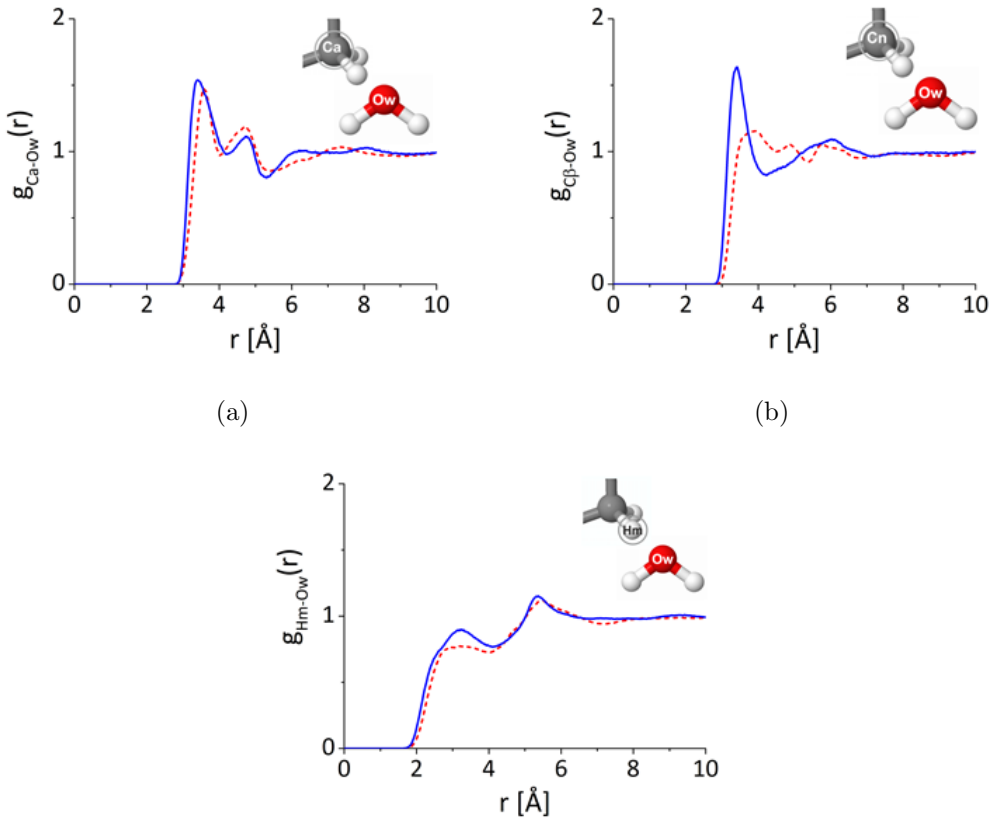


Figure 4.4: The pairwise RDFs for the (a) α-carbon - water (Ca-Ow) interaction, the (b) β-carbon-water interaction (Cm-Ow for amino, Cn-Ow for imino) and the (c) alkyl-hydrogen-water interaction (Hm-Ow) for L-glutamine (red, dashed) and Ngln (blue, solid).

In the work on sarcosine and L-alanine by Uejio *et al* (82), a similar result was observed. For sarcosine, electronic orbitals were not only present for zwitterionic groups, as for L-alanine, but also the alkyl groups. An electronic redistribution could increase the partial charge on these groups, explaining the potentials used and the results observed for Ngln.

4.4.3 Impact on the Backbone Carboxylate

Though the amino and imino acid differ according to changes on the backbone, the backbone carboxylate is common still to both molecules. This subsection aims to determine

how changes to the backbone have altered the hydration of this functional group.

RDFs for the interactions of water oxygen atoms with the backbone carbon, Ob, and the backbone oxygen, Cb, are presented in Figure 4.5. These functions indicate that the hydration of the carboxylate is similar for both molecules. For water molecules coordinated through the Ob atoms, the hydrogen bonding strength is comparable, with the peak position remaining at 2.64 Å. Coordination numbers for this interaction (Table 4.4) suggests that the hydration of these atoms has not changed, with the number of Ow atoms coordinated increasing only by 0.1 for the imino acid.

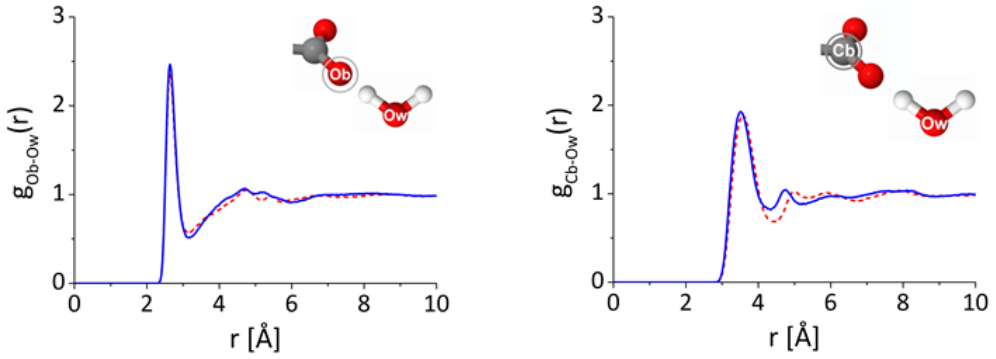


Figure 4.5: The pairwise RDFs for the (a) carboxylate oxygen - water interaction (Ob-Ow) and (b) carboxylate carbon - water (Cb-Ow) interaction, for L-glutamine (red, dashed) and Ngln (blue, solid).

For the Cb atom, there is a subtle increase in hydration, given by a 0.3 increase in coordination number and a 0.06 Å shift in the position of the first coordination shell peak position to 3.51 Å for Ngln. As each of the two Ob coordinates 2.9 Ow atoms, whilst Cb coordinates 9.1, this suggests that water molecules are being brought closer to the molecule not through interactions via the Ob atoms. In studies of carboxylate groups for L-glutamic acid (127) and L-proline (128), the carboxylate oxygen was also able to coordinate water molecules. As this group is attached to the C_α atom that coordinates more water molecules, it is likely it is this that is bringing water molecules closer to Cb. Therefore, despite changes to other portions of the backbone, the hydration of the carboxylate remains fairly consistent between glutamine amino and imino acid

molecules.

Table 4.4: First coordination shell numbers for carboxylate - water interactions for L-glutamine (Gln) and Ngln.

Atom Label	Molecule	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Cb-Ow	Gln	1	4.35	3.57	8.8	0.8
	Ngln	1	4.35	3.51	9.1	0.9
Ob-Ow	Gln	1	3.15	2.64	2.8	1.0
	Ngln	1	3.15	2.64	2.9	0.9

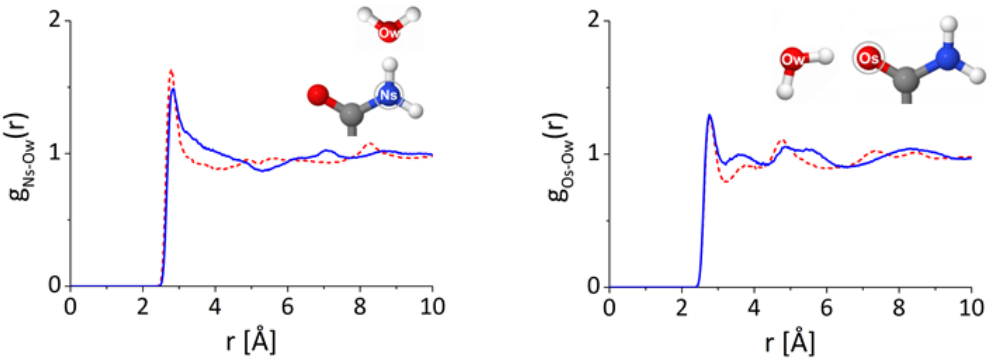
4.4.4 Impact on Side Chain Hydration

A given amino or imino acid is characterised by the side chain group it possesses. In peptoid chains, the loss of the backbone hydrogen makes assembly of peptoid sequences dependent on the side chain groups they contain, which can be tuned to suit their function (34; 36). When exploited in the production of protein mimics, it can potentially be useful for a peptoid sequence to not lose the favourable interactions a side chain may make with its environment or with other biomolecules, which it would have had in peptide form.

This subsection focuses on the interactions the side chain makes with its aqueous environment. Glutamine is characterised by its functional group, - CH_2 - CH_2 - $CONH_2$. The amide group it contains makes this molecule good at hydrogen bonding. The oxygen situated on the carbonyl portion is able to accept hydrogens, whilst the amine portion can donate hydrogen atoms. A lone pair of electrons can also be found on this group. This pair is derived from the nitrogen on the amide group, but is delocalised over the whole group.

These hydrogen bonding interactions can be visualised in radial distribution functions of the interactions of water oxygen atoms with the amide oxygen and nitrogen (Figure 4.6). Both functions suggest that direct interactions between water and the side chain amide remain relatively similar between the two biomolecules, despite changes to

the backbone. Coordination numbers such as that for the side chain nitrogen (Table 4.5), suggest that there has been a 0.8 molecule increase in water coordination for the imino acid. The radial distribution functions suggest there is more water coordination in outer shells for Ngln, which is likely related to the presence of water molecules interacting with other parts of the molecule.



(a) (b)

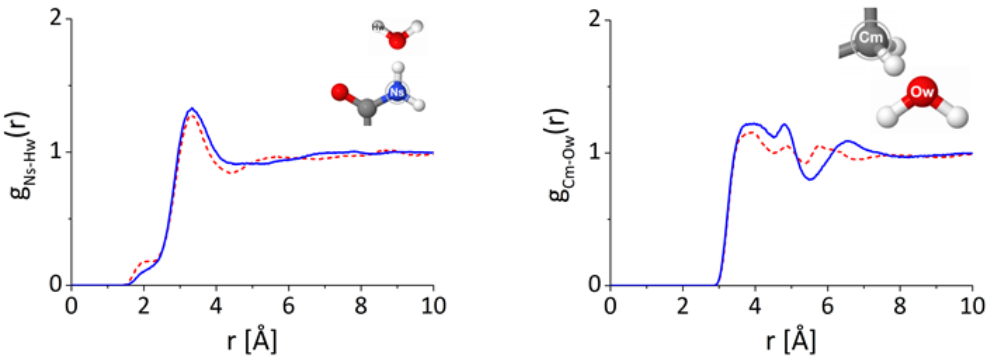
Figure 4.6: The pairwise RDFs for the (a) amide nitrogen - water (Ns-Ow) and the (b) amide oxygen - water interaction (Os-Ow) interaction for L-glutamine (red, dashed) and Ngln (blue, solid).

With the amide group, as well as forming hydrogen bonds with the oxygen atoms and to the amine portion via the hydrogen atoms, coordination can take place via a lone pair sourced from the nitrogen atom that is conjugated over the amide group. Even for Ngln, this bond is still present, signified by short-distance contacts being present in the 2 Å , range, as shown in Figure 4.7(a). Even for the methylene group adjacent to this group, an RDF for this group indicates that more water molecules surround this more hydrophobic group. It should be noted that the Cm-Ow distribution function in Figure 4.4(b) is the average for both Cm atoms on the side chain of L-glutamine, though one of these methylene groups is the equivalent β-carbon to Cn for Ngln, shown previously in Figure 4.4(b). Being an average, including a methylene close to the zwitterionic backbone, this further highlights the difference in hydration state between amino and imino acid forms.

Table 4.5: First coordination numbers for side chain - water interactions for L-glutamine (Gln) and Ngln.

Atom Label	Molecule	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Os-Ow	Gln	1	3.21	2.76	2.3	0.8
	Ngln	1	3.21	2.76	2.3	0.9
Ns-Ow	Gln	1	3.42	2.85	2.8	0.9
	Ngln	1	3.42	2.85	3.6	1.0

Whilst these results suggest that side chain interactions are maintained, this study is limited by the concentration used. The effects observed will also be dependent on characteristics of the side chain group, including hydrophobicity and chain length. In the next chapter, interactions of Ngln at 300 mg ml^{-1} will serve to gauge the effects of concentration on hydration and association through this group. In Chapter 5, when covering the interactions of sarcosine, with a side chain of just a single methyl group, it will also be possible to compare the interactions to those of Ngln to determine any deviance in interactions with varying side chain.



(a) (b)

Figure 4.7: The pairwise RDFs for the (a) amide nitrogen - water (Ns-Hw) interaction and the (b) methylene carbon - water interaction (Cm-Ow) for L-glutamine (red, dashed) and Ngln (blue, solid).

4.5 The Properties of the Glutamine Imino Acid

Following on from the information provided in Section 4.4 on Ngln at $30\,\mathrm{mg\,ml^{-1}}$, this section now presents the interactions of this molecule at a higher concentration of $300\,\mathrm{mg\,ml^{-1}}$. Simulation boxes were set up containing 100 biomolecules to 2700 water molecules, corresponding to a molecular ratio of 1 imino acid to 27 water molecules. The atomic number density used for this simulation was 0.1054 and is derived from density measurements taken for the molecule (Appendix C). In this section, results will be presented for Ngln at 25 °C.

For water, to improve the EPSR fit to the experimental data, the bond length for water was increased slightly to 0.980 Å, compared to 0.976 Å that was used for dilute studies. This change has also been implemented into Ngln simulations at all temperatures (Section 4.5.3), as well as those for sarcosine presented in Chapter 5. This change in the required bond length is due to the increase in biomolecular concentration, where the higher density is slightly altering the intramolecular structure of water.

In studies for L-glutamine at $30\,\mathrm{mg\,ml^{-1}}$, it was not possible to probe details such as the conformation and intermolecular association between L-glutamine molecules. These properties will here be studied instead for Ngln, utilising the information provided from studying this molecule at a higher concentration. Section 4.5.1 will provide details on the molecular conformation, including the fits to the neutron diffraction data. Section 4.5.2 will discuss the impact of concentration on the hydration of Ngln. Finally, Section 4.5.3 will cover the intermolecular association of Ngln in solution.

4.5.1 Conformational Studies

In the previous chapter, the low concentration of L-glutamine resulted in low statistics on the conformation. At a high concentration, it is now possible to probe the conformation in more detail and potentially greater certainty.

Neutron diffraction studies of biomolecules are typically completed without dihedral constraints, with the exception of groups constrained by electrostatic hybridisation, in-

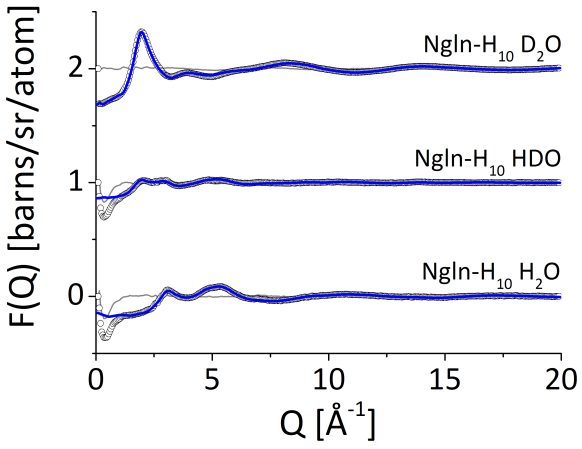
cluding amide and aromatic groups (1; 138–140). Few studies of the structural properties of liquids using neutron diffraction and EPSR have been presented that test the impact of dihedral angles on the fit to the data. The exceptions are studies of glycerol (101; 102) and propanol (117).

For the data at 25 °C, two EPSR simulations were completed, where:

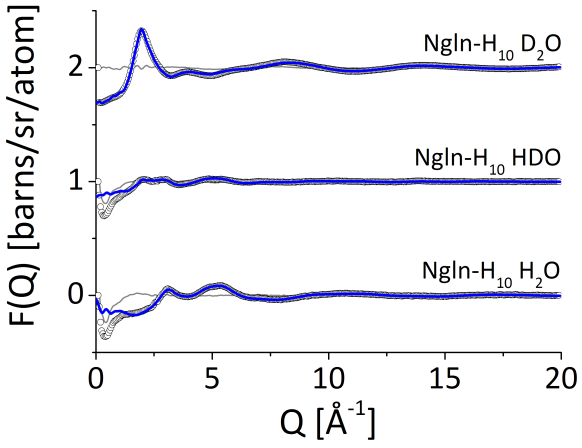
1. All dihedral angles in Ngln were fixed.
2. Only amide and carboxylate dihedral angles in Ngln were fixed.

The dihedral angles constrained in simulation 2 are highlighted in red in Table B.12 of Appendix B. The simulations were accumulated for ∼6000 iterations. The fits for these two simulations are shown in Figure 4.8. The R-factors for these are 4.2×10^{-4} for Simulation 1 and 6.0×10^{-4} for Simulation 2. Using these values and the difference between data and simulation, as shown by the grey lines for each dataset, the results indicate that Simulation 1, a model with fixed dihedral angles, provides a better fit to the data.

The results of this comparison suggest that only certain structures are present in the experimental data. Whilst the fixed-dihedral angle structure used in Simulation 1 is a conformation that could be present in the data, it does not rule out the possibility that other structures could be present also. The output of the TORANGLES routine (Section 2.2.3.5) for select dihedral angles is presented in Figure 4.9, with the value obtained from the AM1 model specified. The result here suggest that the Ngln is affected by the conformation, as was found for sarcosine (82). As referred to previously in Section 3.3.2, the $\alpha\beta$ dihedral conformation was found to produce the best fit to the experimental data for pure glycerol (101).



(a) Simulation 1

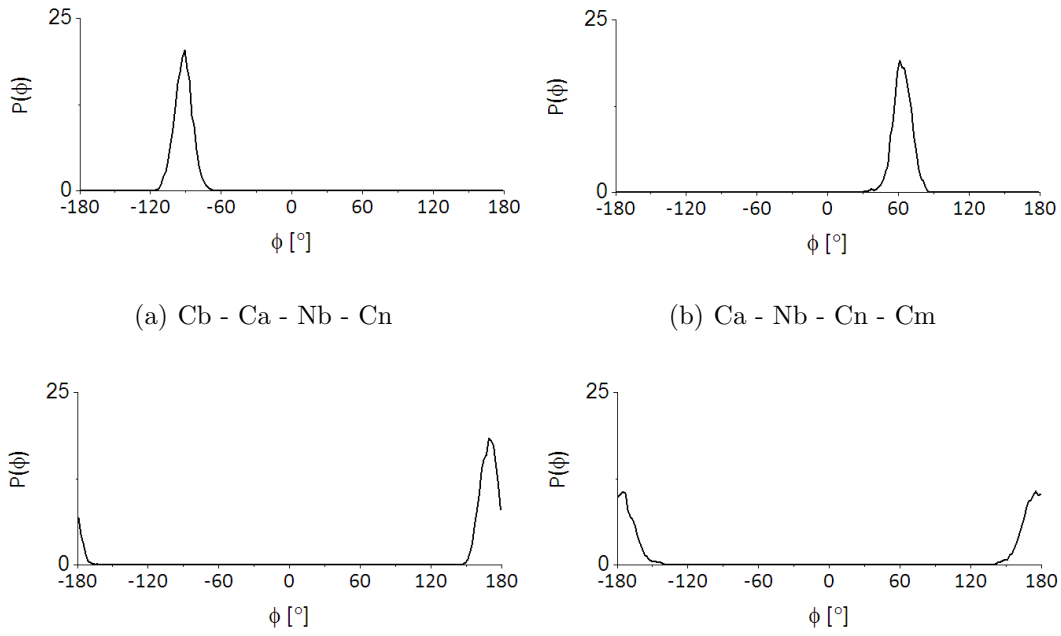


(b) Simulation 2

Figure 4.8: The fit of the total simulated structure factor, $F(Q)$, to the experimental neutron diffraction data, $D(Q)$, for Ngln with dihedral angles (a) fixed and (b) only amide and carboxylate dihedral angles fixed. The data is shown by the black circles, whilst the fit is the line in blue. The difference between data and simulation for each dataset, $D(Q) - F(Q)$ is shown by the grey line.

Like the limitations for L-glutamine discussed in Section 3.3.2, the reason for Simulation 1 producing the better fit could be linked to the way dihedral angles are defined. Using a single equation to control all of the parameters, as is the case for this version of EPSR, is very limiting and could result in certain geometrical parameters being under or over constrained in manner that means the structures sampled do not fit those in the data. By using a pseudo bond to define dihedral angles, the rigidity of the molecule is dependent on the degree of constraint elsewhere in the molecule. For Simulation 2,

having only the amide group constrained may weaken the dihedral angles that have been introduced into the model. This could result in this amide group and molecule taking on physically implausible structures that could produce a poorer fit to the data, given its size and achiral backbone.



(a) Cb - Ca - Nb - Cn

(b) Ca - Nb - Cn - Cm

(c) Nb - Cn - Cm - Cs

(d) Cn - Cm - Cs - Ns

Figure 4.9: Select dihedral angle, ϕ, distributions found for fixed simulations of Ngln. The distributions are for (a) Cb - Ca - Nb - Cn (inputted value -90.3°), (a) Ca - Nb - Cn - Cm (inputted value 61.4°), (a) Nb - Cn - Cm - Cs (inputted value 170.5°) and (a) Cn - Cm - Cs - Ns (inputted value 177.1°) dihedral angles.

4.5.2 The Hydration of the Glutamine Imino Acid with Increasing Concentration

Section 4.3 focused on the hydrophilicity of Ngln and how it is hydration compared with its amino acid equivalent. These studies showed that van der Waals bonds were being formed with alkyl groups in the molecule, potentially explaining Ngln's increased solubility. This section now focuses on how Ngln - water interactions vary with concentration.

Figure 4.10 presents 6 sample RDFs for the hydration of Ngln at both $30\,\mathrm{mg\,ml^{-1}}$ (solid line) and $300\,\mathrm{mg\,ml^{-1}}$ (dashed line). These functions are hydration interactions for the backbone, key alkyl groups known to form van der Waals interactions with water and for the side chain. With the exception of Figure 4.10(a), the height of the first coordination shell has increased for these interactions, suggesting that hydrogen bonds are more likely to be formed at a particular intermolecular distance. This is due there being more of the biomolecule in the solution, increasing the density of the solution. The bond length ascertained from these RDFs either remains consistent for both concentrations or increases for the concentrated solution slightly. For example, the distance between atoms between Ca and Ow atoms increases to 3.48 Å for the $300\,\mathrm{mg\,ml^{-1}}$, compared to 3.39 Å for the dilute mixture. Where the bond length increases, this suggests a weakening of the hydrogen bonds between water and the imino acid, due to the increased concentration. Generally, these subtle changes signify a slight rearrangement of water molecules in the hydration shell.

The one RDF that shows a decrease in the hydrogen bond strength is that for Nb-Ow. This suggests that Ngln preferentially interacts with water by enhancing the bonds made between the carboxylate groups and van der Waals interactions, instead of the ammonium. The greater density of water molecules in other portions of the molecule is indicated by the larger second coordination that has moved from 4.73 Å for the dilute solution to 4.60 Å. In studies of L-glutamine, the carboxylate group was the most successful at coordinating water molecules (1).

Coordination numbers for these groups are presented in Table 4.6. Whilst the RDFs suggest changes to the probability of finding atoms a specific distance, given by the change in the height of the first coordination shell, the coordination numbers, even for Nb-Ow, suggest there is generally little change in the hydration. Interestingly, whilst the number of water oxygen atoms interacting with methylene groups was equal for the dilute system, there is more preference for water oxygen atoms to interact with Ca instead of Cn. The Ca-Ow pairwise interaction has a coordination number of 8.5, whilst the number for Cn is 6.5. Any variation in coordination will be the result of the increased amount of imino acid molecules preventing water from hydrating certain areas so effectively due to access. Overall, the hydration of Ngln is not significantly influenced by the concentration, with any variation being the result of density effects.

Table 4.6: First coordination numbers for backbone - water interactions for Ngln at $30\,\mathrm{mg\,ml^{-1}}$ and $300\,\mathrm{mg\,ml^{-1}}$.

Atom Label	Concentration ($\mathrm{mg\,ml^{-1}}$)	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Nb-Ow	30	1	3.21	2.73	2.2	0.5
	300	1	3.29	2.77	2.1	0.7
Ca-Ow	30	1	4.20	3.39	7.6	1.3
	300	1	4.35	3.48	8.5	1.5
Cn-Ow	30	1	4.23	3.42	7.4	1.2
	300	1	4.20	3.42	6.54	1.3
Ob-Ow	30	1	3.15	2.64	2.9	0.9
	300	1	3.21	2.64	2.7	0.7
Os-Ow	30	1	3.21	2.76	2.3	0.9
	300	1	3.18	2.76	2.1	0.8
Ns-Ow	30	1	3.42	2.85	2.8	0.9
	300	1	3.42	2.85	3.5	1.1

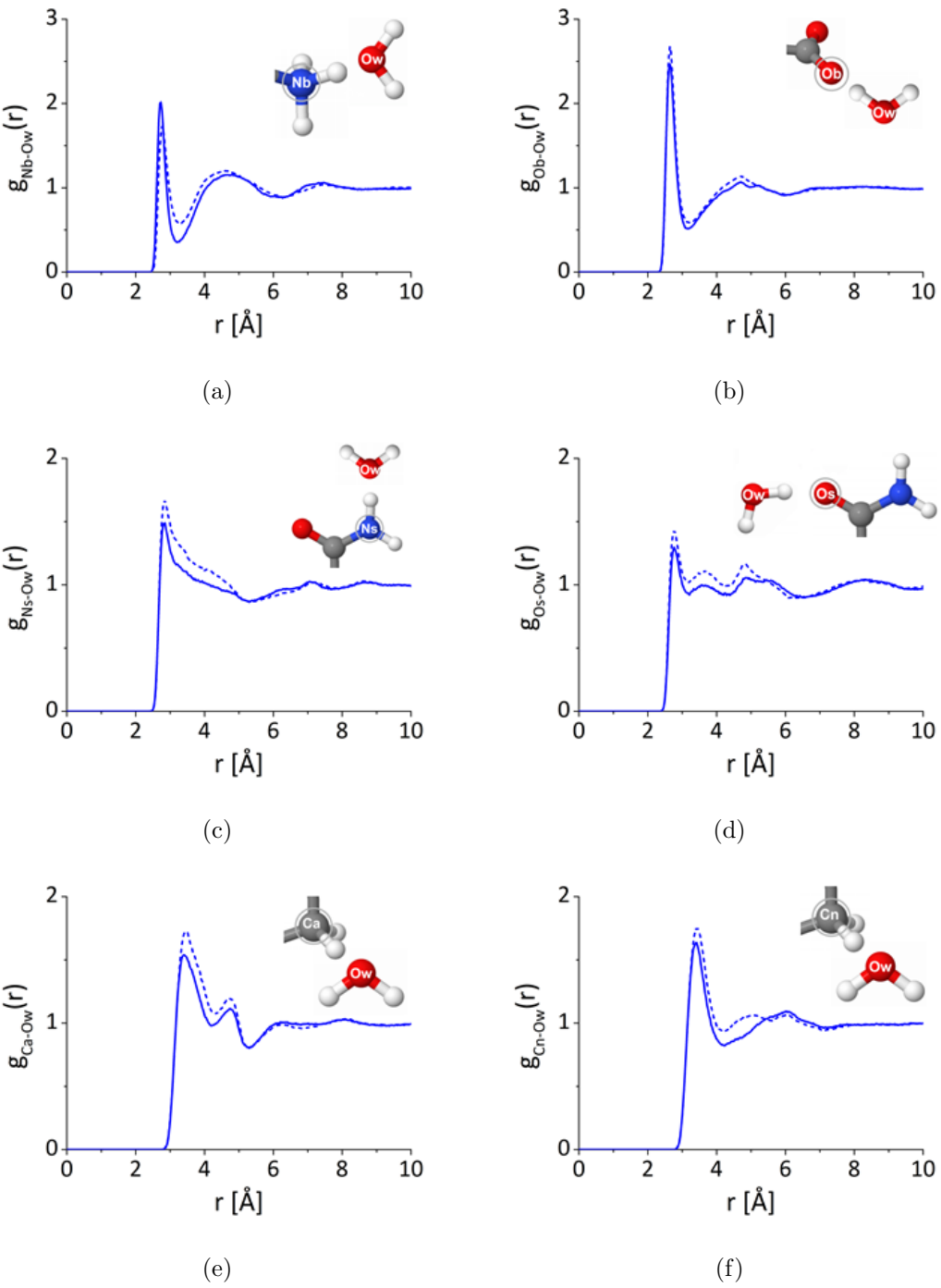


Figure 4.10: The pairwise RDFs for Ngln - water at $30\,\mathrm{mg\,ml^{-1}}$ (solid line) and $300\,\mathrm{mg\,ml^{-1}}$ (dashed line). The functions are for the (a) backbone nitrogen - water (Nb-Ow), (b) backbone oxygen - water (Ob-Ow), (c) side chain nitrogen - water (Ns-Ow), (d) side chain oxygen - water (Os-Ow), (e) α-carbon - water (Ca-Ow) and (f) β-carbon - water (Cn-Ow) interactions.

4.5.3 The Association of Glutamine Imino Acid Molecules in Solution

Whilst Section 4.3 has elucidated a potential cause for the increased aqueous solubility of imino acids, the ability of such molecules to associate or interact with other biomolecules is also of importance. For most applications of relevance for these molecules, including pharmaceutics and nanotechnology, interactions with other matter is vital for their function. In the case of peptoids, they interact through bonds forming with side chain functional groups. Ngln is ideal to study the prevalence of hydrogen bonding between molecules, an interaction not considered to be important for the assembly of peptoid structures.

Association of imino acids through hydrophilic functional groups in Ngln has been investigated. These polar groups are the ammonium cation, the carboxylate anion and the dipolar amide. Two atoms are defined as interacting if they are separated by a distance within the first coordination shell of the pairwise interaction, the maximum distance determined by the respective RDF. As has been done previously, the interactions between an oxygen atom and a hydrogen atom on two separate molecules have been studied. Considering atoms on both the backbone and side chain, there are four interactions that can be studied; backbone-backbone (Ob-Hb), backbone-side chain (Ob-Hs), side chain-backbone (Os-Hb) and side chain-side chain (Os-Hs).

The RDFs for the four interactions are shown in Figure 4.11. All four functions suggest that hydrogen bonds can be formed between hydrophilic groups via all interaction combinations. The molecules readily interact with each other through the backbone oxygen, backbone-backbone and backbone-side chain, with evident hydrogen bond peaks present for these. The backbone-backbone (Ob-Hb) interaction is the strongest with a peak bond length of 1.68 Å. This interaction is comparable to that which would occur between amino or imino acid building blocks to form sequences of residues and this interaction is prevalent as the functional groups are partially charged. Backbone-side chain interactions were also commonly found in studies of aqueous L-glutamine (1), showing that at both low and high concentrations these interactions prevail. Interactions involving the side chain oxygen (Os) are minimal, with no clear peak being present for side chain - side chain interactions. For L-glutamine, interactions through the side chain,

though less dominant than those via the backbone, were more prevalent (Section 3.5).

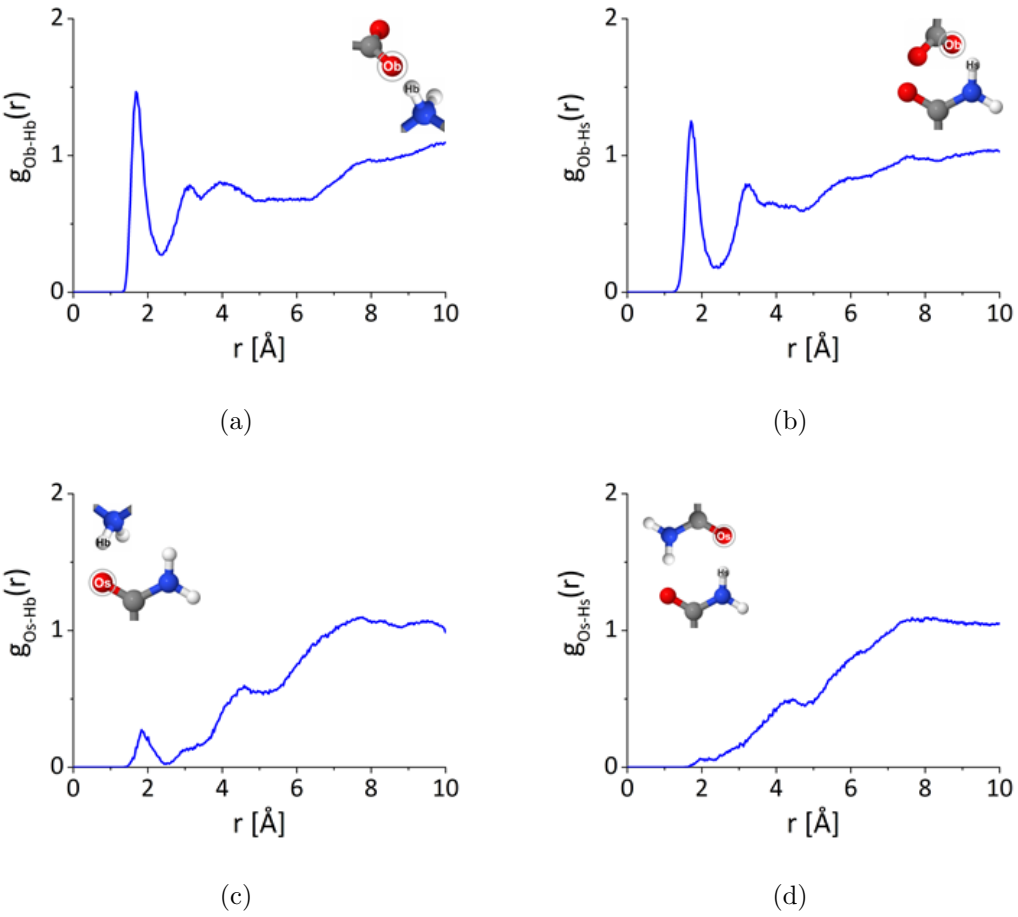


Figure 4.11: The pairwise RDFs for Ngln - Ngln interactions at 25 °C. The functions shown are for the (a) backbone - backbone (Ob-Hb), (b) backbone - side chain (Ob-Hs), (c) side chain - backbone (Os-Hb) and (d) side chain - side chain (Os-Hs) interactions.

The CLUSTERS routine (Section 2.2.3.2)has been used to determine the number of molecules that can associate to form a cluster in the simulation. Figure 4.12 portrays the size of clusters produced through all four hydrophilic interactions. The results show that Ngln clusters of size 2–6 are present in the EPSR simulation. Of these clusters, 64% of the clusters are dimers, with less than a percent attributable to 5- or 6-mers. Despite evidence of self-association, the monomer is the dominant species in the solution. What can also be observed in Figure 4.12 is that the probability of the largest sized clusters is extremely low, with the monomer being the dominant species. Referring to Section 1.2.3 aqueous L-proline at a ratio of 1:10 forms 8-mers in solution (128). For studies

of dipeptide glycyl-L-alanine, glycine-L-proline and L-alanyl-L-proline (131), clusters of up to size 50 were produced for glycyl-L-alanine, the most hydrophilic of the dipeptides, which was the highest cluster size over all the molecules studied. The prevalence of interactions will be dependent on the concentration of the biomolecule, but the result supports the idea that Ngln preferentially interacts with water over other Ngln molecules.

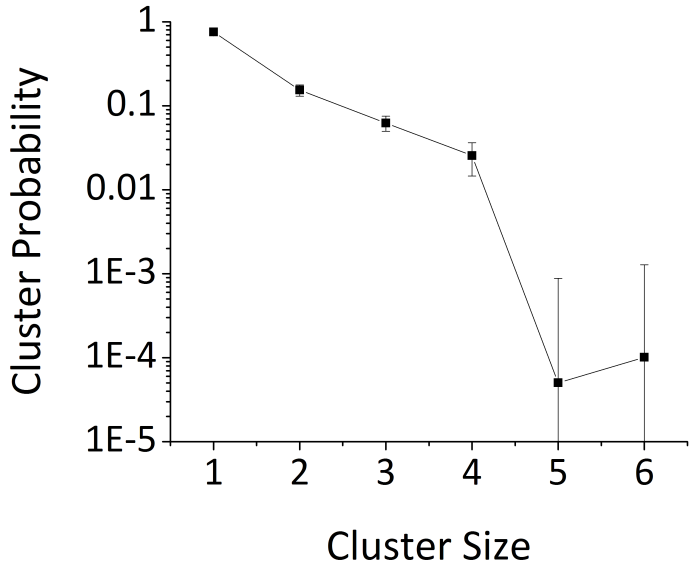


Figure 4.12: Probability of hydrophilic Ngln clusters forming versus the size of the clusters observed, at 25 °C. The probability scale is logarithmic.

In Table 4.7, the coordination numbers for the four association interactions are shown. It should be noted that the coordination numbers are rounded to 1 decimal place and so a coordination number of 0 does not equate to there being 0 atoms within the distance. The highest value is for backbone oxygen (Ob) atoms. With the highest value being 0.1, this signifies that the average imino acid oxygen will not be in close proximity to an imino acid amide or ammonium hydrogen atom. These low numbers suggest that the interactions are transitory, forming and breaking instantly. Stable clusters of the molecule are not found in the solution.

Table 4.7: First coordination shell numbers for Ngln - Ngln interactions.

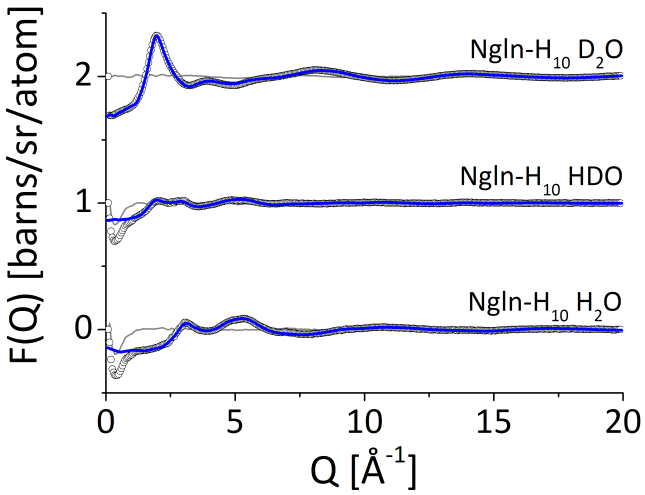
Atom Label	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Ob-Hb	1	2.37	1.68	0.1	0.2
Ob-Hs	1	2.39	1.71	0.1	0.2
Os-Hb	1	2.45	1.83	0	0.1
Os-Hs	1	2.35	1.96	0	0.1

4.6 The Association of the Glutamine Imino Acid with Increasing Temperature

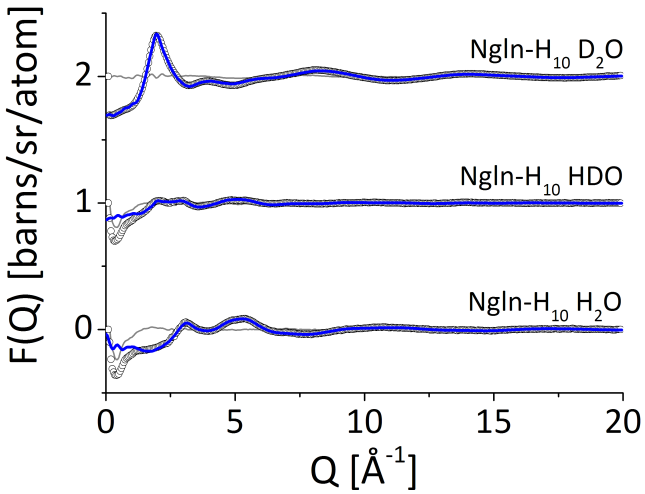
As well as room temperature, experiments have also been completed at 37 °C, physiological temperature, and 60 °C, a typical temperature for protein denaturation to take place (227). Peptoids have the property of increased resistance to extreme temperature (34; 36). The exact mechanism for this is still not fully determined, but is potentially due the lack of hydrogen bond donors on the backbone. Ngln contains both partially-charged zwitterionic groups and a side chain dipole system. Therefore, it will also be possible to observe which types of interactions are dominant at each temperature additionally determining the general clustering ability of the molecules. This may provide insight into what interactions are important for assembly, particularly at higher temperatures.

Simulation boxes for the 300 mg ml^{-1} Ngln system at 37 °C and 60 °C were set up, containing 100 biomolecules to 2700 water molecules, corresponding to a molecular ratio of 1 imino acid to 27 water molecules. The atomic number density used for the 37 °C simulation was 0.1049, and 0.1036 for the 37 °C simulation, both values derived from density measurements taken for the molecule (Appendix C). The fits to the neutron diffraction structure factor data are shown in Figure 4.13. The R-factors were 4.5×10^{-3} and 3.1×10^{-3} for the 37 °C and 60 °C simulations respectively. In this section, the results will be compared to EPSR simulations for the imino acid at 25 °C.

Figure 4.14 presents the RDFs for the four hydrophilic interactions at all temperatures. In all cases, backbone-backbone interactions are the most prevalent. Backbone-backbone interactions are required to produce chains of amino acids and it seems that even in imino form, though this group possesses one less hydrogen and less charge overall, it will readily interact with the carboxylate of another molecule. For each hydrophilic interaction, the lowest prevalence is found at 60 °C.



(a)



(b)

Figure 4.13: The fit of the total simulated structure factor, $F(Q)$, to the experimental neutron diffraction data, $D(Q)$, for Ngln at (a) 37 °C and (b) 60 °C. The data is shown by the black circles, whilst the fit is the line in blue. The difference between data and simulation for each dataset, $D(Q) - F(Q)$ is shown by the grey line.

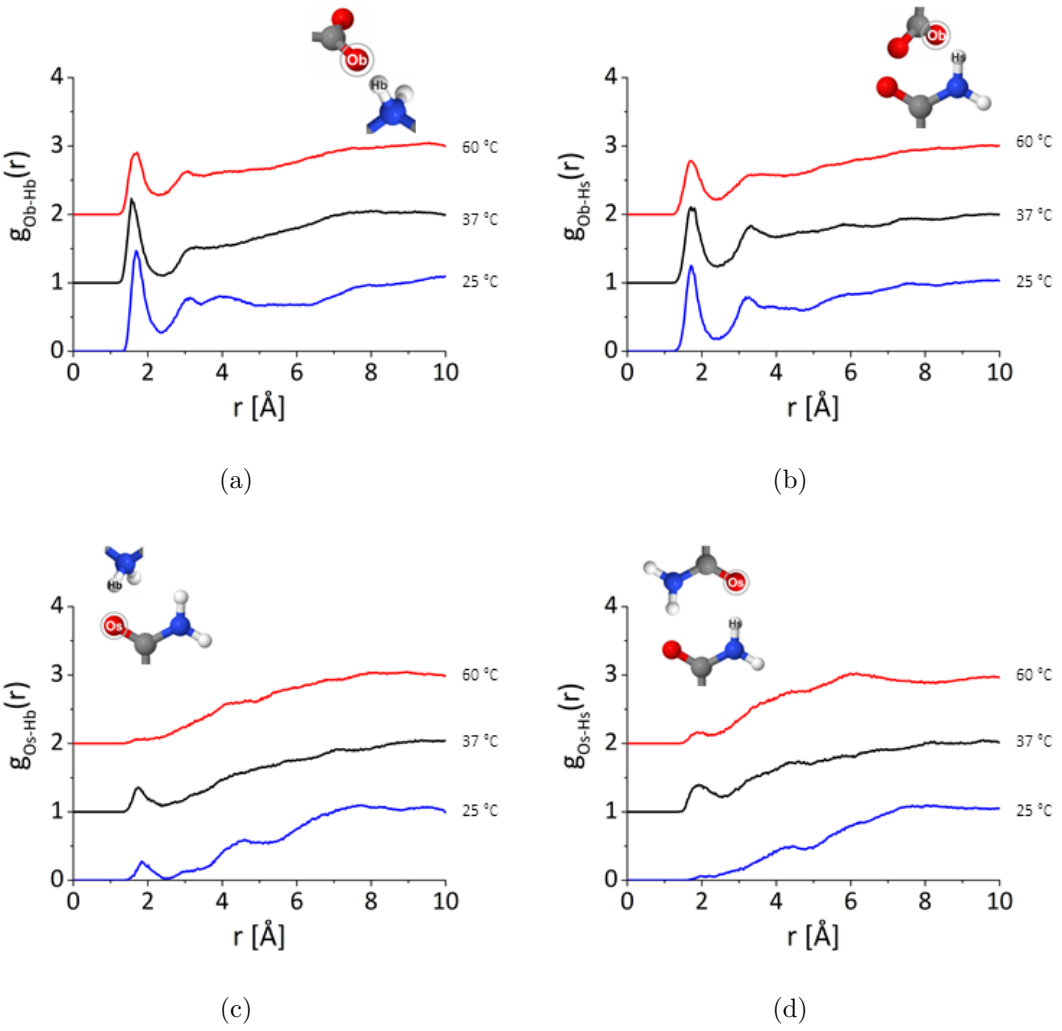
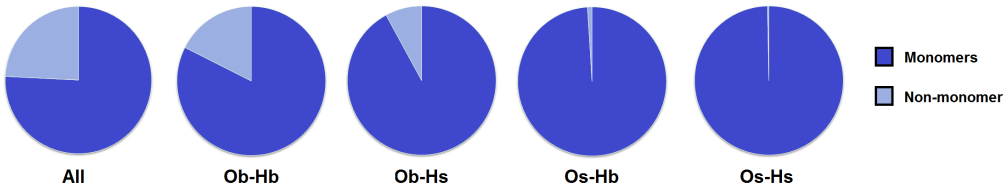


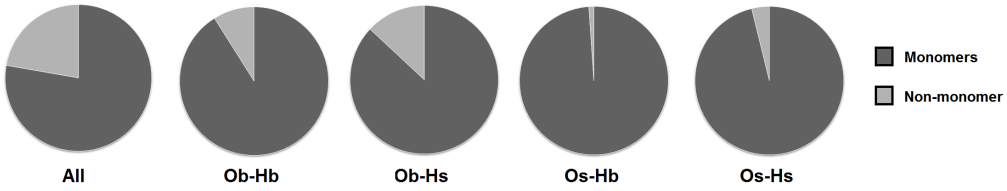
Figure 4.14: The pairwise RDFs for Ngln - Ngln interactions at 25 °C (blue), 37 °C (black) and 60 °C (red). The functions are for the (a) backbone - backbone (Ob-Hb), (b) backbone - side chain (Ob-Hs), (c) side chain - backbone (Os-Hb) and (d) side chain - side chain (Os-Hs) interactions.

For interactions involving the side chain oxygen atom, the situation is more complex. Instead of following a trend with temperature, the interactions appear to fluctuate. Surprisingly though, the RDFs suggest that side chain oxygen interactions prevail at 37 °C, before declining again at 60 °C. Due to these interactions, clusters of size 2–8 can be produced via all interactions at 37 °C, which declines again at 60 °C. Observing the percentage of Ngln molecules involved in a cluster, shown in Figure 4.15, it is evident that the clustering through side chain - side chain (Os-Hs) interactions has increased in percentage (Figure 4.15(b)). For side chain-side chain interactions, this is one that is thought to play a role in neurodegenerative disease. This interaction is purely dominated

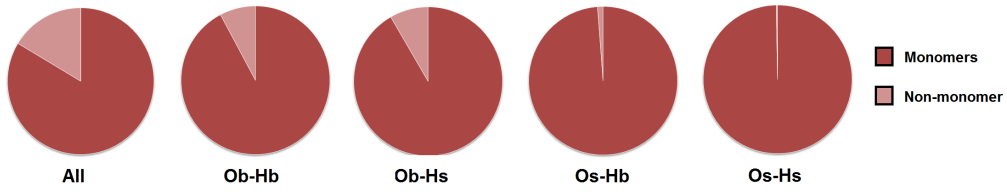
by hydrogen-bonding and suggests that its influence is condition-dependent. In the realms of pharmaceutical research, peptoid chains with polar, uncharged residues have been implemented as diagnostic and therapeutic tools for disease. For these applications, which would take place under physiological conditions, utilising residues with polar, uncharged side chains could be appropriate. Nevertheless, the increase in side chain interactions does not influence the overall percentage of Ngln molecules involved in a cluster, with approximately 25% of molecules involved in a cluster at 25 °C, on a similar scale to studies of L-glutamine (Section 3.5), despite the concentration difference, with a notable decrease seen between 37 °C and 60 °C.



(a) 24 °C



(b) 37 °C



(c) 60 °C

Figure 4.15: Percentage of Ngln molecules existing as monomers and in hydrophilic intermolecular clusters of 2 or more (non-monomer), at (a) 25 °C (blue), (b) 37 °C (grey) and (c) 60 °C (red).

The persistence of hydrophilic interactions was discussed previously in Chapter 3 for L-glutamine in aqueous solution. For L-glutamine, there was no strong evidence to suggest a change in the percentage of molecules involved in clusters with increased

temperature. Here, the enthalpy and entropy contributions were discussed, where positive entropy was driving the formation of clusters, as there is not enough energy had been provided to make the change in enthalpy positive and cause hydrogen bonds to weaken. For Ngln, it is possible that a threshold has been reached at 60 °C for all interactions, where the thermal energy is sufficient to weaken certain intermolecular interactions between imino acids. This threshold appears to be dependent on the interaction type, where side chain interactions prevail and even increase before reaching an energy threshold, potentially being driven by entropy prior. An increase of hydrophilic interactions at higher temperatures was also an observation seen in tertiary-butanol solutions (106), with the conclusion of this work indicating that there is a complex balance between enthalpic and entropic contributions governing association. In the case of molecules containing an amide group, the transfer of energy between the molecule and its environment is facilitated due to one of the vibrational energy bands for this group overlapping with one for water (151). This facilitated transfer of energy is thought to have a role in the ability of amide-residues to aggregate. Upon reaching a transition point though, enough heat is supplied to reduce the stability of some of the polar bonds present. As hydrophilic interactions are less prevalent at 60 °C, which was not observed for the amino acid, the energy required to weaken such bonds is also dependent on the stereochemistry of the molecules.

4.7 Conclusions

This chapter has determined the properties of Ngln, including the source of its increased hydrophilicity and solubility, the molecular conformation, the impact of varying the concentration, association and the effect of varying the temperature.

In Section 4.4 the hydration of Ngln was compared to the results presented Section 3.4 for its amino acid equivalent, L-glutamine, to relate its structural features to its functional properties. It was determined that whilst this change decreases direct water interactions with the ammonium group there is instead greater hydration of the adjacent hydrocarbons, despite them being stereotypically ‘hydrophobic’ groups. This hydrophilic behaviour though does not profoundly influence direct interactions of the

side chain and backbone carboxylate with water, with a small increase in hydration observed.

Section 4.4 presented details for the conformation, hydration and association for Ngln at $300\,\mathrm{mg\,ml^{-1}}$. Section 4.5.1 determined that a fixed dihedral model was better for representing the molecule in EPSR simulations. It was determined that increasing the biomolecular concentration did not significantly influence the number of water molecules the imino acids coordinate or the strength of the hydrogen bonds formed, as presented in Section 4.5.2. Section 4.5.3 highlighted that Ngln is able to associate via the hydrophilic groups it possesses, with the biomolecules associating through all four possible hydrophilic interaction combinations, allowing instantaneous clusters to be formed.

It has also been possible to complete studies at higher temperatures to understand the interactions that govern the properties of Ngln, at physiological and extreme conditions. Side chain interactions were found to initially increase when the temperature is raised to 37 °C, though this change does not affect the trend of the percentage of molecules involved in a cluster decreasing. This suggests that each interaction has its own entropic-enthalpic balance and energy threshold to overcome. The interactions that prevail include those important for peptoid applications and under the right conditions could be used to stabilise structures.

Chapter 5

The Impact of the Side Chain on Imino Acid Properties

Chapter 4 presented a comparison between an amino and an imino acid, as well as discussing the hydration, association and temperature-based properties of the novel Ngln molecule. In the development of peptoid-based materials, much emphasis is placed on the side chain controlling the self-assembly of peptoid chains (36; 37). Thus, it is necessary to consider how the side chain functional group present could affect the properties of peptoids, even down to imino acid level.

This chapter will present a comparison between Ngln with the imino acid sarcosine; a molecule that has been investigated previously. As described in Section 1.3.3, sarcosine is the only naturally-occurring, non-hybrid imino acid recognised biochemically as an osmoprotectant, capable of inhibiting the effect of denaturants such as urea on protein structure (61; 67; 175). Sarcosine was used as a candidate molecule in the first structural comparison between an amino and imino acid, as discussed previously Section 4.2. The structural information supplied in this study was mostly qualitative and there has been no study that determines whether the results found are also observed for imino acids with other side chains. It has been suggested that the increased hydrophilicity seen for sarcosine is correlated to the small size of the side chain, which is inefficient at shielding the backbone nitrogen of the residue (47). It is possible to speculate, then, that the electrostatics and structural properties could vary with side chain functional

group even for an imino acid. Consequently, this chapter will also analyse the hydration and association of sarcosine in aqueous solution, seeking to provide quantitative data to compare with previous literature and determine whether water molecules hydrate both hydrophilic and alkyl groups.

Section 5.2 will present the neutron experiment completed on sarcosine and details of the simulations completed. Section 5.3 will determine the effect of dihedral angles on fits to the neutron diffraction data and assess the flexibility of the molecule. The hydration of the backbone groups for sarcosine will be presented in Section 5.4 and compared to Ngln. The backbone groups were previously hypothesised as being important in the improved properties of imino acid building blocks. However, as the side chain is the varying functional group between Ngln and sarcosine, the hydration of this group (Section 5.4) and a comparison of the hydrophilic and hydrophobic interactions important for association (Section 5.5) will be presented. Given the osmoprotective behaviour of sarcosine, its impact on water structure will be covered in Section 5.6, as well as an analysis on whether Ngln potentially shares the same properties.

5.1 The Osmolytic and Previously Studied Structural Properties of Sarcosine

Osmolytes are small organic molecules that are able provide protection against the effects osmotic stress (63). These molecules are produced readily by cells and organisms and can help to resist against the effects of temperature, pressure and pH that could influence the osmotic balance.

Osmolytes can be categorised according to molecule type. These categories include polyols and sugars, amino acids and methylamines. Example amino acid and methylamine osmolytes are presented in Figure 5.1. Sarcosine is considered a type of methylamine; a molecule containing both a methyl and amine functional group (62; 175). Methylamines are recognised as protein stabilisers and can counteract the effects of denaturing osmolytes such as urea (61; 67; 175). Urea has a dominant role in the metabolism of nitrogen-containing compounds (236), and is accumulated by some species as a key

osmolyte, despite its ability to denature proteins. Other molecules in this category include trimethylamine-oxide (TMAO) and betaine, with numerous studies being carried out on how these particular molecules can counteract the destabilising effect of urea (237–240). The exact mechanism for this is still debated, but it is thought that these molecules act as an osmoprotectant by strengthening the bulk structure of water, by interacting with urea, or both (61; 67; 175).

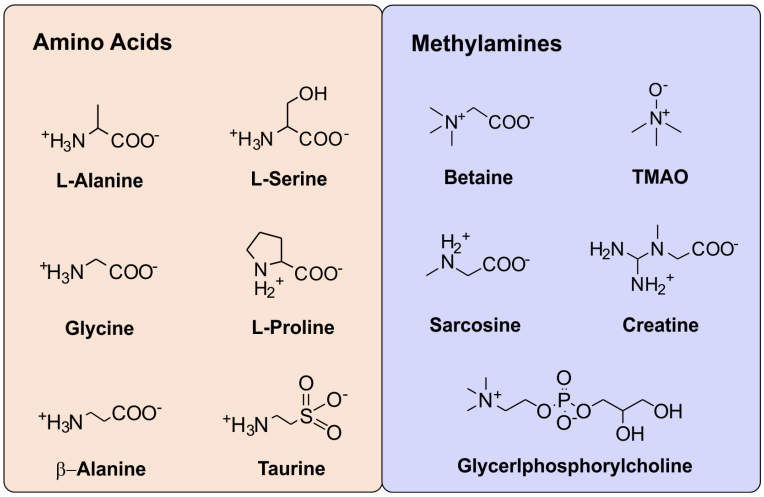


Figure 5.1: Example amino acid and methylamine protecting osmolytes.

Literature on the structure and interactions of sarcosine is select. The crystal structure (241) has been obtained and the molecule has been studied using methods such as Raman spectroscopy (69) and computational modelling (67; 242). The crystallographic structure presented by Mostad and Natarajan determined showed that sarcosine generally took on an *antiperiplanar* form. Antiperiplanar as a term comes from the Klyne - Prelog system (243) for describing conformations about a single bond, and refers to larger angles from -150° to 150°. The Cb - Ca - Nb - Cs was quoted as -166.3°. Structured in the antiperiplanar range has been used as a standard starting model in studies of sarcosine (67; 82) at 180° (*trans* conformation), though altering or freeing dihedral angles has been found to produce good fits to experimental data. Computational work of the zwitterion in the gas-phase suggested that the angle could deviate to 107.3°. Computational studies of the sarcosine-water complex suggested angles of 76 - 81°, though the neutral form of the molecule was utilised (242), whilst NMR research sarcosine in

aqueous solution with lanthanide ions predicted the dihedral was taking on cis angles (-30° to 30°) (244), though a schematic Newman projection in the same paper showed gauche angles (approximately $\pm 60°$). The structure of sarcosine consequently appears to vary.

Radial distribution functions for sarcosine in the liquid phase were obtained by Kumar *et al* in 2013 (67). This group completed MD simulations with the aim of studying the impact of sarcosine on urea in solution, with increasing concentration. Simulations were completed on sarcosine, urea and sarcosine-urea mixtures in solutions, at varying concentrations. With the addition of sarcosine, the hydrogen bonds formed between water molecules increased in probability and strength. Hydrogen bonding between sarcosine molecules was noted, taking place between the carboxylate and ammonium groups of two separate molecules. Intermolecular interactions were observed in the solutions, which decreased with increasing concentration of sarcosine, leading them to question whether sarcosine forms clusters in solution. Cluster formation was investigated by calculating the number of water molecules in the hydration shell of sarcosine, as well as the solute density around a sarcosine molecule. Using these calculations, they concluded that sarcosine did not form clusters in solution. In terms of its osmolytic capabilities, sarcosine was found to interact strongly with urea molecules, potentially preventing urea from interacting with proteins.

What was not in the work by Kumar *et al*, or in other studies, was quantitative information on the interactions of the side chain group with water and other sarcosine molecules. Furthermore, whilst the number of molecules in the hydration shell was presented, no coordination shell numbers were given for the individual pairwise interactions. This chapter will build on this study by determining quantitative details on sarcosine interactions and obtaining coordination numbers for the separate pairwise interactions.

5.2 Experimental Procedure

Neutron diffraction experiments have been completed on an aqueous solution of sarcosine under ambient conditions of 298 K temperature and 1 bar pressure. The molecular ratio of the solution was 1 sarcosine to 27 water molecules; a ratio identical to that used for the studies of Ngln presented in Chapter 4.

Table 5.1: Isotopic substitution experiments completed for Sarcosine (Sar)

Sample No.	Sample Name	Description
i	Sar-D_2O	Fully protiated Sar in D_2O
ii	Sar-HDO	Fully protiated Sar in 50% Milli-Q water and 50% D_2O
iii	Sar-H_2O	Fully protiated Sar in Milli-Q water

For sarcosine, three isotopic substitution experiments were completed, as shown in Table 5.1. These experiments study protiated (Sigma Aldrich, purity 98 %) sarcosine. These samples were studied in milli-Q water, deuterium oxide (Sigma Aldrich, purity 98 %) and a $50-50$% mix of H_2O and D_2O. It had been originally intended that isotopic substitution experiments would also be completed on a methyl-deuterated version of sarcosine (Sarcosine-methyl-D3, Sigma Aldrich, purity 99 %). However, at the time of completing the experiments, the sarcosine-D3 sample was not miscible in protiated or deuterated water and samples containing this solute consequently were not utilised. Speculative reasons for the sample's poor miscibility include impurities, incorrect isotopic substitution or an error in sample preparation.

EPSR simulations of the system were constructed with 100 sarcosine molecules and 2700 waters. The atomic number density of 0.1024 used has been derived from taking density measurements of the sample studied with neutron diffraction (Appendix 3). In the simulations, there are 8 distinct atomic components defined for sarcosine. These components are labelled in Figure 5.2. There are also two components for water in each simulation. The hydrogen atoms in both simulations are labelled as Hb, Ha and Hs, referring to hydrogens on the backbone ammonium, α-carbon and the side chain methyl respectively. The oxygen atoms are Ob, for the two found on the backbone carboxylate

anion. Nb refers to the ammonium nitrogen. For the carbon atoms, Cb is that found on the carboxylate, Ca the α-carbon and Cs the side chain methyl carbon atom. The labels used for water are the same as those shown in Figure 3.1.

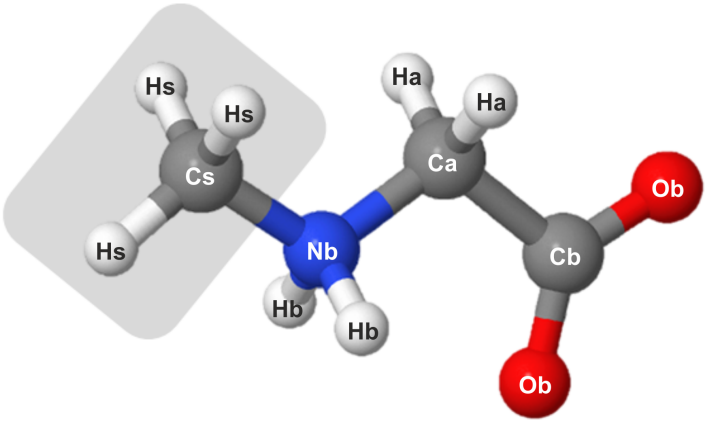


Figure 5.2: Diagram of the sarcosine molecule. The side chain of the molecule is highlighted in grey. C=O double bonds are not shown in this diagram. The labels are those used to identify each of the atomic components in the EPSR simulations. There are 8 distinct atomic components for sarcosine.

5.3 Conformational Studies

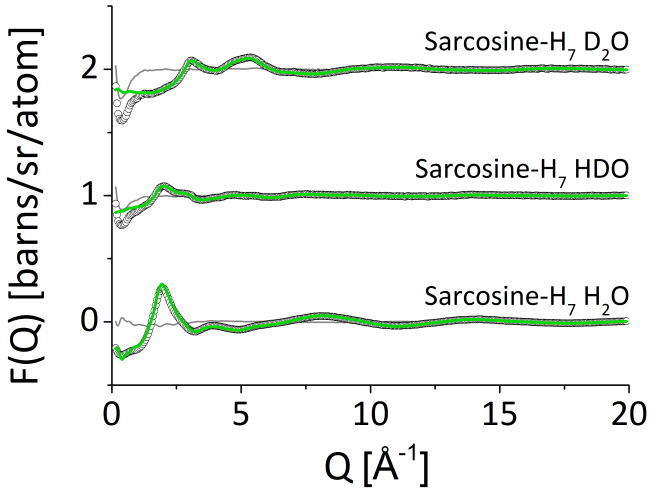
In Uejio *et al*'s NEXAFS study, a comparison was completed between simulations with varying levels of dihedral angle constraint. With the neutron diffraction work completed here, it is possible to undertake a similar comparison. It should be noted that the concentration used in the NEXAFS is 1 M, equivalent to a molecular ratio of 1 sarcosine to approximately 56 water molecules. This is more than half the concentration used for the neutron diffraction study, which is at a concentration of 2.1 M. As a consequence, it is feasible that this study may deviate not due to experimental error, but simply because of conformation restraints imposed by the molecule being in a more concentrated environment.

Two EPSR simulations were completed to test the effects of fixing dihedral angles;

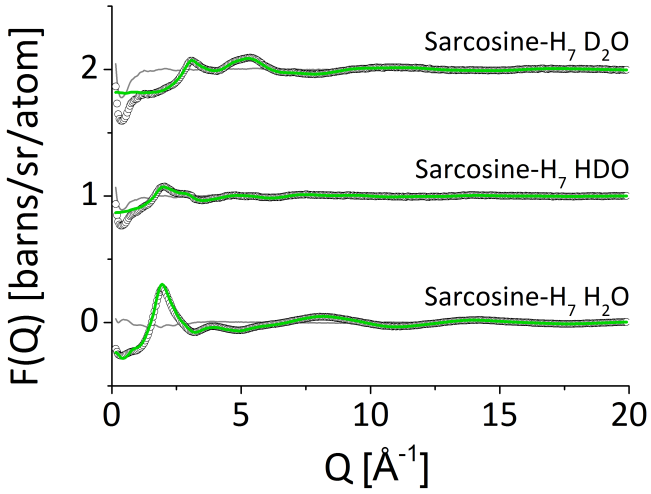
1. All dihedral angles in sarcosine were fixed.
2. No dihedral angles in sarcosine were fixed.

These simulations were accumulated for ∼6000 iterations. The resulting fits to the data are plotted for both simulations in Figure 5.3. The R-factors for these two simulations are very similar, with a Simulation 1 being 4.7×10^{-3} and Simulation 2 being 4.3×10^{-3}, though the slight improvement for the latter suggests that a flexible model can give a slightly improved fit which is in support of Uejio *et al* hydration study.

The Cb - Ca - Nb - Cs dihedral angles sampled are shown in Figure 5.4, with example snapshots presentedfrom the simulation box. Due to the nature of EPSR being able to sample disorder in a system and also produce structures that are not thermodynamically stable, it is expected that higher energy conformations will be observed in the EPSR simulation (138), hence the broadness of the distribution. Whilst it is possible for this angle to be at approaching ∼180°, as defined in the fixed simulation, it is also possible for the molecule to take conformations at small angles. This is comparable to results seen for solution NMR studies of sarcosine (244). For the NEXAFS work, it is possible that being able to sample such conformational populations allowed for a better fit to the experimental data. However, as described by Uejio *et al*, having no constraints is potentially unphysical. The dihedral angle distribution shown is broad, making it hard to differentiate realistic populations in the simulation. Therefore, as well as to be consistent with studies of Ngln described previously (Chapter 4), this chapter will present the results for the simulation with no fixed dihedral angles.



(a) Simulation 1



(b) Simulation 2

Figure 5.3: The fit of the total simulated structure factor, $F(Q)$, to the experimental data, $D(Q)$, for the sarcosine with dihedral angles (a) fixed (Simulation 1) and (b) not fixed (Simulation 2). The data is shown by the black circles, whilst the fit is the line in green. The difference between data and simulation for each dataset, $D(Q) - F(Q)$ is shown by the grey line.

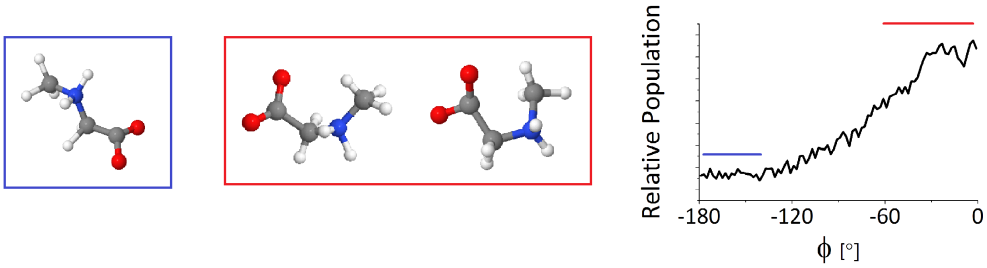


Figure 5.4: Conformations of sarcosine observed in the EPSR simulation.

5.4 The Hydration of Sarcosine

This section will cover the interactions of sarcosine with its aqueous environment. Throughout this section, comparisons will also be made with those presented for Ngln, to ascertain the impact of side chain on interactions.

5.4.1 Hydration of the Backbone Nitrogen

In Chapter 4, the results for the hydration of the backbone nitrogen were first presented. This was the region with greatest structural change between amino and imino acid forms of glutamine, which is thought to be the source for the improved solubility of imino acids. Using Nb-Ow as a first step again, it is evident that the differing side chain has an impact on the hydration of the overall molecule. The Nb-Ow RDF is shown in Figure 5.5 for both sarcosine and Ngln, the latter being the same RDF as presented in Section 4.5.2. The first coordination number for sarcosine is 1.8 (Table 5.2). With the coordination number for each Hb atom being 0.8, this suggests the Nb-Ow first peak comes from interactions between these 2 ammonium hydrogen atoms with a water oxygen atom each. These peaks also correlate to similar bonding distances and thus strengths, with sarcosine at 2.79 Å and Ngln at 2.77 Å.

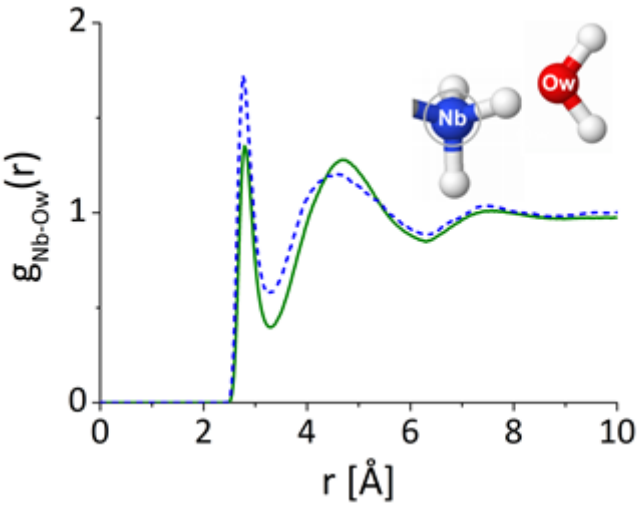


Figure 5.5: The pairwise RDF for the backbone nitrogen - water (Nb-Ow) interaction for sarcosine (green, solid) and Ngln (blue, dashed).

Sarcosine possesses the same increased coordination in the outer shells as for Ngln. The increased density of water in outer shells suggests that water molecules are forming bonds with other parts of the biomolecules in both cases. 26.3 water oxygen atoms can be found in the second coordination of sarcosine, which spans a distance range of 3.30 - 6.30 Å. Interestingly, the position of the second coordination shell for sarcosine has moved by 0.08 Å to 4.70 Å, compared to 4.62 Å for Ngln. This may be linked either to the varying distance of other functional groups in sarcosine that may coordinate water molecules or the differing influence of sarcosine on bulk water structure.

Table 5.2: First coordination shell numbers for backbone - water interactions for sarcosine (Sar) and Ngln.

Atom Label	Molecule	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Nb-Ow	Sar	1	3.30	2.79	1.7	0.8
	Ngln	1	3.29	2.77	2.1	0.7
Ca-Ow	Sar	1	4.23	3.53	6.4	1.6
	Ngln	1	4.35	3.48	8.5	1.5
Cb-Ow	Sar	1	4.47	3.55	8.9	1.8
	Ngln	1	4.32	3.51	8.3	1.3
Ob-Ow	Sar	1	3.10	2.64	2.2	0.8
	Ngln	1	3.21	2.64	2.7	0.7

Table 5.3: Second coordination shell numbers for backbone nitrogen - water interactions for sarcosine (Sar) and Ngln.

Atom Label	Molecule	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Nb-Ow	Sar	3.30	6.30	4.70	26.3	2.9
	Ngln	3.29	6.30	4.62	25.7	2.4

5.4.2 Hydration of the α-Carbon

It was determined in Chapter 4 that the difference between the glutamine amino and imino acid was linked to the structure altering the charge distribution over the molecule. By converting the tertiary ammonium to a secondary ammonium and making the α-carbon achiral, the charge redistributed in a way such as to encourage weak hydrogen bonds to form between water and, in particular, the adjacent methylene groups. In

sarcosine, the backbone nitrogen is attached to the α-carbon methylene still, but also the side chain methyl instead of a second methylene group. This subsection will focus only in hydration of the α-carbon, whilst the hydration of the side chain will be discussed in Section 5.4.4.

Figure 5.6 shows the RDF for Ca-Ow for both sarcosine and Ngln. For both, a prominent coordination shell is present, along with a succession of outer coordination shells, correlating to other functional groups with each biomolecule forming hydrogen bonds with water. 6.4 water oxygen atoms can be found in the first coordination shell for sarcosine (Table 5.2), notably lower than for Ngln that coordinates 8.5 Ow atoms through this group. Whilst the side chain for sarcosine is shorter, this provides fewer surrounding water molecules that could have overlapping first coordination shells. It is likely that the coordination number for Ngln is reflective of the numerous functional groups it possesses that form bonds with water molecules.

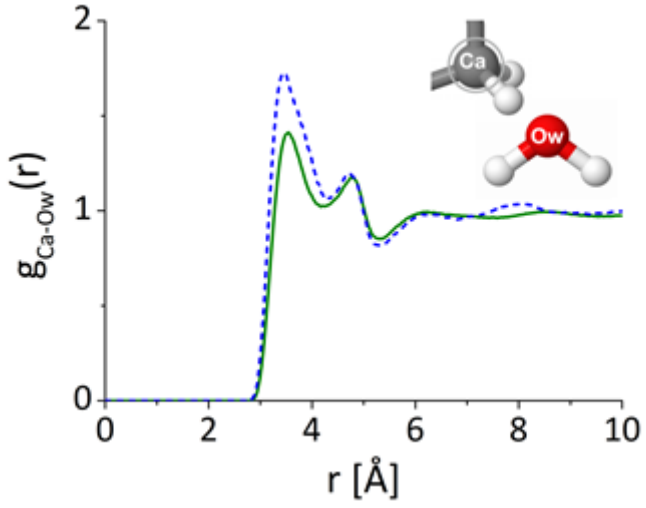


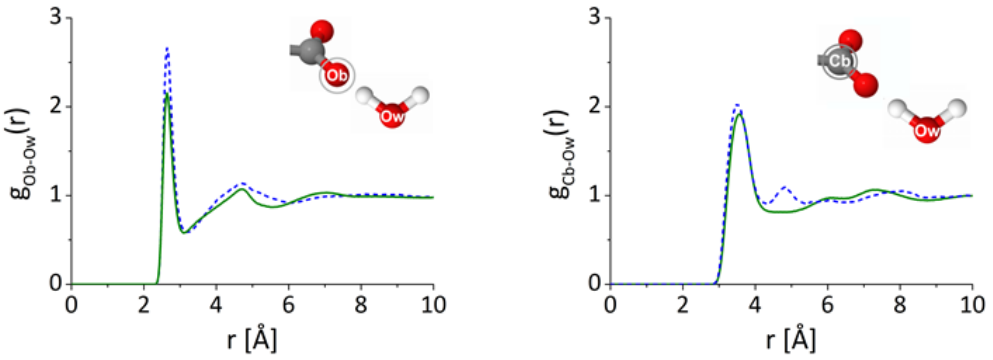
Figure 5.6: The pairwise RDF for the α-carbon - water (Ca-Ow) interaction for sarcosine (green, solid) and Ngln (blue, dashed).

5.4.3 Hydration of the Backbone Carboxylate

Common to all structures in this thesis is the presence of the backbone carboxylate group. In Section 4.4.3, it was determined that amino and imino acid versions of glu-

tamine coordinated water molecules similarly, with notable changes being observed for outer shells due the differing hydration shell over each molecule.

Figure 5.7 shows the RDFs for the Cb-Ow and Ob-Ow interactions for both sarcosine and Ngln. In Figure 5.7(a), the heights for the first and second coordinations shells indicate that the Ngln backbone oxygen atoms coordinate more water molecules than for sarcosine. Coordination numbers for the first shell of sarcosine (Table 5.2) indicate though that the difference is not significant, with an Ob atom interacting with 2.2 Ow atoms, as compared with 2.7 for Ngln. When correlated to an RDF for Cb-Ow (Figure 5.7(b)), the hydration in the first shell appears similar, with only a small variance in peak position, being 3.55 Å for sarcosine and 3.51 Å for Ngln.



(a) (b)

Figure 5.7: The pairwise RDFs for the (a) backbone oxygen-water interaction (Ob-Ow) and the (b) backbone carbon-water (Cb-Ow) interaction for sarcosine (green, solid) and Ngln (blue, dashed).

5.4.4 Hydration of the Side Chain

As described in Section 4.4.2, a change in electrostatics caused a redistribution of partial charge over Ngln. This in turn caused water molecules to form weak hydrogen bonds with methylene groups of the side chain, particularly the Cn methylene group. The corresponding group for sarcosine is the methyl side chain.

Figure 5.8 shows the RDF for the interaction of side chain carbon, Cs with water

oxygen atoms. The peak at 3.51 Å signifies that the side chain interacts with water molecules. The coordination number for this peak is 13.6 (Table 5.4). Whilst it is possible some of this could be attributed to water molecule coordination from other groups, the magnitude suggests that this group is also able to coordinate water molecules also. This capability comes from the presence of greater partial charge now being found on this methyl group.

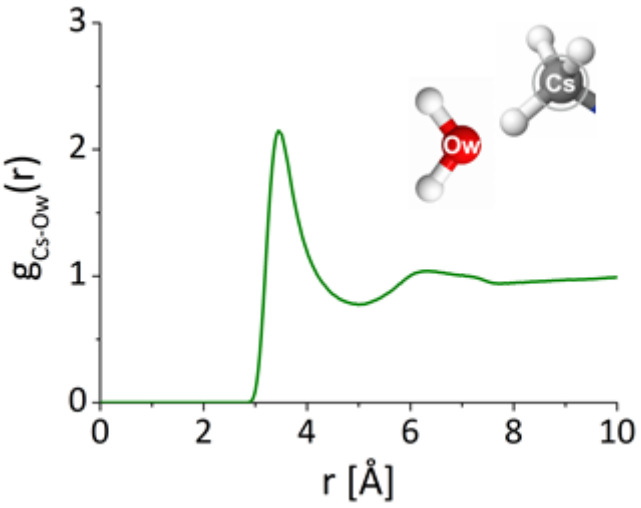
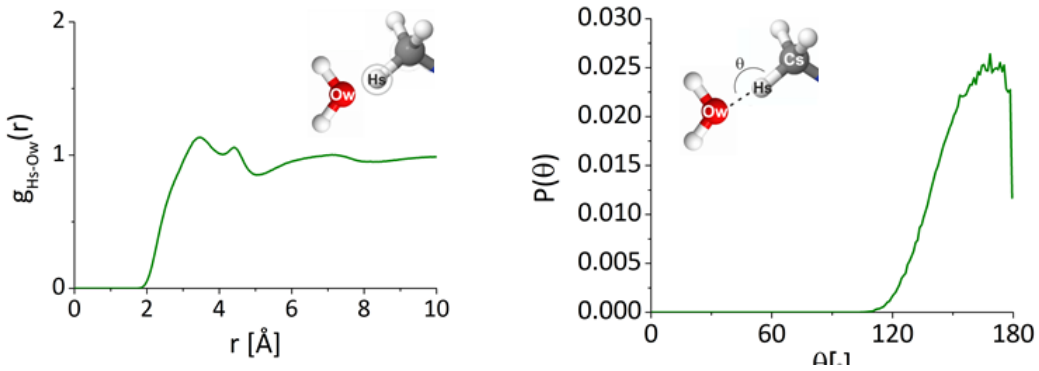


Figure 5.8: The pairwise RDF for the side chain carbon - water (Cs-Ow) interaction for sarcosine.

To determine if water molecules directly interact with the methyl group, the TRIANGLES routine (Section 2.2.3.4) has been used to determine the Cs-Hs· · ·Ow angle. The Hs-Ow RDF (Figure 5.9(a)) shows no defined peak for direct interactions, likely due to the presence of water molecules from other functional groups. Therefore, 2.4 Åwas used as the maximum distance between Hs and Ow, to determine the angle of the most direct interaction. The distribution (Figure 5.9(b))shows that direct interactions can occur between Hs and Ow predominantly between angles of 165–175°, indicating that the methyl group can coordinate water molecules.



(a) (b)

Figure 5.9: (a) The pairwise RDF for the side chain hydrogen - water (Hs-Ow), and (b) the bond angle, θ, distribution for Cs-Hs$\cdots$Ow, for sarcosine.

This result is comparable to that seen for Ngln, where Cn was able to coordinate water molecules. In reference to the Ngln side chain, this group was not strongly impacted by the changes closer to the backbone, with the side chain amide group showing similar levels of hydration for amino and imino acid. What the results for sarcosine suggest is that the electrostatic difference fully impacts the hydration of the side chain, likely due to the length of this functional group. Whilst the glutamine side chain consists of a chain of two methylenes and the amide group, sarcosine constitutes merely a short methyl group. Therefore, this indicates that bulky side chains are more likely to be hydrated and interact more similarly to its amino acid equivalent than those which have shorter side chains. In shorter side chains, a greater percentage of the total side chain takes the impact of the electrostatic change and does not have bulk functional groups to shield this behaviour. Even in chains, this behaviour is also unshielded, with possible increased partial charge at the exposed extremities of the side chain. This could account for the increased hydrophilicity seen in the Tang and Deber experiment. To properly test this, it would be useful to do a comparison of sarcosine with L-alanine, to determine quantitatively how hydration of the methyl side chain differs between them. This is discussed further in Section 6.3.

Table 5.4: First coordination shell numbers for side chain sarcosine - water interactions.

Atom Label	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Cs-Ow	1	4.97	3.45	13.3	1.8

5.5 The Association of Sarcosine

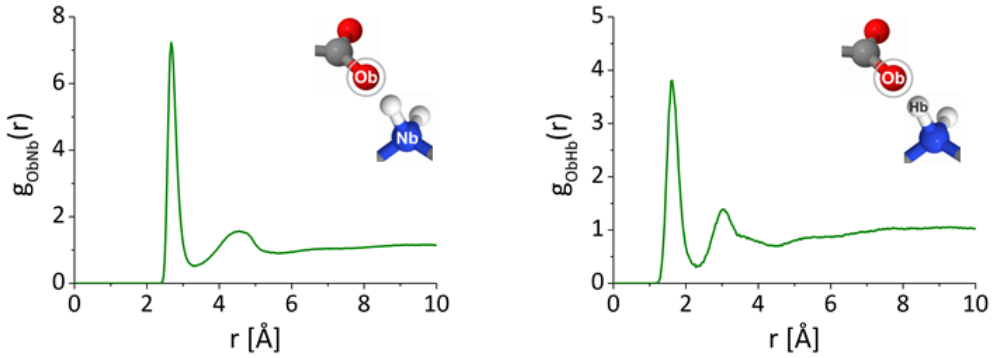
This section covers the interactions by which sarcosine molecules may associate through in solution. As ascertained in Chapter 3, association between amino acids may be driven not by hydrophobic interactions, but by those between its hydrophilic groups. This challenges the view of hydrophobicity being the driving force association, as indicated by researchers such as Kauzmann (22). For sarcosine, there are fewer functional groups, meaning that a comparison of interactions through hydrophilic and hydrophobic groups can be completed. The terminal ammonium and carboxylate stand as hydrophilic groups. The hydrocarbon methyl in sarcosine would usually represent a hydrophobic group, though work in the previous section highlights its increased hydrophilicity compared to its amino acid counterpart and other residues. In tackling this functional group, it will be possible to ascertain the impact of the backbone change and electrostatic redistribution on important hydrophobic interactions.

5.5.1 Association through Terminal Groups

To study the interactions between terminal groups, it is possible to first look at the RDFs for the Ob atom of one sarcosine interacting with the Nb and Hb atoms of another. Unlike glutamine-based molecules, there are no other oxygen and nitrogen atoms to consider.

For dilute systems studied in this thesis, the interaction between oxygen and amide or ammonium hydrogens was utilised (see Chapter 3). In studies such as that by Kumar, the Ob-Nb interaction was studied. Figure 5.10 presents the Ob-Nb and Ob-Hb

interaction for this system. The first peak in each signifies a direct hydrogen bond being made between the oxygen and hydrogen atoms of two molecules. The second peak is indicative of the the second hydrogen on the ammonium being in close proximity to the hydrogen bonding interaction taking place. The peak distance between Ob and Nb is 2.67 Å and is correlated to an Ob-Hb distance of 1.68 Å. This distance is shorter than the hydrogen bonds found in pure water or between water and sarcosine, indicating that these bonds are stronger and allow association to take place.



(a) (b)

Figure 5.10: The pairwise RDFs for the (a) backbone oxygen-backbone nitrogen (Ob-Nb) and the (b) backbone oxygen-backbone ammonium hydrogen (Ob-Hb) interactions for sarcosine.

A similar RDF was determined by Kumar *et al* (67). These RDFs are not identical, though this is the result of the present simulation being corrected to fit experimental data. The Kumar model for instance utilises SPC/E geometries for water. In this work, they determined that sarcosine, though forming interactions with other sarcosine molecules, exists in monomeric state.

In the present study, the CLUSTERS routine (Section 2.2.3.2) has been used to determine the size of clusters produced via the Ob-Hb interaction. Analysis of the first coordination shell for Ob-Hb suggests, that whilst monomers are the most dominant species in the solution, around 37.5% of the molecules will exist as dimers or more complex structures. However, the monomer is sill the dominant species in the solution. The results of the CLUSTERS routine is presented in Figure 5.11 and indicates that

clusters of up to 6 sarcosine molecules can be produced via this interaction. This is comparable to the results for L-proline, with a distinct hydrophilic and alkyl region, where 6-mers were produced through the zwitterionic groups. In these self-associative clusters, 54% are in the form of dimers, with approaching 9% attributable to 5- or 6-mers.

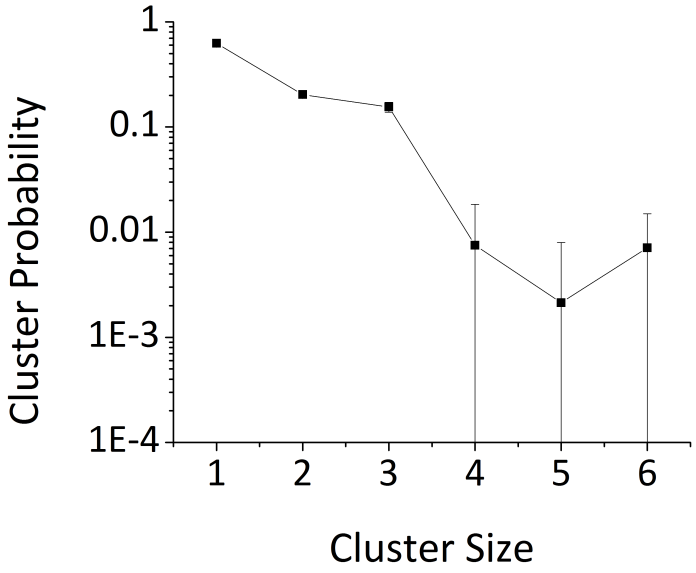


Figure 5.11: Probability of hydrophilic Sar clusters forming versus the size of the clusters observed, at 25 °C. The probability scale is logarithmic.

In Figure 5.10(b), the first coordination shell was represented by a prominent peak. However, a coordination number for this group equates to only 0.2 Hb molecules being found around a given Ob atom (Table 5.5). This suggests that the clusters produced are not stable, instead being transitory. The calculation of Kumar and co-workers was able to successfully determine that monomers are the most stable species of sarcosine to exist in solution. The present study has now extended this to show that it is possible for sarcosine to associate and form instantaneous clusters of size 2-6.

Table 5.5: First coordination shell numbers for backbone carboxylate oxygen - backbone ammonium hydrogen interactions for sarcosine.

Atom Label	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Ob-Hb	1	2.40	1.68	0.2	0.4

5.5.2 Association through the Methyl Side Chain

In Section 5.4.4, it was ascertained that the methyl side chain was surrounded by a shell of water molecules. Due to an increase in hydration around this group, it is possible that the ability of association to occur through the methyl groups could be affected. The Cs-Cs RDF is presented in Figure 5.12. From approximately 4.5 Å and lower, a shoulder is present that correlates to short-range interactions between the methyl groups. Whilst this suggests that association can exist through the side chain, it has been affected by the electronic state of the molecule. This further supports the work by Tang and Deber of sarcosine being hydrophilic, as this molecule preferentially is interacting with water over hydrophobic bimolecular interactions.

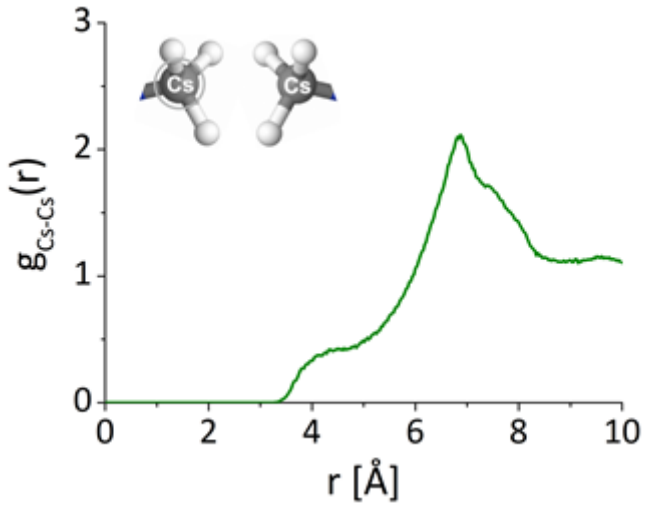


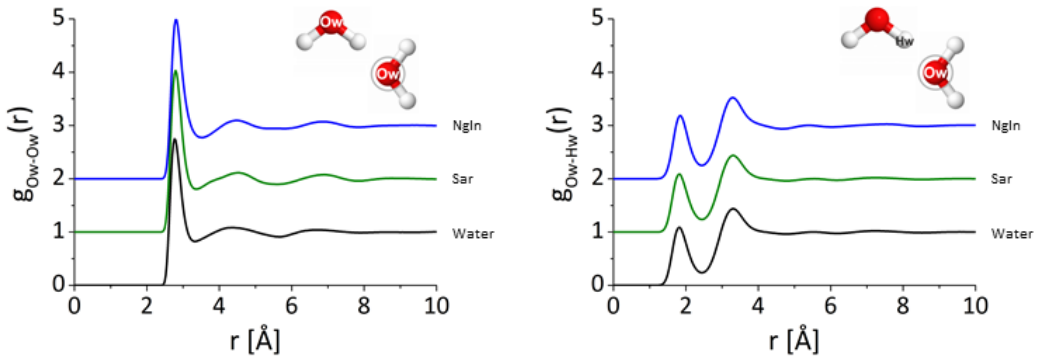
Figure 5.12: The pairwise RDF for the methyl carbon - methyl carbon (Cs-Cs) interaction for sarcosine.

5.6 The Impact of Imino Acids on Bulk Water

As discussed in depth in in Section 5.1, sarcosine is a recognised as an osmoprotectant. How it is speculated to function is by enhancing the structure of water. As other imino acids like L-proline are also known to have this capability, it is interesting to investigate the possibility of Ngln exhibiting osmolytic characteristics. This section will discuss the effect of both sarcosine and Ngln on the structure of water, looking at RDFs in Section 5.6.1 and tetrahedrality in Section 5.6.2

5.6.1 The Structure of Water in Imino Acids Solutions

The RDFs for the Ow-Ow and Ow-Hw interactions are presented in Figure 5.13. In both figures, the functions for sarcosine are now being compared with those for pure water at 25 °C. The main conclusion from these RDFs is that the structure of water is not significantly altered by the presence of sarcosine. The hydrogen bond length, represented by the first coordination shell of Ow-Hw remains the same at 1.83 Å, whilst the Ow-Ow distance increases only by 0.3 Å to 2.79 Å. The coordination numbers (Table 5.6) are found to stay near consistent for these shells, with any deviance attributable to the increased density of sarcosine in the solution influencing the ability of water to hydrogen bond within the bulk water network.



(a) (b)

Figure 5.13: The pairwise RDFs for the (a) water oxygen - oxygen interaction (Ow-Ow) and the (b) water oxygen - hydrogen (Ow-Hw) interaction for pure water (black), sarcosine (green) and Ngln (blue). The distributions have been shifted by 1 for clarity.

For Ngln, the Ow-Hw bond distance is comparable to that for pure water and sarcosine. The coordination number for this interaction, at 1.7, is also similar to the other two systems. Interestingly, coordination number has increased to 4.8. This is due to an increase in the bond distance encapsulated within the first coordination shell, where the r_{max} is 3.54 Å, compared to 3.33 Å for pure water and 3.36 Å for sarcosine. Furthermore, the likelihood of a bond forming, indicated by the peak height has increased, like for sarcosine. As the Ow-Hw distance is consistent, this would suggest that Ngln altering the structure of the water network. This may not only be due to Ngln being a larger molecule that would inflict a more notable density, but also there are a number of hydrophilic groups capable of bonding with water, which could in turn disturb the bulk network.

Table 5.6: Coordination numbers for water - water interactions for pure water, sarcosine (Sar) and Ngln at 25 °C.

Atom Label	Molecule	r_{min} (Å)	r_{max} (Å)	Peak Position (Å)	Coordination Number	Standard Deviation
Ow-Ow	Water	1	3.33	2.76	4.4	0.9
	Sar	1	3.36	2.79	4.3	1.1
	Ngln	1	3.54	2.82	4.8	1.1
Ow-Hw	Water	1	2.43	1.83	1.8	0.6
	Sar	1	2.46	1.83	1.8	0.7
	Ngln	1	2.46	1.83	1.7	0.7

5.6.2 The Tetrahedrality of Water in Imino Acids Solutions

In the work by Kumar and co-workers, one of the ways sarcosine was considered to act as an osmoprotectant was by enhancing the tetrahedral network of water molecules. It has been possible to test this for the present study by analysing the Ow$\cdots$Ow$\cdots$Ow triplet angle, using the TRIANGLES routine (described in Section 2.2.3.4).

A tetrahedral angle is defined as:

$$cos(\overline{ABC}) = -\frac{1}{3} \tag{5.1}$$

This equates to the angle for $\overline{ABC}$ being 109.5°. The closer the structure of bulk water conforms to this angle, the more tetrahedral the network. The Ow$\cdots$Ow$\cdots$Ow angle distributions for pure water, sarcosine and Ngln in Figure 5.14, with the ideal tetrahedral angle shown by a grey line. For pure water, there are peaks at 55.5 ° and 98.5 °. The 98.5 ° corresponds to bulk water, whilst the 55.5 ° peak is correlated to interstitial water molecules that do not conform to bulk water structure. In aqueous sarcosine, there is an increase in the the probability of interstitial water molecules, though the angle remains comparable to that for water. The change in the interstitial water peak height is potentially linked to water molecules forming a hydration layer around sarcosine, re-

sulting in more water molecules not conforming to bulk water structure. Interestingly, the bulk water peak has now moved outwards to 101.5 °, and the presence of sarcosine has encouraged bulk water to take on a more tetrahedral-like structure, supporting the findings of Kumar. In studies of Huang and co-workers (64) on the cryoprotective properties of peptoid chains, those containing sarcosine were found to lower the freezing point of water, though not to the same extremity as molecules containing hydroxyl groups. By making the structure of water more tetrahedral, it makes it more ice-like. Thus, whilst sarcosine may exist in monomer form and break up the continuity of bulk water, its effect on the tetrahedrality could explain why sarcosine was less effective as a cryoprotectant, with polyols such as glycerol typically disrupting the tetrahedrality of bulk water (188).

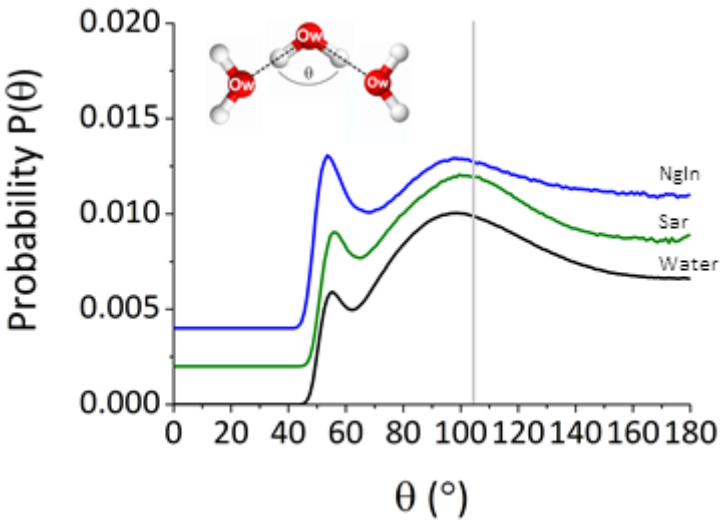


Figure 5.14: The Ow$\cdots$Ow$\cdots$Ow angle, θ, distribution for pure water (black), sarcosine (green) and Ngln (blue). The distributions have been shifted by 0.002 for clarity. The grey line shows the ideal tetrahedral angle, 109.5 °.

Section 5.6.1 hypothesised that there was a notable deviance in first coordination shell of the Ow-Ow interaction for Ngln due to more water molecules not conforming to the bulk water network. This is now supported by a notable increased interstitial water peak at 53.5 °. This will be due to the increased number of Ngln - water hydrogen bonds being formed. Consequently, the proportion of water molecules in the bulk structure has decreased, indicated by a 0.0011 reduction in peak height compared to water. Interest-

ingly, though fewer water molecules form the bulk, the most probable Ow$\cdots$Ow$\cdots$Ow angle for these molecules is 98.5 °, comparable to that for pure water. Though fewer water molecules are part of the bulk structure, the hydrogen bonding network has not been affected by the increase of interstitial waters. These studies have only been completed for one concentration and this result could be concentration-dependent as has been observed for other molecules (188), but nevertheless result potentially highlights that structures containing a common secondary ammonium structure tend towards not reducing the Ow$\cdots$Ow$\cdots$Ow bond angle. The impact of the imino acid though is dependent on the side chain functional group, with those containing methyl/alkyl groups better at supporting bulk water structure.

5.7 Conclusions

This chapter analysed the hydration and association of sarcosine in aqueous solution. As part of this study, comparison was made to previous experimental and computational analyses of this molecule for the conformation, hydration and bulk species of sarcosine in the solution, which was found to be supportive of previous investigations (67; 82). It also presented new information on coordination numbers for hydration and the size of sarcosine clusters that could exist in the solution, as well as discussing aspects such as hydrophilic versus hydrophobic association. In Section 5.2, it was determined that the difference between flexible and rigid versions of sarcosine gave similar fits to the neutron diffraction data, with a slight improvement for the flexible version with no dihedral angles fixed as opposed to a simulation with dihedral angles fixed. As part of this study, comparisons were made to Ngln and how the differing side chain had an impact on the hydration of certain backbone groups, as presented in Section 5.4. In this section, it was also determined that the methyl side chain was coordinating water molecules as a result of the charge distribution over the molecule. In Section 5.5.1, it was determined that hydrophilic association through the backbone allows for transitory clusters of up to size 6 to be produced, whilst the ability of the side chain to form weak hydrogen bonds reduces the likelihood of observing side chain - side chain association in aqueous solutions. Finally, the impact of imino acids on water structure was also

introduced in Section 5.6, where imino acids are capable of maintaining a tetrahedral tetrahedral network, even if the side chain affects the number of molecules that do not conform to bulk structure. This is the first structural comparison of two imino acids to be completed and the similarities and differences observed in their interactions could have implications on how they are utilised in more complex peptoid structures.

Chapter 6

Conclusions and Future Work

Chapters 3 - 5 presented the results for this thesis. This final chapter summarises the main conclusions and impact of these results on peptoid, biomolecule and EPSR-based research. The results highlighted in Section 6.1 are on single amino and imino acid molecules and Section 6.2 discusses how these could relate to the understanding of more complex peptoid structures for a variety of applications. Finally, Section 6.3 will detail future directions that can be taken for the work presented in this thesis.

6.1 Concluding Remarks

This thesis has presented a bottom-up-approach to ascertaining the structural properties of the class of peptidomimetics known as peptoids. Peptoids have been noted for their tunable properties and are being exploited for many biological, pharmaceutical and nanotechnology applications. Previous literature has focused on the properties of peptoid chains and how they can fold to form complex structures, given the functional groups they contain (36; 37). Any structural information on peptoids has been achieved only through crystallographic studies or computational simulation (34; 35). What has been lacking in the literature is direct, structural information on the inter- and intramolecular interactions in the liquid phase. Very little research has also been completed on the properties of the individual residues that make up these chains, single imino acids, which in themselves are exploited as osmoprotectants (61–64). As properties such as

increased hydrophilicity are observed not only in chains but also in individual imino acids, it is thought that the properties seen in peptoids are the direct result of changes to the building blocks themselves.

This thesis has completed structural investigations on imino acid building blocks. As part of this study, a comprehensive study has been completed on the molecule N-(2-carboxamidoethyl)glycine or Ngln, the imino acid version of L-glutamine. This molecule is synthetically produced and this thesis describes the first structural study to be completed on this building block, and represents the first study of any synthetic peptoid monomer. The intermolecular interactions that this molecule makes with varying concentration and temperature, as well as its impact on bulk water, have all been investigated for this project. Previously, Ngln has been subject to limited studies and only within the context of peptoid sequences, where it is thought to be a useful residue for neuropeptide mimics (171), but incapable of forming the bonds required to assemble more complex structures such as superhelices (44). This is in contrast to the amino residue version, of which chains are thought to be able to collapse and form stable aggregates that are associated with disease (154; 155), as well as having a role in forming the secondary structures in proteins (152). This thesis presented information on the hydration and association of three molecules in aqueous solution: Ngln, the amino acid equivalent L-glutamine so as to complete a structural comparison between the two and determine reasons for the improved properties that the imino acid possesses, and sarcosine, a natural-occurring imino acid that allowed for the study of the effect of changing the side chain on imino interactions.

In Chapter 3, results were presented for L-glutamine. Fitting its hydrophilic nature, it was determined that L-glutamine readily hydrogen bonds with water. It is able to form bonds with water as a result of the zwitterionic groups located on the backbone and the polar amide group on the side chain. This amino acid was found to associate with other glutamine molecules via all four hydrophilic interaction types, though backbone-backbone interactions were the most prevalent. From the first initial study completed (1), developments have been introduced to obtain a better representation of this system, though the conclusions of the initial study remain unchanged and quantitative information generally remained consistent within error. Due to concentration limitations, it was

concluded that detailed information on the conformation is not attainable. Furthermore, functions with little or any weighting in the data such as association could be subject to inadequate sampling. As a way of determining the extent of clustering in L-glutamine solutions, a larger simulation box was created to elucidate the prevalence of clustering in the solutions. Association of L-glutamine molecules through hydrophilic groups was found to persist at all temperatures studied, 24 °C, 37 °C and 60 °C. As many as 15–20% of the molecules could be involved in a cluster at each temperature, though both neutron diffraction and SANS suggest that the clusters produced are transitory, being formed and breaking apart instantaneously. Whilst hydrophilic clustering was present in the solution, interactions via hydrophobic entities were negligible, existing only *in situ* with a hydrophilic intermolecular interaction. The results show that the dominant interactions to exist between glutamines are hydrophilic, be them dipoles or possessing partial charges, and it is possible to speculate that hydrophobic interactions do not play a significant role until a biomolecule becomes no longer miscible in solution.

The topic of hydrophilicity has been a recurring topic in this thesis and Chapter 4 presented the properties of Ngln, unique compared to other imino acid building blocks studied in the literature for its hydrophilic side chain. This chapter utilised the findings of Chapter 3 to complete a structural comparison of amino acid and imino acid building blocks to elucidate the source of properties such as its increased solubility. With the change from a primary to secondary ammonium group, the backbone nitrogen for an imino acid coordinates one less water molecule than an amino acid. However, the results of this study have suggested that van der Waals interactions are present between water and the methylene groups adjacent to the ammonium. The backbone carboxylate and side chain functional groups only experience subtle changes in hydration, but with any differences favouring the imino acid being more readily hydrated. In line with other literature comparing amino and imino acid (82), the change in backbone structure alters the electronic and charge distribution over the molecule, resulting in alkyl groups losing their hydrophobic character. This means that more of the imino acid is surrounded by water molecules and this could cause the greater solubility and hydrophilicity seen for imino acids.

The solubility of Ngln is at least 10 times greater than that of its amino acid version.

Thus it serves as a model system to understand some general properties about glutamine-based molecules and their association, which could not be probed for L-glutamine. Upon increasing the concentration to 300 mg ml^{-1}, the hydrogen bonds between the imino acid and water remain similar to the 30 mg ml^{-1} sample. At this higher concentration and with increasing temperature, the imino acid is able to interact via the four possible polar combinations of hydrogen bond and allow 6-mers to form in solution at 25 °C. The most prevalent of these were those interactions involving the backbone carboxylate oxygen, which is possible due to the partial charge on this functional group. As part of this study, the prevalance of the four hydrophilic interactions studied was observed with temperature. For all interactions at 60 °C, the level of hydrogen bonding was seen to decline. Whilst it is not possible to obtain thermodynamic parameters from the data, it can be speculated that these hydrogen bonds become destabilised with increasing temperature due to enough thermal energy being present to weaken them. Suprisingly, whilst Ngln-Ngln interactions involving the backbone zwitterionic groups showed a trend of decreasing with increasing temperature, interactions involving the side chain amide group were found to become more prevalent initially at 37 °C. The side chain groups are common to both imino acid building blocks and peptoid chains. As these interactions exist and are potentially even enhanced at certain temperatures, this highlights their potential importance in applications that require physiological conditions, as well as potentially having a role in disease-causative proteins.

Ngln was compared to the naturally-occurring alanine imino acid, sarcosine, which has been subject to a select number of investigations on its structure and interactions (67; 69; 82; 241; 242; 244). Whilst sarcosine possesses the same backbone structure as Ngln, it is evident that the differing side chain has an impact on the overall hydration of the molecule. Furthermore, the side chain has been shown to interact with water molecules, indicating that the change in backbone structure has affected the charge distribution over the molecule and caused it to lose the hydrophobic character associated with methyl groups. This was supported by evidence from the relevant RDF that methyl groups of separate sarcosine molecules rarely interacted in solution. Though a comparison has not been completed with L-alanine for this thesis, it can be speculated that the number of water molecules surrounding this group could attributed to the

increased hydrophilicity of sarcosine, seen in single monomer studies (82) and in chains containing sarcosine residues (47). Whilst the work generally supports that of previous studies (67; 82), this study gave insight into the number of molecules surrounding a given sarcosine molecule and the fact that it is possible to obtain instantaneous clusters in solution, even if the dominant species is the monomer state, as shown previously.

Imino acids like sarcosine and the amino-imino acid hybrid L-proline are all known osmolytes and are recognised as protein stabilisers, able to counteract the effects of denaturants (62; 63; 67; 128). This work concluded by determining the effect of sarcosine on water structure, as well as explored the possibility that Ngln may also possess osmolytic properties, given its backbone structure. It was observed that whilst sarcosine maintained similar levels of coordination to that of pure water, Ngln was increasing the number interstitial water molecules in the solution. It was also determined that both molecules either maintained or increased the Ow$\cdots$Ow$\cdots$Ow angle of water. Whilst this indicates that the common secondary ammonium backbone has a role in preserving the bulk water network, much dependence is on the side chain group present, making certain molecules more suitable for osmolytic applications than others.

6.2 Implications for more Complex Peptoid Structures

A key result ascertained from Section 6.1 was that imino acid differences concern not only the nitrogen, but also adjacent alkyl groups. In literature that speculates about interactions in these areas for peptoid chains (46; 47), no consideration has previously been taken into account regarding interactions with these groups. The results also suggest that interactions vary depending on the side chain present, with sarcosine and Ngln exhibiting differing hydration levels.

What it is important to appreciate is that Section 6.1 states the results for simple imino acids and provides only a fundamental understanding of the properties of complex peptoid structures. Thus, this section considers the ways to predict the behaviour in more complex systems based on the results presented in this thesis.

When investigating the association of sarcosine, it was clear that interaction through the zwitterion terminals dominated over those through the methyl group. The increase in hydration and the lack of methyl-methyl interactions could explain the drastically increased hydrophilicity observed in chromatography experiments for sarcosine (47). With association not driven by side chain interactions, it is possible that these interactions would be less dominant also in peptoids, causing these chains to take on more open structures. Studies such as that by Rosales *et al* (70) determined that peptoids containing hydrophobic moieties, though forming helices in *cis* form, were still flexible and could form a variety of structures in solution.

As discussed in Section 1.3.2, research by Murnen *et al* determined that sequences of Ngln could not be used to form superhelical structures (44). In Chapter 4, it was found that though interactions through all four hydrophilic combinations studied were present, interactions involving the side chain oxygen atom were minimal. Therefore, superhelices may not have been produced using Ngln residues as hydrogen bonding through side chain groups does not prevail. However, though still not dominant compared to backbone interactions, side chain hydrogen bonding interactions were shown to increase when the temperature was raised to 37 °C, and could mean that this residue is important for systems in the physiological temperature range. Therefore, it is possible that more complex structures could be formed using this residue if the temperature was varied. This is a topic worthy of further investigation, particularly for hydrogen bonding that is not thought to have a dominant role in the self-assembly of peptoid structures (34).

The evident next step is to try and appreciate how the interactions seen in this thesis could vary in a peptoid chain setting. A method for doing this is to make predictions based on the reference potentials. EPSR, though making its own alterations using the empirical potential added to the simulation, uses the reference potential as a guide and the assumption made is that the reference potential is a reasonable estimation of the parameters in the system.

Figure 6.1 presents the OPLS potentials for model peptoids. The figure looks at the numerical charge distribution in sarcosine, imino acids with complex chains, L-proline and tripeptoids. For the tripeptoid, they maintain a repeating pattern of having higher

charge density on the methylene groups surrounding the ammonium group. The α-carbon has a charge of +0.19, balancing out effects from the backbone carboxylate, whilst the β-carbon has a charge of +0.29. If this were an alkyl group attached to an α-carbon, the charge would be 0. For a tripeptoid, this charge pattern is only seen for the N-terminal residue. For the other two residues in the tripeptoid, there is only a negative backbone nitrogen, with no available hydrogens to bond with and the charge on each adjacent methylene groups is only +0.07. Whilst it may be possible that van der Waals interactions could exist, particularly for sarcosine residues where the side chain with this charge is fully exposed, the extent to which these interactions occur may be decreased. Much of this will be dependent on how well the backbone nitrogen and adjacent alkyl groups are shielded by the side chain. This could mean some of the more extreme differences seen between amino and imino acid could be further reduced, making these chains much more favourable for applications such as mimicry.

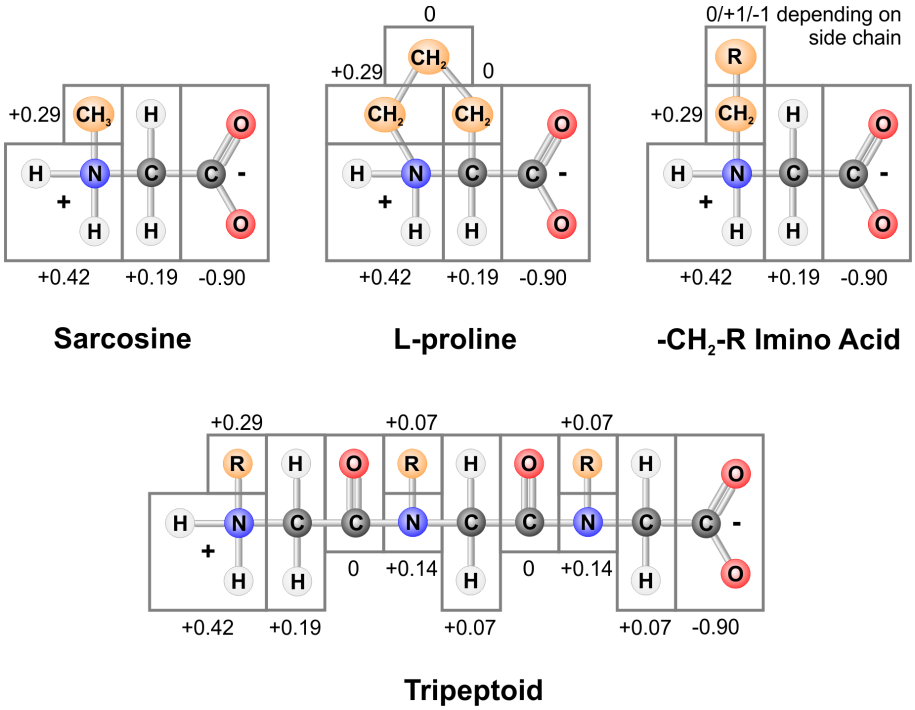


Figure 6.1: OPLS charges for zwitterionic imino acids, L-Proline and Tripeptoids, taken from (91; 92; 235). The **-CH_2-R Imino Acid** image is a generalised schematic representing a given imino acid with a longer side chain group. Even in a peptoid chain, charge is now present on the side chain, which will affect the alkyl groups adjacent to the backbone nitrogen atoms.

6.3 Future Work

The work detailed in this thesis addresses the structure and interactions of a novel peptidomimetic building block in solution. Following the bottom-up approach, this work can be used as a starting point for future investigations into the properties, interactions and assembly of more complex peptoid structures. This subsection lists some suggestions of investigations that could be carried out that would enhance the scientific importance of the work detailed in this thesis.

6.3.1 Further Incorporation of Auxiliary Routines and Data Analyses

In this account, structural data has been presented in the form of Radial Distribution Functions (RDFs) as well as exploiting the auxiliary routines listed in Table 2.3. As well as the analyses reported, there are ways these systems could be explored.

Equation 2.28 described how the partial structure factor, S(Q), was related to the RDF, g(r). An RDF gives real-space information on common bond distances between atoms, but it is also possible to analyse pairwise interactions in the system in terms of S(Q). By analysing the partial structure factors, it is possible to determine which interactions give rise to specific signatures in F(Q). S(Q) plots have been used readily in structures on liquid propanol (117); by comparing to F(Q), it was possible to see what interactions have a greater weighting in the data and at what Q values they appear at in the data. For example, it was determined that the main peak in F(Q) for propanol was due to intermolecular interactions of carbon with other atoms in the system.

In Chapter 3, the CLUSTERS routine (Section 2.2.3.2) was utilised to determine the impact of hydrogen-bonding on association. By setting the charges to zero, this switched off hydrogen-bonding, allowing for a comparison in the clustering to be made with the original EPSR simulation. It is possible to also completed such tests for the imino acid molecules studies, to determine the role of hydrogen bonding in their association. If the clustering decreased for polar groups, this would highlights the relevance of hydrogen bonding, whilst if it increased this would support the idea that these molecules preferentially interact with water. Furthermore, as sarcosine possesses a methyl side chain,

the CLUSTERS routine could be used to explore association through more hydrophobic groups.

6.3.2 Analysis using Newer Versions of EPSR

The work presented in this thesis uses version 18 of the EPSR software. During the latter stage of this project developments were made externally to improve the EPSR software. The changes implemented include making it possible to run the simulation in parallel on a computer. This means the program can use more processors, improving the running time and computational cost. The way that dihedral angles are defined has also been altered; as opposed to calculating a pseudo bond length to represent this parameter the angle between planes is calculated explicitly. In EPSR18, geometrical parameters were all constrained using the one relation for the intramolecular potential (Equation 2.37) and were controlled within the program by specifying a single energy parameter. Bond lengths, bond angles and dihedral angles are now separately controlled in newer versions of the software. It would be of interest to study the systems presented in this thesis with a newer version of EPSR to determine the impact of these improvements on the quantitative results and molecular conformations produced.

Due to time constraints, it has not been possible to thoroughly analyse all the systems presented in this thesis with newer versions of the software. In this thesis, simulations have been completed using version 18 of the software. However, in Section 3.5, results were presented for the association and clustering of L-glutamine with increasing temperature, trialling version 24 of the software. This simulation exploited the fact that the computational improvements allow larger simulations boxes to be produced. It should be noted though that running this one simulation, with only the CLUSTERS routine, the simulation took many weeks to run and to attain additional quantitative information would require many more weeks or months of computational time or further alterations to improve the running speed of the program.

Figure 6.2 shows example water-biomolecule and water-water RDFs for L-glutamine, comparing the two versions of the software. Interestingly, the newer version shows a more disordered picture for the system, though the key features are still observed for both

simulations. The smearing present could be related to the resolution being probed now in newer EPSR, but could simply represent disorder in the system. EPSR produces a plausible result that fits the data, though not necessarily the only result possible and these RDFs suggest that not only one outcome is possible for this system. The newer versions, post version 18, are beta versions that are still undergoing significant development. Nevertheless, it would be of benefit to simulate the systems in this thesis to determine the disordered variance in the solutions and identify necessary improvements needed for the EPSR software.

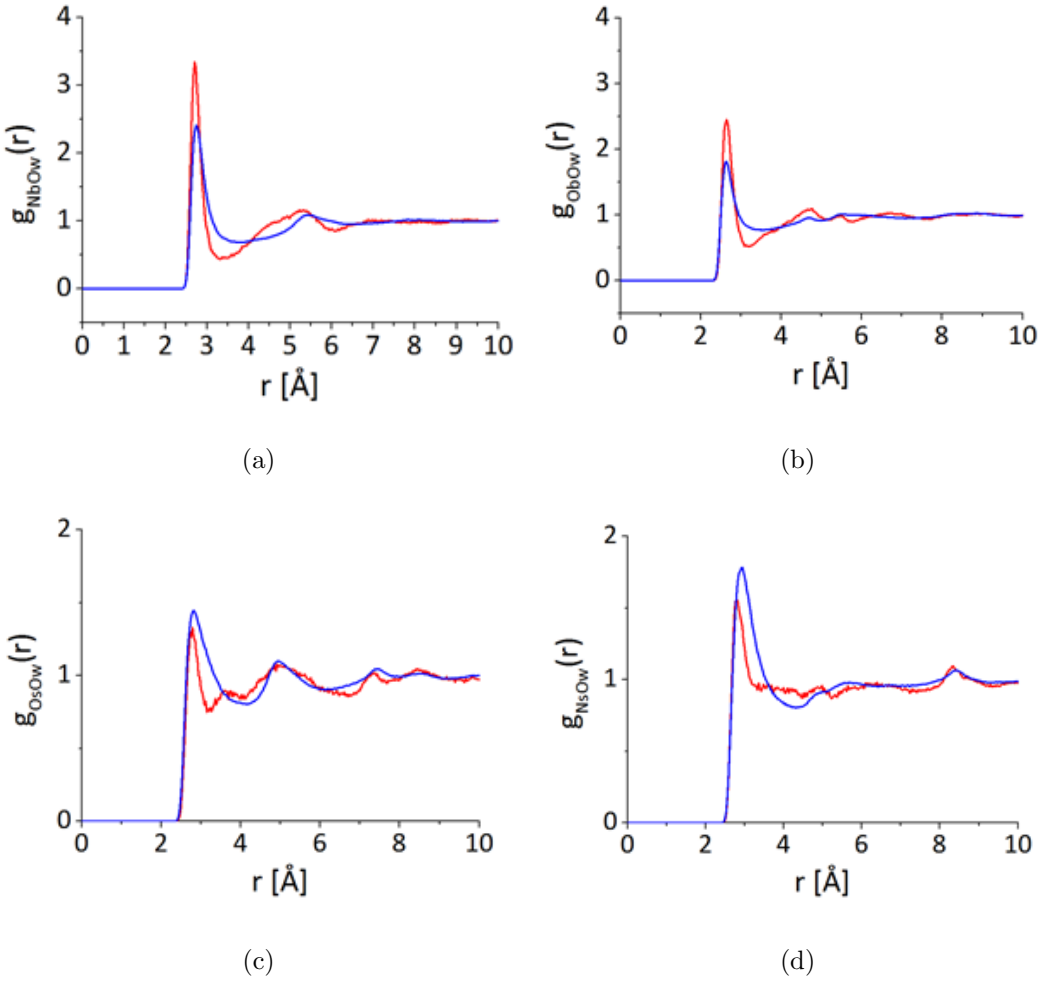


Figure 6.2: Sample pairwise RDFs for L-glutamine comparing the outputs from EPSR18 (red) and EPSR24 (blue). The functions shown are for the (a) backbone nitrogen - water (Nb-Ow), (b) backbone oxygen - water (Ob-Ow), (c) side chain nitrogen - water (Ns-Ow) and (d) side chain - water (Os-Ow) interactions.

6.3.3 Comparison with Amino Acid Equivalents and the Impact of Other Side Chains

As part of this thesis, a comparison was completed between Ngln with L-glutamine. By also studying sarcosine, it was possible to determine the impact of imino acid side chain and it was possible to speculate on the effect of side chain type and length on intermolecular interactions. To extend this study, it would be ideal to compare sarcosine with its amino acid counterpart L-alanine to determine quantitatively the extent of altering the backbone on side chain interactions.

This study has mainly focused on the hydrogen bonds made by these imino acid molecules when hydrated and when associating in solution. Whilst hydrogen bonding has been shown to be important in these systems, previous studies have indicated that interactions such as hydrophobic, electrostatic and n-π^* stacking are important for peptoid self-assembly (34–37). The current literature on the peptoid folding process was described in detail in Section 1.1.4.3. By looking at systems with different side chain groups, it will be possible to structurally examine the impact of these different interactions on peptoid assembly.

6.3.4 Understanding More Complex Peptoid Structures

The work completed in this study has looked at the building blocks of the peptoids, imino acids. Studying the building blocks has given structural insight into the source of the properties of imino acids, but it is a first step in understanding the properties of complex peptoid structures. Upon forming bonds, the nature of these building blocks is expected to be modified, as indicated in Section 6.2.

To understand how the chemistry changes upon converting from an imino acid into a more complex form, structural studies could be completed. With this method, ensuring appropriate solubility, it is possible to complete studies on small chains, so work on dipeptides or tripeptides would be of benefit. Neutron diffraction combined with EPSR modelling has been used to study small tripeptide sequences such as glycine-proline-glycinamide (139; 140) and glutathione, a sequence of glutamate-cysteine-glycine (141).

Alternatively, the amide group found on the backbone of peptoid chains could be introduced into a single imino acids by capping at least one of the zwitterionic terminals. Possible caps include flanking an acetyl group at the N-terminus or an amide group at the C-terminus. These caps are often also flanked with methyl groups. Capped amino acids have been studied previously via the methods exploited in this thesis (143; 145; 245–247), where the cap can even improve the solubility of the molecule. Indeed, the studies completed on peptoids, even on small tripeptides (64), use capped sequences, to inhibit the behaviour of the charged terminals and potentially observe the behaviour of backbone groups.

6.3.5 The Use of Supporting Methodologies to Understand Biomolecular Structure and Interactions

This thesis has provided quantitative insight into the hydration of natural and synthetic biomolecules in solution. Neutrons were used as they are sensitive to the location of hydrogen, making it a more sensitive probe to the hydrogen bonds present in the system. As highlighted in Chapter 3 and touched on throughout the thesis, there have been limitations also to what the method can provide. In particular for dilute systems, which are the lowest concentration studies to ever be completed using this combination of methods, the limitation comes from information on association and conformation not being present in the data, as well as restrictions with the modelling stage in defining certain constraints and it being computationally expensive to produce large enough boxes to sample rare events.

Within Chapter 3, SANS was introduced to determine if any stable clusters of L-glutamine were being formed in solution. An evident next step would be to introduce more techniques to support the results presented in this thesis. Whilst these methods could solve the issues seen for the most dilute solutions described in this thesis, they may also be utilised to provide information on the more concentrated biomolecular solutions. For the imino acid solutions, SANS could be exploited to determine whether the monomer state is the predominant form existing in solution and that there are no stable clusters being produced through association.

There are a number of structural methods that provide supporting information. This work provided a rationale for prioritising the use of neutrons over X-rays. Nevertheless, X–ray diffraction data could provide supplementary information on heavier elements in the system such as oxygen and carbon. Literature by Soper has noted the ability of X-rays in determining the distance of water oxygen atoms (248). EPSR is able to analyse both neutron and X–ray data simultaneously.

Other non-scattering methods can also be utilised to provide supplementary structural information on these biomolecules. NMR spectroscopy is sensitive to the location of non-exchangeable hydrogen atoms and the sample concentrations used can be as low as 10 μM, approximately 5% of the concentration used for the dilute systems presented in this thesis (83). If diffusion coefficients are calculated, it is possible to extrapolate the average mass of the species in solution and determine the extent of clustering in a given system. Two-dimensional Overhauser effect experiments are sensitive to the spatial arrangement of atoms within a molecule and can detect atomic pair-wise correlations. Literature exists that uses the results of one- and two-dimensional NMR experiments to obtain quantitative information on the hydration shell surrounding a molecule (84). Information on the hydration could allow for a comparison to be completed between neutron diffraction and NMR spectroscopy.

When considering differences in association with temperature, reference was made to thermodynamic parameters such as entropy and enthalpy. Methods such as ITC can be used to investigate the thermodynamics behind biomolecular associations, and used to obtain parameters such as the enthalpy of association (249) in the systems studied in this thesis.

Appendix A

Synthesis and Characterisation of the Glutamine Imino Acid

A.1 Synthesis

The glutamine imino acid can be produced via a simple two-step method. Stage one requires producing the intermediate, acrylonitrile, with the method first reported by McKinney et al (233). The second stage requires adding the side chain group to the intermediate, producing N-(2-carboxamidoethyl)glycine. The second stage is described in 1971 literature by Stewart (234). The two synthesis stages are shown in Figure A.1.

The intermediate is synthesised by reacting acrylonitrile, glycine and potassium hydroxide in a 1:1:1 molar ratio. Aqueous glycine and aqueous potassium hydroxide are stirred together whilst the flask is held within a container of ice. This cool environment is necessary as potassium hydroxide is known to heat upon reaction. Upon sufficient miscibility, acrylonitrile is added gradually and left to mix for 4 hours. Water is evaporated from the mixture to a volume where the concentration equates to 200 ml per mole of solution. Ethanol is added to encourage crystallisation and the solution refrigerated at 4 °C to allow N-(2-cyanoethyl)glycine crystals to form. The crystals are filtered and washed with copious ethanol and dried in a dessicator. 1 ml of hydrogen bromide in acetic acid is then added to every 1 g of N-(2-cyanoethyl)glycine. This solution is

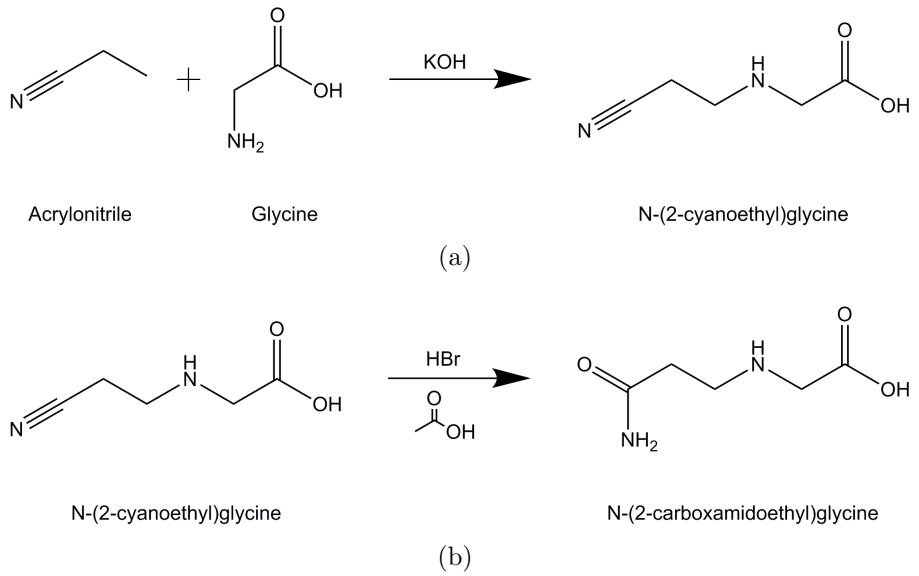


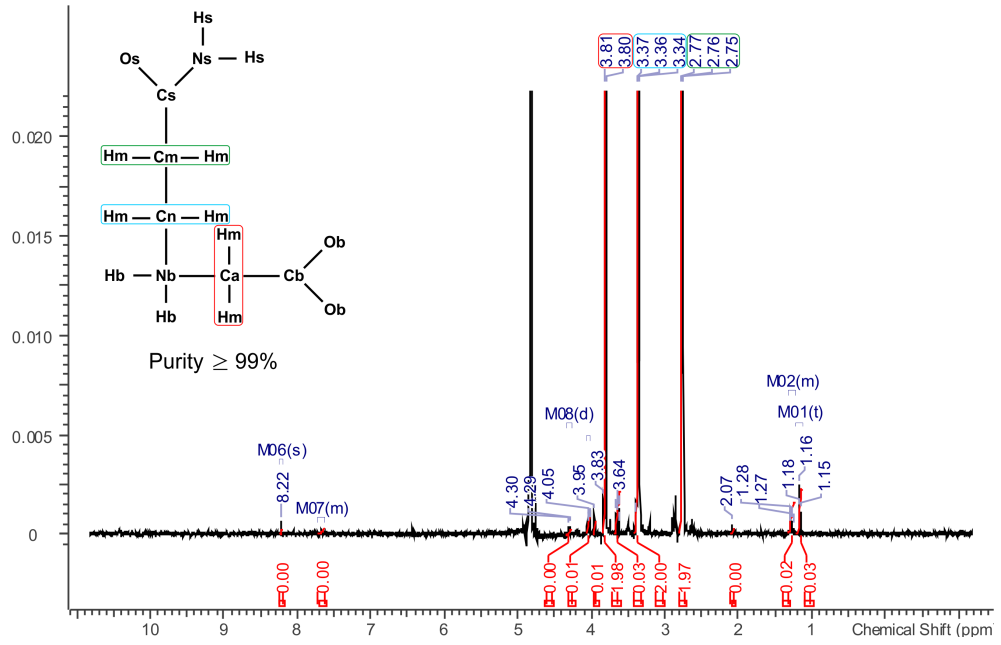
Figure A.1: The two step method to produce the glutamine-imino acid, N-(2-carboxamidoethyl)glycine. (a) The intermediate, N-(2-cyanoethyl)glycine, made from reacting acrylonitrile (C_3H_3N) and glycine ($C_2H_5NO_2$) with potassium hydroxide (KOH). (b) The N-(2-cyanoethyl)glycine is then reacted with hydrogen bromide (HBr) in acetic acid ($C_2H_4O_2$) to produce the imino acid.

stirred and then filtered to remove any unreacted material. Ether is added to form the product. The product is cleaned in ethanol and dissolved in ethanol and water in preparation for being refrigerated overnight to allow N-(2-carboxamidoethyl)glycine crystals to form. The crystals are cleaned in ethanol. The crystallisation process is repeated until the crystals are of $\geq 98\%$ purity. The purity can be assessed by obtaining an NMR spectrum of the sample.

A.2 NMR Characterisation

Spectra for the glutamine imino acid were taken using a Bruker 500 MHz NMR spectrometer.

Peak height ratios are labelled in red. With sample peaks having a height value of 2 and impurity peaks having a value of no greater than 0.03, this satisfies the pruity requirement of $\geq 98\%$.



Appendix B

Reference Potentials and EPSR Parameters

B.1 L-glutamine Reference Potentials and Parameters

For atom type Ob, Hm, Hb and Hs, these labels have been allocated to more than one atom. Whilst all atoms with a given label will have the same potentials, the bond lengths, angles and dihedrals are specified according the specific atoms. For geometric parameters, atoms of a given type also have numbers, e.g. Ob1 and Ob2.

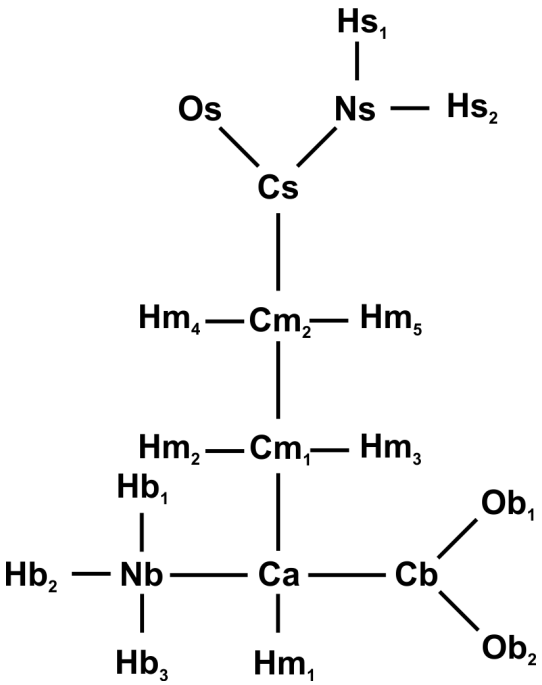


Table B.1: L-glutamine potentials

Atom Label	ϵ (kJ mol^{-1})	σ (Å)	Mass (a.m.u.)	q(e)	Atom Type
Cb	0.43932	3.750	12	0.700	Carbon
Ca	0.33472	3.800	12	0.210	Carbon
Cm	0.49371	3.905	12	0.000	Carbon
Cs	0.43932	3.750	12	0.500	Carbon
Nb	0.71128	3.250	14	-0.300	Nitrogen
Ns	0.71128	3.250	14	-0.850	Nitrogen
Ob	0.87864	2.960	16	-0.800	Oxygen
Os	0.87864	2.960	16	-0.500	Oxygen
Hb	0.000	0.000	1	0.330	Hydrogen
Hm	0.000	0.000	1	0.000	Hydrogen
Hs	0.000	0.000	1	0.425	Hydrogen

Table B.2: L-glutamine bond lengths

Atom 1	Atom 2	Bond Length (Å)
Cb	Ca	1.514
Cb	Ob1	1.373
Cb	Ob2	1.212
Ca	Cm	1.551
Ca	Nb	1.469
Nb	Hb1	1.044
Nb	Hb2	1.043
Nb	Hb3	1.045
Ca	Hm1	1.117
Cm1	Cm2	1.541
Cm1	Hm2	1.110
Cm1	Hm3	1.111
Cm2	Hm4	1.112
Cm2	Hm5	1.108
Cm2	Cs	1.508
Cs	Os	1.222
Cs	Ns	1.415
Ns	Hs1	1.029
Ns	Hs2	1.028

Table B.3: L-glutamine bond angles

Atom 1	Atom 2	Atom 3	Angle (°)
Ca	Cb	Ob1	120.97
Ca	Cb	Ob2	120.09
Ob1	Cb	Ob2	118.92
Cm1	Ca	Cb	112.91
Cm1	Ca	Nb	110.95
Cm1	Ca	Hm1	106.69
Cb	Ca	Nb	111.27
Cb	Ca	Hm1	106.83
Nb	Ca	Hm1	107.86
Ca	Cm1	Hm2	110.29
Ca	Cm1	Hm3	106.14
Ca	Cm1	Cm2	114.74
Cm2	Cm1	Hm2	108.86
Cm2	Cm1	Hm3	109.12
Hm2	Cm1	Hm3	107.42
Ca	Nb	Hb1	110.03
Ca	Nb	Hb2	111.73
Ca	Nb	Hb3	109.66
Hb1	Nb	Hb2	109.12
Hb1	Nb	Hb3	108.00
Hb2	Nb	Hb3	108.21
Cm1	Cm2	Hm4	109.59
Cm1	Cm2	Hm5	111.38
Cm1	Cm2	Cs	110.48
Hm4	Cm2	Hm5	108.28
Hm4	Cm2	Cs	111.38
Hm5	Cm2	Cs	105.88
Cm2	Cs	Os	118.67
Cm2	Cs	Ns	122.14
Os	Cs	Ns	119.16

Atom 1	Atom 2	Atom 3	Angle (°)
Cs	Ns	Hs1	121.02
Cs	Ns	Hs2	119.86
Hs1	Ns	Hs2	119.11
Os	Cs	Ns	119.16
Cs	Ns	Hs1	121.02
Cs	Ns	Hs2	119.86
Hs1	Ns	Hs2	119.11

Table B.4: L-glutamine dihedral angles

Atom 1	Atom 2	Atom 3	Atom 4	Angle (°)
Cb	Ca	Cm1	Hm2	61.19
Cb	Ca	Cm1	Cm2	-62.17
Cb	Ca	Cm1	Hm3	177.25
Nb	Ca	Cm1	Hm2	-173.11
Nb	Ca	Cm1	Cm2	63.54
Nb	Ca	Cm1	Hm3	-57.05
Hm1	Ca	Cm1	Hm2	-55.87
Hm1	Ca	Cm1	Cm2	-179.22
Hm1	Ca	Cm1	Hm3	60.19
Cm1	Ca	Cb	Ob1	122.10
Cm1	Ca	Cb	Ob2	-59.43
Nb	Ca	Cb	Ob1	-3.43
Nb	Ca	Cb	Ob2	175.04
Hm1	Ca	Cb	Ob1	-120.93
Hm1	Ca	Cb	Ob2	57.54
Cm1	Ca	Nb	Hb1	52.60
Cm1	Ca	Nb	Hb2	-68.77
Cm1	Ca	Nb	Hb3	171.25
Cm1	Ca	Nb	Hb1	52.60
Cm1	Ca	Nb	Hb2	-68.77
Cm1	Ca	Nb	Hb3	171.25
Cb	Ca	Nb	Hb1	179.22
Cb	Ca	Nb	Hb2	57.84
Cb	Ca	Nb	Hb3	-62.13
Hm1	Ca	Nb	Hb1	-63.91
Hm1	Ca	Nb	Hb2	174.71
Hm1	Ca	Nb	Hb3	54.74
Ca	Cm1	Cm2	Hm4	70.11
Ca	Cm1	Cm2	Hm5	-49.55
Ca	Cm1	Cm2	Cs	-166.80

Atom 1	Atom 2	Atom 3	Atom 4	Angle (°)
Hm2	Cm1	Cm2	Hm4	-54.01
Hm2	Cm1	Cm2	Hm5	-173.67
Hm2	Cm1	Cm2	Cs	69.08
Hm3	Cm1	Cm2	Hm4	-170.97
Hm3	Cm1	Cm2	Hm5	69.38
Hm3	Cm1	Cm2	Cs	-47.87
Cm1	Cm2	Cs	Os	61.60
Cm1	Cm2	Cs	Ns	-120.42
Cm4	Cm2	Cs	Os	-176.36
Cm4	Cm2	Cs	Ns	1.62
Cm5	Cm2	Cs	Os	-58.87
Cm5	Cm2	Cs	Ns	119.11
Cm2	Cs	Ns	Hs1	1.42
Cm2	Cs	Ns	Hs2	-178.11
Os	Cs	Ns	Hs1	179.39
Os	Cs	Ns	Hs2	-0.14

Table B.5: L-glutamine EPSR parameters 25-biomolecule simulation

Xg	Temperature (K)	No. Gln	No. Water	No. Atoms	ρ (Atoms $\AA^{-3}$)
0.00370	297	25	6725	20,675	0.1000

Table B.6: L-glutamine EPSR box sizes 25-biomolecule simulation

Xg	Temperature	Box Length (Å)	Volume ($\AA^3$)
0.00370	297	59.13	206,750

Table B.7: L-glutamine EPSR parameters 200-biomolecule simulation

Xg	Temperature (K)	No. Gln	No. Water	No. Atoms	ρ (Atoms $\mathring{A}^{-3}$)
0.00370	297	200	53,800	165,400	0.1000
0.00370	310	200	53,800	165,400	0.0997
0.00370	333	200	53,800	165,400	0.0987

Table B.8: L-glutamine EPSR box sizes 200-biomolecule simulation

Xg	Temperature	Box Length (Å)	Box Width (Å)	Box Depth (Å)	Volume ($\mathring{A}^3$)
0.00370	297	187.73	93.86	93.86	1,654,000
0.00370	310	188.29	94.14	94.14	1,658,977
0.00370	333	190.20	95.10	95.10	1,675,785

B.2 Glutamine Imino Acid Reference Potentials and Parameters

For atom type Ob, Hm, Hb and Hs, these labels have been allocated to more than one atom. Whilst all atoms with a given label will have the same potentials, the bond lengths, angles and dihedrals are specified according the specific atoms. For geometric parameters, atoms of a given type also have numbers, e.g. Ob1 and Ob2.

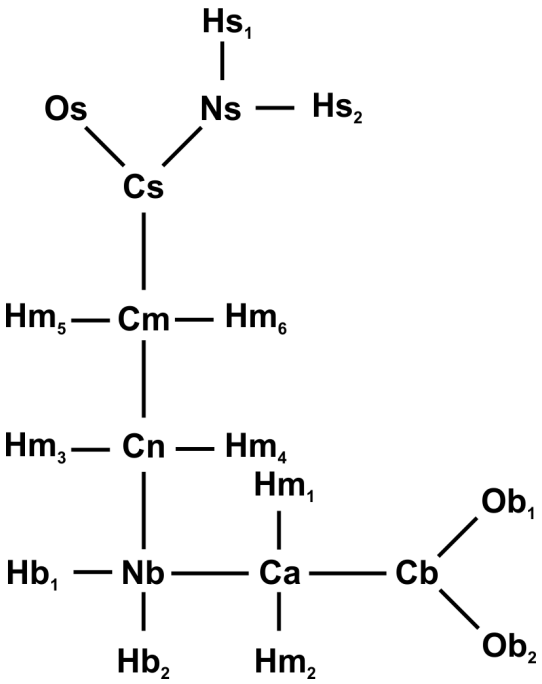


Table B.9: Glutamine imino acid potentials

Atom Label	ϵ (kJ mol^{-1})	σ (Å)	Mass (a.m.u.)	q(e)	Atom Type
Cb	0.43932	3.750	12	0.700	Carbon
Ca	0.49371	3.800	12	0.190	Carbon
Cn	0.49371	3.800	12	0.290	Carbon
Cm	0.49371	3.905	12	0.000	Carbon
Cs	0.43932	3.750	12	0.500	Carbon
Nb	0.71128	3.250	14	-0.200	Nitrogen
Ns	0.71128	3.250	14	-0.850	Nitrogen
Ob	0.87864	2.960	16	-0.800	Oxygen
Os	0.87864	2.960	16	-0.500	Oxygen
Hb	0.000	0.000	1	0.310	Hydrogen
Hm	0.000	0.000	1	0.000	Hydrogen
Hs	0.000	0.000	1	0.425	Hydrogen

Table B.10: Glutamine imino acid bond lengths

Atom 1	Atom 2	Bond Length (Å)
Cb	Ca	1.588
Cb	Ob1	1.251
Cb	Ob2	1.251
Ca	Nb	1.488
Ca	Hm1	1.124
Ca	Hm2	1.124
Cn	Nb	1.488
Nb	Hb1	1.030
Nb	Hb2	1.030
Cn	Cm	1.625
Cn	Hm3	1.124
Cn	Hm4	1.124
Cm	Cs	1.516
Cm	Hm5	1.126
Cm	Hm6	1.126
Cs	Ns	1.367
Cs	Os	1.251
Ns	Hs1	0.989
Ns	Hs2	0.989

Table B.11: Glutamine imino acid bond angles

Atom 1	Atom 2	Atom 3	Angle (°)
Ca	Cb	Ob1	115.10
Ca	Cb	Ob2	115.10
Ob1	Cb	Ob2	129.79
Cb	Ca	Nb	110.38
Cb	Ca	Hm1	108.70
Cb	Ca	Hm2	108.70
Nb	Ca	Hm1	108.78
Nb	Ca	Hm2	110.29
Hm1	Ca	Hm2	109.96
Ca	Nb	Cn	113.19
Ca	Nb	Hb1	108.00
Ca	Nb	Hb2	108.00
Cn	Nb	Hb1	109.66
Cn	Nb	Hb2	107.90
Hb1	Nb	Hb2	110.06
Nb	Cn	Cm	112.35
Nb	Cn	Hm3	108.48
Nb	Cn	Hm4	108.49
Cm	Cn	Hm3	109.73
Cm	Cn	Hm4	109.47
Hm3	Cn	Hm4	108.22
Cn	Cm	Cs	109.04
Cn	Cm	Hm5	110.42
Cn	Cm	Hm6	110.41
Cs	Cm	Hm5	110.46
Cs	Cm	Hm6	110.23
Hm5	Cm	Hm6	106.26
Cm	Cs	Ns	117.57
Cm	Cs	Os	121.50

Atom 1	Atom 2	Atom 3	Angle (°)
Ns	Cs	Os	120.93
Cs	Ns	Hs1	120.39
Cs	Ns	Hs2	120.39
Hs1	Ns	Hs2	119.21

Table B.12: Glutamine imino acid dihedral angles (red for dihedrals constrained in partially-fixed simulations)

Atom 1	Atom 2	Atom 3	Atom 4	Angle (°)
Ob1	Cb	Ca	Nb	-12.75
Ob1	Cb	Ca	Hm1	106.48
Ob1	Cb	Ca	Hm2	-133.84
Ob2	Cb	Ca	Nb	166.35
Ob2	Cb	Ca	Hm1	-74.42
Ob2	Cb	Ca	Hm2	45.26
Cb	Ca	Nb	Cn	-90.27
Cb	Ca	Nb	Hb1	148.12
Cb	Ca	Nb	Hb2	29.12
Hm1	Ca	Nb	Cn	150.54
Hm1	Ca	Nb	Hb1	28.93
Hm1	Ca	Nb	Hb2	-90.07
Hm2	Ca	Nb	Cn	29.86
Hm2	Ca	Nb	Hb1	-91.74
Hm2	Ca	Nb	Hb2	149.26
Ca	Nb	Cn	Cm	61.38
Ca	Nb	Cn	Hm3	-60.11
Ca	Nb	Cn	Hm4	-177.45

Atom 1	Atom 2	Atom 3	Atom 4	Angle (°)
Hb1	Nb	Cn	Cm	-177.95
Hb1	Nb	Cn	Hm3	60.56
Hb1	Nb	Cn	Hm4	-56.78
Hb2	Nb	Cn	Cm	-58.07
Hb2	Nb	Cn	Hm3	-179.56
Hb2	Nb	Cn	Hm4	63.10
Nb	Cn	Cm	Cs	170.51
Nb	Cn	Cm	Hm5	-67.95
Nb	Cn	Cm	Hm6	49.26
Hm3	Cn	Cm	Cs	-68.72
Hm3	Cn	Cm	Hm5	52.82
Hm3	Cn	Cm	Hm6	170.03
Hm4	Cn	Cm	Cs	49.91
Hm4	Cn	Cm	Hm5	171.45
Hm4	Cn	Cm	Hm6	-71.34
Cn	Cm	Cs	Ns	177.06
Cn	Cm	Cs	Os	-3.14
Hm5	Cm	Cs	Ns	55.55
Hm6	Cm	Cs	Ns	-61.58
Hm5	Cm	Cs	Os	-124.66
Hm6	Cm	Cs	Os	118.22
Cm	Cs	Ns	Hs1	179.77
Cm	Cs	Ns	Hs2	-0.04
Os	Cs	Ns	Hs1	-0.03
Os	Cs	Ns	Hs2	-179.84

Table B.13: Glutamine imino acid EPSR parameters

Xg	Temperature (K)	No. Ngln	No. Water	No. Atoms	ρ (Atoms $\mathring{A}^{-1}$)
0.00369	298	25	6750	20,750	0.1000
0.03571	298	100	2700	10,100	0.1054
0.03571	310	100	2700	10,100	0.1049
0.03571	333	100	2700	10,100	0.1036

Table B.14: Glutamine imino acid EPSR box sizes

Xg	Temperature	Box Length (Å)	Volume ($\mathring{A}^3$)
0.00369	298	59.20	207,500
0.03571	298	45.76	95,825
0.03571	310	45.83	96,282
0.03571	333	46.02	97,490

B.3 Sarcosine Reference Potentials and Parameters

For atom type Ob, Ha, Hb and Hs, these labels have been allocated to more than one atom. Whilst all atoms with a given label will have the same potentials, the bond lengths, angles and dihedrals are specified according the specific atoms. For geometric parameters, atoms of a given type also have numbers, e.g. Ob1 and Ob2.

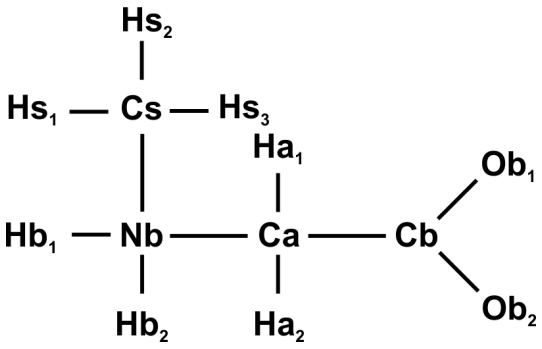


Table B.15: Sarcosine potentials

Atom Label	ϵ (kJ mol^{-1})	σ (Å)	Mass (a.m.u.)	q(e)	Atom Type
Cb	0.43932	3.750	12	0.700	Carbon
Ca	0.49371	3.800	12	0.070	Carbon
Cs	0.66944	3.910	12	0.110	Carbon
Nb	0.71128	3.250	14	-0.200	Nitrogen
Ob	0.87864	2.960	16	-0.800	Oxygen
Hb	0.000	0.000	1	0.310	Hydrogen
Ha	0.12558	2.500	1	0.060	Hydrogen
Hs	0.12558	2.500	1	0.060	Hydrogen

Table B.16: Sarcosine bond lengths

Atom 1	Atom 2	Bond Length (Å)
Cb	Ca	1.597
Cb	Ob1	1.249
Cb	Ob2	1.249
Ca	Nb	1.487
Ca	Ha1	1.119
Ca	Ha2	1.119
Nb	Cs	1.468
Nb	Hb1	1.030
Nb	Hb1	1.030
Cs	Hs1	1.122
Cs	Hs2	1.122
Cs	Hs3	1.122

Table B.17: Sarcosine bond angles

Atom 1	Atom 2	Atom 3	Angle (°)
Ca	Cb	Ob1	114.55
Ca	Cb	Ob2	114.55
Ob1	Cb	Ob2	130.90
Cb	Ca	Nb	109.66
Cb	Ca	Ha1	109.03
Cb	Ca	Ha2	109.03
Nb	Ca	Ha1	109.49
Nb	Ca	Ha2	109.50
Ha1	Ca	Ha2	110.12
Ca	Nb	Cs	114.29
Ca	Nb	Hb1	106.79
Ca	Nb	Hb2	106.79
Cs	Nb	Hb1	111.29
Cs	Nb	Hb2	111.28
Hb1	Nb	Hb2	105.92
Nb	Cs	Hs1	109.58
Nb	Cs	Hs2	109.57
Nb	Cs	Hs3	109.57
Hs1	Cs	Hs2	109.57
Hs1	Cs	Hs3	109.97
Hs2	Cs	Hs3	109.56

Table B.18: Sarcosine dihedral angles

Atom 1	Atom 2	Atom 3	Atom 4	Angle (°)
Ob1	Cb	Ca	Nb	-0.06
Ob1	Cb	Ca	Ha1	119.81
Ob1	Cb	Ca	Ha2	-119.94
Ob2	Cb	Ca	Nb	179.94
Ob2	Cb	Ca	Ha1	-60.19
Ob2	Cb	Ca	Ha2	60.06
Cb	Ca	Nb	Cs	179.84
Cb	Ca	Nb	Hb1	56.32
Cb	Ca	Nb	Hb2	-56.66
Ha1	Ca	Nb	Cs	60.26
Ha1	Ca	Nb	Hb1	-63.27
Ha1	Ca	Nb	Hb2	-176.24
Ha2	Ca	Nb	Cs	-60.57
Ha2	Ca	Nb	Hb1	175.91
Ha2	Ca	Nb	Hb2	62.93
Ca	Nb	Cs	Hs1	59.82
Ca	Nb	Cs	Hs2	-179.92
Ca	Nb	Cs	Hs3	-59.69
Hb1	Nb	Cs	Hs1	-179.11
Hb1	Nb	Cs	Hs2	-58.86
Hb1	Nb	Cs	Hs3	61.38
Hb2	Nb	Cs	Hs1	-61.22
Hb2	Nb	Cs	Hs2	59.03
Hb2	Nb	Cs	Hs3	179.27

Table B.19: Sarcosine EPSR parameters

Xg	Temperature (K)	No. Sar	No. Water	No. Atoms	ρ (Atoms $\mathring{A}^{-3}$)
0.03571	298	100	2700	9,300	0.1024

Table B.20: Sarcosine EPSR box sizes

Xg	Temperature	Box Length (Å)	Volume ($\mathring{A}^3$)
0.03571	298	44.95	90,820

B.4 Water Reference Potentials and Parameters

Table B.21: Water potentials

Atom Label	ϵ (kJ mol^{-1})	σ (Å)	Mass (a.m.u.)	q(e)	Atom Type
Ow	0.650	3.166	16	-0.8476	Oxygen
Hw	0.000	0.000	1	0.4238	Hydrogen

Table B.22: Water bond lengths

Atom 1	Atom 2	Bond Length (Å)	Concentration (mg ml^{-1})
Ow	Hw	0.976	30
Ow	Hw	0.980	300

Table B.23: Water bond angle

Atom 1	Atom 2	Atom 3	Angle (°)
Hw	Ow	Hw	104.50

Appendix C

Density Readings

C.1 Method for taking Density Measurements

Density readings have been taken for all 300 mg ml^{-1} solutions, equivalent in concentration to those used for the neutron diffraction experiments. These measurements were obtained using an Anton Paar DMA 4100 densitometer. The densitometer is able to record readings between 273 to 368 K, meaning density readings can be taken for all samples at all temperatures studied with neutron diffraction.

The densitometer is initially calibrated with respect to current air pressure and water density. The water density is for a sample of degassed water at 293 K, set to have a reading of 0.99820 g cm^{-3}. After cleaning the cuvette with water and ethanol, the degassed sample is put into the cuvette. Once the densitometer has equilibrated to reach the desired temperature, a reading can be taken. The densitometer has taken readings in 'density1' mode. All readings have been taken a minimum of 3 times and averaged to obtain the final value. The density readings are taken with 0.0001 g cm^{-3} accuracy and the temperature to 0.05 K accuracy.

C.2 Density Readings

Table C.1: Glutamine imino acid density readings

Xg	Temperature (K)	Density ($g\,cm^{-3}$)
0.03571	298	1.0962
0.03571	310	1.0910
0.03571	333	1.0780

Table C.2: Sarcosine density readings

Xg	Temperature (K)	Density ($g\,cm^{-3}$)
0.03571	298	1.0360

To convert between density (D) and atomic number density (ρ), the following equation is used:

$$\rho = \frac{DN_A}{A} \tag{C.1}$$

where N_A is Avogadro's constant and A the average atomic mass of an atom in the system.

References

[1] N.H. Rhys, A.K. Soper, and L. Dougan, "The Hydrogen-Bonding Ability of the Amino Acid Glutamine Revealed by Neutron Diffraction Experiments.", *The Journal of Physical Chemistry B* **116**(45), pp. 13308–13319 (2012).

[2] N.H. Rhys and L. Dougan, "The emerging role of hydrogen bond interactions in polyglutamine structure, stability and association", *Soft Matter* **9**(8), pp. 2359–2364 (2013).

[3] N.H. Rhys, A.K. Soper, and L. Dougan, "Hydrophilic association in a dilute glutamine solution is persistent with increasing temperature", *Submitted manuscript* (2015).

[4] T.M. Devlin, *Textbook of biochemistry: with clinical correlations*, Wiley-Liss New York (2002).

[5] P.J. Reeds, "Dispensable and Indispensable Amino Acids for Humans", *Journal of Nutrition* **130**(7), pp. 1835–1840 (2000).

[6] J. Clayden, N. Greeves, and S. Warren, *Organic Chemistry*, Oxford University Press Oxford (2012).

[7] J.M. Berg, J.L. Tymoczko, and L. Stryer, *Biochemistry*, W H Freeman New York 5th edition (2002).

[8] S. Vijay-Kumar, C.E. Bugg, and W.J. Cook, "Structure of ubiquitin refined at 1.8 Å resolution", *The Journal of Molecular Biology* **194**(3), pp. 531–544 (1987).

[9] J.N. Israelachvili, *Intermolecular and Surface Forces*, Academic Press London (2010).

[10] S.K. Burley and G.A. Petsko, "Weakly polar interactions in proteins.", *Advances in Protein Chemistry* **39**, pp. 125–89 (1988).

[11] K.A. Sharp and S.W. Englander, "How much is a stabilizing bond worth?", *Trends in Biochemical Sciences* **19**(12), pp. 526–529 (1994).

[12] K-T. Tang and J.P. Toennies, "Johannes Diderik van der Waals: a pioneer in the molecular sciences and Nobel Prize Winner in 1910.", *Angewandte Chemie (International ed. in English)* **49**(50), pp. 9574–9 (2010).

[13] F. London, "Zur Theorie und Systematik der Molekularkräfte", *Zeitschrift für Physik* **63**(3-4), pp. 245–279 (1930).

[14] P. Debye, "'Van der Waals' Cohesive Force (Die van der Waalsschen Kohasioskrafte)", *Physiks Zeitschrift* **21**, pp. 178–187 (1920).

[15] W.H. Keesom, "On the deduction from Boltzmanns entropy principle of the second virial-coefficient of material particles (in the limit of rigid spheres of central symmetry) which exert central forces upon each other and for rigid spheres of central symmetry containing", *Proceedings of the Royal Academy of Amsterdam* **15**, pp. 256–273 (1912).

[16] T. Cserháti and M. Szögyi, "Role of hydrophobic and hydrophilic forces in peptide-protein interaction: New advances", *Peptides* **16**(1), pp. 165–173 (1995).

[17] H. Meyer, "Zur theorie der alkoholnarkose: I. Welche eigenschaft der anaesthetika bedingt ihre narkotische wirkung?", *Archiv for Experimentelle Pathologie und Pharmakologie* **42**(2-4), pp. 109–118 (1899).

[18] E. Overton, "Über die allgemeinen osmotischen eigenschaften der zelle, ihre vermutlichen ursachen und ihre bedeutung für die physiologie.", *Vierteljahrsschr. Naturforsch. Ges. Zürich.* **44**, pp. 87–136 (1899).

[19] A. Sarkar and G.E. Kellogg, "Hydrophobicity–shake flasks, protein folding and drug discovery.", *Current Topics in Medicinal Chemistry* **10**(1), pp. 67–83 (2010).

[20] J.A.V. Butler, "The energy and entropy of hydration of organic compounds", *Transactions of the Faraday Society* **33**, pp. 229 (1937).

[21] H.S. Frank and M.W. Evans, "Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes", *The Journal of Chemical Physics* **13**(11), pp. 507 (1945).

[22] W Kauzmann, "Some factors in the interpretation of protein denaturation.", *Advances in Protein Chemistry* **14**, pp. 1–63 (1959).

[23] D. Chandler, "Interfaces and the driving force of hydrophobic assembly", *Nature* **437**, pp. 640–647 (2005).

[24] A. Grauer and B. König, "Peptidomimetics - A Versatile Route to Biologically Active Compounds", *European Journal of Organic Chemistry* **2009**(30), pp. 5099–5111 (2009).

[25] T.A. Martinek and F. Fülöp, "Peptidic foldamers: ramping up diversity.", *Chemical Society Reviews* **41**(2), pp. 687–702 (2012).

[26] I. Avan, C.D. Hall, and A.R. Katritzky, "Peptidomimetics via modifications of amino acids and peptide bonds.", *Chemical Society Reviews* **43**(10), pp. 3575–3594 (2014).

[27] A.E. Rawlings, J.P. Bramble, and S.S. Staniland, "Innovation through imitation: biomimetic, bioinspired and biokleptic research", *Soft Matter* **8**(25), pp. 6675–6679 (2012).

[28] J.C. Johnson and L.T.J. Korley, "Enhanced mechanical pathways through nature's building blocks: amino acids", *Soft Matter* **8**(45), pp. 11431–11442 (2012).

[29] E. Ko, J. Liu, L.M. Perez, G. Lu, A. Schaefer, and K. Burgess, "Universal peptidomimetics.", *Journal of the American Chemical Society* **133**(3), pp. 462–477 (2011).

[30] J.C. Johnson, N.D. Wanasekara, and L.T.J. Korley, "Utilizing peptidic ordering in the design of hierarchical polyurethane/ureas.", *Biomacromolecules* **13**(5), pp. 1279–1286 (2012).

[31] R.J. Simon, R.S. Kania, R.N. Zuckermann, V.D. Huebner, D.A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, and C.K. Marlowe, "Peptoids: a modular approach to drug discovery.", *Proceedings of the National Academy of Sciences of the United States of America* **89**(20), pp. 9367–9371 (1992).

[32] P.S. Farmer and E.J. Ariens, "Speculations on the design of nonpeptidic peptidomimetics", *Trends in Pharmacological Sciences* **3**, pp. 362–365 (1982).

[33] R.N. Zuckermann, J.M. Kerr, S.B.H. Kent, and W.H. Moos, "Efficient method for the preparation of peptoids [oligo(N-substituted glycines)] by submonomer solid-phase synthesis", *Journal of the American Chemical Society* **114**(26), pp. 10646–10647 (1992).

[34] S.A. Fowler and H.E. Blackwell, "Structure-function relationships in peptoids: recent advances toward deciphering the structural requirements for biological function.", *Organic and Biomolecular Chemistry* **7**(8), pp. 1508–1524 (2009).

[35] J. Seo, B-C. Lee, and R.N. Zuckermann, "Peptoids - Synthesis, Characterization, and Nanostructures", In P. Ducheyne, editor, *Comprehensive Biomaterials* volume 2 , pp. 53–76. Elsevier Ltd. Salt Lake City (2011).

[36] A.M. Rosales, R.A. Segalman, and R.N. Zuckermann, "Polypeptoids: a model system to study the effect of monomer sequence on polymer properties and self-assembly", *Soft Matter* **9**(35), pp. 8400–8414 (2013).

[37] J. Sun and R.N. Zuckermann, "Peptoid polymers: a highly designable bioinspired material.", *ACS Nano* **7**(6), pp. 4715–4732 (2013).

[38] G.P. Moss, P.A.S. Smith, and D. Tavernier, "Glossary of class names of organic compounds and reactivity intermediates based on structure (IUPAC Recommendations 1995)", *Pure and Applied Chemistry* **67**(8-9), pp. 1307–1375 (1995).

[39] J.A.W. Kruijtzer, *Synthesis of Peptoid Peptidomimetics*, Universiteit Utrecht (1996).

[40] J.S. Laursen, J. Engel-Andreasen, P. Fristrup, P. Harris, and C.A. Olsen, "Cis-trans amide bond rotamers in β-peptoids and peptoids: evaluation of stereoelec-

tronic effects in backbone and side chains.", *Journal of the American Chemical Society* **135**(7), pp. 2835–2844 (2013).

[41] S.M. Miller, R.J. Simon, Si. Ng, R.N. Zuckermann, J.M. Kerr, and W.H. Moos, "Proteolytic studies of homologous peptide and N-substituted glycine peptoid oligomers", *Bioorganic and Medicinal Chemistry Letters* **4**(22), pp. 2657–2662 (1994).

[42] S.M. Miller, R.J. Simon, Si. Ng, R.N. Zuckermann, J.M. Kerr, and W.H. Moos, "Comparison of the proteolytic susceptibilities of homologous L-amino acid, D-amino acid, and N-substituted glycine peptide and peptoid oligomers", *Drug Development Research* **35**(1), pp. 20–32 (1995).

[43] B. Sanii, R. Kudirka, A. Cho, N. Venkateswaran, G.K. Olivier, A.M. Olson, H. Tran, R.M. Harada, Li. Tan, and R.N. Zuckermann, "Shaken, not stirred: collapsing a peptoid monolayer to produce free-floating, stable nanosheets.", *Journal of the American Chemical Society* **133**(51), pp. 20808–20815 (2011).

[44] H.K. Murnen, A.M. Rosales, J.N. Jaworski, R.A. Segalman, and R.N. Zuckermann, "Hierarchical self-assembly of a biomimetic diblock copolypeptoid into homochiral superhelices.", *Journal of the American Chemical Society* **132**(45), pp. 16112–16119 (2010).

[45] T.J. Sanborn, C.W. Wu, R.N. Zuckermann, and A.E. Barron, "Extreme stability of helices formed by water-soluble poly-N-substituted glycines (polypeptoids) with alpha-chiral side chains.", *Biopolymers* **63**(1), pp. 12–20 (2002).

[46] N.C. Tan, P. Yu, Y-U. Kwon, and T. Kodadek, "High-throughput evaluation of relative cell permeability between peptoids and peptides.", *Bioorganic and Medicinal Chemistry* **16**(11), pp. 5853–5861 (2008).

[47] Y-C. Tang and C.M. Deber, "Hydrophobicity and helicity of membrane-interactive peptides containing peptoid residues.", *Biopolymers* **65**(4), pp. 254–262 (2002).

[48] W.M. Haynes, editor, *CRC Handbook of Chemistry and Physics, 92nd Edition*, CRC Press Boca Raton, FL (2011).

[49] R.N. Zuckermann and T. Kodadek, "Peptoids as potential therapeutics.", *Current Opinion in Molecular Therapeutics* **11**(3), pp. 299–307 (2009).

[50] D.G. Udugamasooriya, S.P. Dineen, R.A. Brekken, and T. Kodadek, "A peptoid "antibody surrogate" that antagonizes VEGF receptor 2 activity.", *Journal of the American Chemical Society* **130**(17), pp. 5744–5752 (2008).

[51] D.G. Udugamasooriya, G. Dunham, C. Ritchie, R.A. Brekken, and T. Kodadek, "The pharmacophore of a peptoid VEGF receptor 2 antagonist includes both side chain and main chain residues.", *Bioorganic and medicinal chemistry medicinal chemistry letters* **18**(22), pp. 5892–5894 (2008).

[52] G.A. Eggimann, H.L. Bolt, P.W. Denny, and S.L. Cobb, "Investigating the Anti-leishmanial Effects of Linear Peptoids.", *ChemMedChem* **10**(2), pp. 233–237 (2015).

[53] S.L. Seurynck-Servoss, M.T. Dohm, and A.E. Barron, "Effects of including an N-terminal insertion region and arginine-mimetic side chains in helical peptoid analogues of lung surfactant protein B.", *Biochemistry* **45**(39), pp. 11809–11818 (2006).

[54] N.J. Brown, C.W. Wu, S.L. Seurynck-Servoss, and A.E. Barron, "Effects of hydrophobic helix length and side chain chemistry on biomimicry in peptoid analogues of SP-C.", *Biochemistry* **47**(6), pp. 1808–1818 (2008).

[55] J.A. Patch and A.E. Barron, "Helical peptoid mimics of magainin-2 amide.", *Journal of the American Chemical Society* **125**(40), pp. 12092–12093 (2003).

[56] N.P. Chongsiriwatana, J.A. Patch, A.M. Czyzewski, M.T. Dohm, A. Ivankin, D. Gidalevitz, R.N. Zuckermann, and A.E. Barron, "Peptoids that mimic the structure, function, and mechanism of helical antimicrobial peptides.", *Proceedings of the National Academy of Sciences of the United States of America* **105**(8), pp. 2794–1799 (2008).

[57] M.M. Reddy, R. Wilson, J. Wilson, S. Connell, A. Gocke, Li. Hynan, D. German, and T. Kodadek, "Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening.", *Cell* **144**(1), pp. 132–142 (2011).

[58] B-C. Lee, T.K. Chu, K.A. Dill, and R.N. Zuckermann, "Biomimetic nanostructures: creating a high-affinity zinc-binding site in a folded nonbiological polymer.", *Journal of the American Chemical Society* **130**(27), pp. 8847–8855 (2008).

[59] C.A. Olsen, "Beta-peptoid "foldamers"–why the additional methylene unit?", *Biopolymers* **96**(5), pp. 561–566 (2011).

[60] P.H. Yancey, M.E. Clark, S.C. Hand, R.D. Bowlus, and G.N. Somero, "Living with Water-Stress - Evolution of Osmolyte Systems", *Science* **217**(4566), pp. 1214–1222 (1982).

[61] T. Arakawa and S.N. Timasheff, "The stabilization of proteins by osmolytes.", *Biophysical Journal* **47**(3), pp. 411–414 (1985).

[62] P.H. Yancey, "Water Stress, Osmolytes and Proteins", *Integrative and Comparative Biology* **41**(4), pp. 699–709 (2001).

[63] P.H. Yancey, "Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses.", *The Journal of Experimental Biology* **208**(Pt 15), pp. 2819–2830 (2005).

[64] M.L. Huang, D. Ehre, Q. Jiang, C. Hu, K. Kirshenbaum, and M.D. Ward, "Biomimetic peptoid oligomers as dual-action antifreeze agents.", *Proceedings of the National Academy of Sciences of the United States of America* **109**(49), pp. 19922–19927 (2012).

[65] Z. Ignatova and L.M. Gierasch, "Extended polyglutamine tracts cause aggregation and structural perturbation of an adjacent beta barrel protein.", *The Journal of Biological Chemistry* **281**(18), pp. 12959–12967 (2006).

[66] Z. Ignatova and L.M. Gierasch, "Effects of osmolytes on protein folding and aggregation in cells.", *Methods in Enzymology* **428**, pp. 355–372 (2007).

[67] N. Kumar and N. Kishore, "Structure and effect of sarcosine on water and urea by using molecular dynamics simulations: Implications in protein stabilization.", *Biophysical Chemistry* **171**, pp. 9–15 (2013).

[68] C.W. Wu, T.J. Sanborn, R.N. Zuckermann, and A.E. Barron, "Peptoid Oligomers with α-Chiral, Aromatic Side Chains: Effects of Chain Length on Secondary Structure", *Journal of the American Chemical Society* **123**(13), pp. 2958–2963 (2001).

[69] J.T. Edsall, "Raman Spectra of Amino Acids and Related Compounds. VI. Sarcosine, Ethanolamine, Choline, Betaine and Betaine Derivatives1", *Journal of the American Chemical Society* **65**(9), pp. 1767–1770 (1943).

[70] A.M. Rosales, H.K. Murnen, S.R. Kline, R.N. Zuckermann, and R.A. Segalman, "Determination of the persistence length of helical and non-helical polypeptoids in solution", *Soft Matter* **8**(13), pp. 3673–3680 (2012).

[71] C.W. Wu, K. Kirshenbaum, T.J. Sanborn, J.A. Patch, K. Huang, K.A. Dill, R.N. Zuckermann, and A.E. Barron, "Structural and spectroscopic studies of peptoid oligomers with alpha-chiral aliphatic side chains.", *Journal of the American Chemical Society* **125**(44), pp. 13525–13530 (2003).

[72] P. Armand, K. Kirshenbaum, A. Falicov, R.L. Dunbrack, K.A. Dill, R.N. Zuckermann, and F.E Cohen, "Chiral N-substituted glycines can form stable helical conformations.", *Folding and Design* **2**(6), pp. 369–375 (1997).

[73] C Hammond, *The Basics of Crystallography and Diffraction*, Oxford University Press Oxford (2001).

[74] B.T.M. Willis and C.J. Carlile, *Experimental Neutron Scattering*, Oxford University Press Inc. New York, NY (2009).

[75] D S Sivia, *Elementary Scattering Theory: For X-ray and Neutron Users*, Oxford University Press Oxford (2011).

[76] J.R. Ferraro, K. Nakamoto, and C.W. Brown, *Introductory Raman Spectroscopy*, Academic Press Amsterdam (2003).

[77] M. Heyden, J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith, and D. Marx, "Dissecting the THz spectrum of liquid water from first principles via correlations in time and space.", *Proceedings of the National Academy of Sciences of the United States of America* **107**(27), pp. 12068–12073 (2010).

[78] K.M. Tych, C.D. Wood, A.D. Burnett, A.R. Pearson, A.G. Davies, E.H. Linfield, and J.E. Cunningham, "Probing temperature- and solvent-dependent protein dynamics using terahertz time-domain spectroscopy", *Journal of Applied Crystallography* **47**(1), pp. 146–153 (2013).

[79] K. Shiraga, T. Suzuki, N. Kondo, and Y. Ogawa, "Hydration and hydrogen bond network of water around hydrophobic surface investigated by terahertz spectroscopy", *The Journal of Chemical Physics* **141**(23), pp. 235103 (2014).

[80] G.N. Graham and I.J. Pickering, "X-ray Absorption Spectroscopy of Metals in Biology", In *Encyclopedia of Biophysics* , pp. 2762–2767. Springer Berlin Heidelberg Berlin, Heidelberg (2013).

[81] B. K. Teo and D. C. Joy, editors, *EXAFS Spectroscopy*, Springer US Boston, MA (1981).

[82] J.S. Uejio, C.P. Schwartz, A.M. Duffin, A. England, D. Prendergast, and R.J. Saykally, "Monopeptide versus monopeptoid: insights on structure and hydration of aqueous alanine and sarcosine via X-ray absorption spectroscopy.", *The Journal of Physical Chemistry. B* **114**(13), pp. 4702–4709 (2010).

[83] H. Günther, *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*, Wiley-VCH Verlag Weinheim (2013).

[84] P.G. Takis, K.D. Papavasileiou, L.D. Peristeras, V.S. Melissas, and A.N. Troganis, "Probing micro-solvation in "numbers": the case of neutral dipeptides in water.", *Physical Chemistry Chemical Physics* **15**(19), pp. 7354–7362 (2013).

[85] K. Huang, C.W. Wu, T.J. Sanborn, J.A. Patch, K. Kirshenbaum, R.N. Zuckermann, A.E. Barron, and I. Radhakrishnan, "A threaded loop conformation adopted by a family of peptoid nonamers.", *Journal of the American Chemical Society* **128**(5), pp. 1733–1738 (2006).

[86] T. Schlick, R. Collepardo-Guevara, L.A. Halvorsen, S. Jung, and Xi. Xiao, "Biomolecularmodeling and simulation: a field coming of age.", *Quarterly Reviews of Biophysics* **44**(2), pp. 191–228 (2011).

[87] M.J S. Dewar, E.G. Zoebisch, E.F. Healy, and J.J.P. Stewart, "Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model", *Journal of the American Chemical Society* **107**(13), pp. 3902–3909 (1985).

[88] T. Hassinen and M. Peräkylä, "New energy terms for reduced protein models implemented in an off-lattice force field", *Journal of Computational Chemistry* **22**(12), pp. 1229–1242 (2001).

[89] 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", *Journal of Computational Chemistry* **4**(2), pp. 187–217 (1983).

[90] P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, "A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules", *Journal of the American Chemical Society* **117**(19), pp. 5179–5197 (1995).

[91] W.L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin", *Journal of the American Chemical Society* **110**(6), pp. 1657–1666 (1988).

[92] 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", *Journal of the American Chemical Society* **118**(45), pp. 11225–11236 (1996).

[93] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, "The missing term in effective pair potentials", *The Journal of Physical Chemistry* **91**(24), pp. 6269–6271 (1987).

[94] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, "Comparison of simple potential functions for simulating liquid water", *The Journal of Chemical Physics* **79**(2), pp. 926–935 (1983).

[95] W.L. Jorgensen and J.D. Madura, "Temperature and size dependence for Monte Carlo simulations of TIP4P water", *Molecular Physics* **56**(6), pp. 1381–1392 (2006).

[96] R.O. Dror, R.M. Dirks, J.P. Grossman, H. Xu, and D.E. Shaw, "Biomolecular simulation: a computational microscope for molecular biology.", *Annual Review of Biophysics* **41**, pp. 429–52 (2012).

[97] M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press Oxford (1987).

[98] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines", *The Journal of Chemical Physics* **21**, pp. 1087–1092 (1953).

[99] B. Guillot, "A reappraisal of what we have learnt during three decades of computer simulations on water", *Journal of Molecular Liquids* **101**(1-3), pp. 219–260 (2002).

[100] A.K. Soper, "Probing the structure of water around biological molecules: concepts, constructs and consequences", *Physica B: Condensed Matter* **276-278**(1-3), pp. 12–16 (2000).

[101] J.J. Towey, A.K. Soper, and L. Dougan, "The structure of glycerol in the liquid state: A neutron diffraction study", *Physical Chemistry Chemical Physics* **12**, pp. 9347–9397 (2011).

[102] J.J. Towey, A.K. Soper, and L. Dougan, "Preference for isolated water molecules in a concentrated glycerol-water mixture.", *The Journal of Physical Chemistry. B* **115**(24), pp. 7799–7807 (2011).

[103] J.J. Towey and L. Dougan, "Structural Examination of the Impact of Glycerol on Water Structure", *The Journal of Physical Chemistry B* **116**(5), pp. 1633–1641 (2012).

[104] J.J. Towey, A.K. Soper, and L. Dougan, "Molecular Insight into the Hydrogen Bonding and Micro-segregation of a Cryoprotectant Molecule", *The Journal of Physical Chemistry B* **116**(47), pp. 13898–13904 (2012).

[105] J.J. Towey, A.K. Soper, and L. Dougan, "What Happens to the Structure of Water in Cryoprotectant Solutions?", *Faraday Discussions* **167**, pp. 159–176 (2013).

[106] D. T. Bowron, A. K. Soper, and J. L. Finney, "Temperature dependence of the structure of a 0.06 mole fraction tertiary butanol-water solution", *The Journal of Chemical Physics* **114**(14), pp. 6203 (2001).

[107] J.L. Finney, D.T. Bowron, and A.K. Soper, "The structure of aqueous solutions of tertiary butanol" (2000).

[108] D.T. Bowron, J.L. Finney, and A.K. Soper, "Structural Investigation of Solute - Solute Interactions in Aqueous Solutions of Tertiary Butanol", *The Journal of Physical Chemistry B* **102**(18), pp. 3551–3563 (1998).

[109] S Imberti and D Bowron, "Formic and Acetic Acid Aggregation in the Liquid State", *Journal of Physics: Condensed Matter* **22**(40), pp. 404212 (2010).

[110] F. Meersman, D. Bowron, A.K. Soper, and M.H.J. Koch, "Counteraction of Urea by Trimethylamine N-Oxide Is Due to Direct Interaction", *Biophysical Journal* **97**(9), pp. 2559–2566 (2009).

[111] A.K. Soper and A. Luzar, "Orientation of Water Molecules around Small Polar and Nonpolar Groups in Solution: A Neutron Diffraction and Computer Simulation Study", *The Journal of Physical Chemistry* **100**(4), pp. 1357–1367 (1996).

[112] S.E. McLain, A.K. Soper, and A. Luzar, "Orientational correlations in liquid acetone and dimethyl sulfoxide: A comparative study", *The Journal of Chemical Physics* **124**(7), pp. 74502 (2006).

[113] L. Dougan, J. Crain, J.L. Finney, and A.K. Soper, "Molecular self-assembly in a model amphiphile system.", *Physical Chemistry Chemical physics* **12**(35), pp. 10221–10229 (2010).

[114] L. Dougan, S.P. Bates, R. Hargreaves, J.P. Fox, J. Crain, J.L. Finney, V. Reat, and A.K. Soper, "Methanol-water solutions: A bi-percolating liquid mixture", *Journal of Chemical Physics* **121**(13), pp. 6456–6462 (2004).

[115] L. Dougan, R. Hargreaves, S.P. Bates, J.L. Finney, V. Reat, A.K. Soper, and J. Crain, "Segregation in aqueous methanol enhanced by cooling and compression", *The Journal of Chemical Physics* **122**(17), pp. 174514 (2005).

[116] A.K. Soper, L. Dougan, J. Crain, and J.L. Finney, "Excess Entropy in Alcohol-Water Solutions: A Simple Clustering Explanation", *The Journal of Physical Chemistry B* **110**(8), pp. 3472–3476 (2006).

[117] P. Sillrén, J. Swenson, J. Mattsson, D. Bowron, and A. Matic, "The temperature dependent structure of liquid 1-propanol as studied by neutron diffraction and EPSR simulations.", *The Journal of Chemical Physics* **138**(21), pp. 214501 (2013).

[118] A.J. Johnston, Y.R. Zhang, S. Busch, L.C. Pardo, S. Imberti, and S.E. McLain, "Amphipathic Solvation of Indole: Implications for the Role of Tryptophan in Membrane Proteins.", *The Journal of Physical Chemistry B* **119**(19), pp. 5979–5987 (2015).

[119] R. Mancinelli, A. Botti, F. Bruni, M.A. Ricci, and A.K. Soper, "Perturbation of water structure due to monovalent ions in solution.", *Physical Chemistry Chemical Physics* **9**(23), pp. 2959–2967 (2007).

[120] A.K. Soper and K. Weckström, "Ion solvation and water structure in potassium halide aqueous solutions.", *Biophysical Chemistry* **124**(3), pp. 180–191 (2006).

[121] S. Diaz-Moreno, D.T. Bowron, and J. Evans, "Structural investigation of the bridged activated complex in the reaction between hexachloroiridate(iv) and pentacyanocobaltate(ii)", *Dalton Transactions* **23**, pp. 3814 (2005).

[122] S. Díaz-Moreno, S. Ramos, and D.T. Bowron, "Solvation structure and ion complexation of La3+ in a 1 molal aqueous solution of lanthanum chloride.", *The Journal of Physical Chemistry. A* **115**(24), pp. 6575–6781 (2011).

[123] D.T. Bowron and S. Diaz-Moreno, "Local structure refinement of disordered material models: ion pairing and structure in YCl3 aqueous solutions.", *The Journal of Physical Chemistry B* **111**(39), pp. 11393–11399 (2007).

[124] D.T. Bowron, E.C. Beret, E. Martin-Zamora, A.K. Soper, and E.S. Marcos, "Axial structure of the Pd(II) aqua ion in solution.", *Journal of the American Chemical Society* **134**(2), pp. 962–967 (2012).

[125] D.T. Bowron and S. Díaz-Moreno, "Solvent structure and the extended range hydration of Cr3+ in aqueous solution.", *The Journal of Physical Chemistry B* **113**(35), pp. 11858–11864 (2009).

[126] A.K. Soper, "The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa", *Chemical Physics* **258**(2-3), pp. 121–137 (2000).

[127] S.E. McLain, A.K. Soper, and A. Watts, "Structural studies on the hydration of L-glutamic acid in solution.", *The Journal of Physical Chemistry B* **110**(42), pp. 21251–21258 (2006).

[128] S.E. McLain, A.K. Soper, A.E. Terry, and A. Watts, "Structure and hydration of L-proline in aqueous solutions.", *The Journal of Physical Chemistry B* **111**(17), pp. 4568–4580 (2007).

[129] R.Z. Troitzsch, G.J. Martyna, S.E. McLain, A.K. Soper, and J. Crain, "Structure of aqueous proline via parallel tempering molecular dynamics and neutron diffraction.", *The Journal of Physical Chemistry B* **111**(28), pp. 8210–8222 (2007).

[130] S.E. McLain, A.K. Soper, and A. Watts, "Water structure around dipeptides in aqueous solutions.", *European Biophysics Journal* **37**(5), pp. 647–655 (2008).

[131] S.E. McLain, A.K. Soper, I. Daidone, J.C. Smith, and A. Watts, "Charge-based interactions between peptides observed as the dominant force for association in aqueous solution.", *Angewandte Chemie (International ed. in English)* **47**(47), pp. 9059–9062 (2008).

[132] S.E. Pagnotta, M.A. Ricci, F. Bruni, S.E. McLain, and S. Magazu, "Water structure around trehalose", *Chemical Physics* **345**(2-3), pp. 159–163 (2008).

[133] S.E. Pagnotta, S.E. McLain, A.K. Soper, F. Bruni, and M.A. Ricci, "Water and Trehalose: Flow Much Do They Interact with Each Other?", *The Journal of Physical Chemistry B* **114**(14), pp. 4904–4908 (2010).

[134] F. Foglia, M.J. Lawrence, C.D. Lorenz, and S.E. McLain, "On the hydration of the phosphocholine headgroup in aqueous solution.", *The Journal of Chemical Physics* **133**(14), pp. 145103 (2010).

[135] A.P. Dabkowska, F. Foglia, M.J. Lawrence, C.D. Lorenz, and S.E. McLain, "On the solvation structure of dimethylsulfoxide/water around the phosphatidylcholine head group in solution.", *The Journal of Chemical Physics* **135**(22), pp. 225105 (2011).

[136] R.J. Gillams, J.V. Busto, S. Busch, F.M. Goñi, C.D. Lorenz, and S.E. McLain, "Solvation and hydration of the ceramide headgroup in a non-polar solution.", *The Journal of Physical Chemistry B* **119**(1), pp. 128–139 (2015).

[137] E.C. Hulme, A.K. Soper, S.E. McLain, and J.L. Finney, "The hydration of the neurotransmitter acetylcholine in aqueous solution.", *Biophysical Journal* **91**(6), pp. 2371–2380 (2006).

[138] S.K. Callear, A.J. Johnston, S.E. McLain, and S. Imberti, "Conformation and interactions of dopamine hydrochloride in solution.", *The Journal of Chemical Physics* **142**(1), pp. 014502 (2015).

[139] S. Busch, C.D. Bruce, C. Redfield, C.D. Lorenz, and S.E. McLain, "Water mediation is essential to nucleation of β-turn formation in peptide folding motifs.", *Angewandte Chemie (International ed. in English)* **52**(49), pp. 13091–13095 (2013).

[140] S. Busch, L.C. Pardo, W.B. O'Dell, C.D. Bruce, C.D. Lorenz, and S.E. McLain, "On the structure of water and chloride ion interactions with a peptide backbone in solution.", *Physical Chemistry Chemical Physics* **15**(48), pp. 21023–21033 (2013).

[141] E. Scoppola, A. Sodo, S.E. McLain, M.A. Ricci, and F. Bruni, "Water-peptide site-specific interactions: a structural study on the hydration of glutathione.", *Biophysical Journal* **106**(8), pp. 1701–1709 (2014).

[142] R. Hargreaves, D.T. Bowron, and K. Edler, "Atomistic Structure of a Micelle in Solution Determined by Wide Q-Range Neutron Diffraction", *Journal of the American Chemical Society* **133**(41), pp. 16524–16536 (2011).

[143] T. Head-Gordon, J.M. Sorenson, A. Pertsemlidis, and R.M. Glaeser, "Differences in hydration structure near hydrophobic and hydrophilic amino acids.", *Biophysical Journal* **73**(4), pp. 2106–2115 (1997).

[144] J.M. Sorenson, G. Hura, A.K. Soper, Al. Pertsemlidis, and T. Head-Gordon, "Determining the Role of Hydration Forces in Protein Folding", *The Journal of Physical Chemistry B* **103**(26), pp. 5413–5426 (1999).

[145] A. Pertsemlidis, A.K. Soper, J.M. Sorenson, and T. Head-Gordon, "Evidence for microscopic, long-range hydration forces for a hydrophobic amino acid.", *Proceedings of the National Academy of Sciences of the United States of America* **96**(2), pp. 481–486 (1999).

[146] S. Busch, C.D. Lorenz, J. Taylor, L.C. Pardo, and S.E. McLain, "Short-Range Interactions of Concentrated Proline in Aqueous Solution.", *The Journal of Physical Chemistry B* **118**(49), pp. 14267–14277 (2014).

[147] I. Lobo, "Biological Complexity and Integrative Levels of Organization", *Nature Education* **1**(1), pp. 141 (2008).

[148] W-J. Lee, R.A. Hawkins, J.R. Vina, and D.R. Peterson, "Glutamine transport by the blood-brain barrier: a possible mechanism for nitrogen removal", *American Journal of Physiology - Cell Physiology* **274**(4), pp. C1101–1107 (1998).

[149] P. Newsholme, "Why Is L-Glutamine Metabolism Important to Cells of the Immune System in Health, Postinjury, Surgery or Infection?", *The Journal of Nutrition* **131**(9), pp. 2515S–2522S (2001).

[150] X. Li, C. Lin, and P.B. O'Connor, "Glutamine deamidation: differentiation of glutamic acid and gamma-glutamic acid in peptides by electron capture dissociation.", *Analytical Chemistry* **82**(9), pp. 3606–3615 (2010).

[151] L. Cruzeiro, "Why are proteins with glutamine- and asparagine-rich regions associated with protein misfolding diseases?", *Journal of Physics: Condensed Matter* **17**(50), pp. 7833–7844 (2005).

[152] B.J. Stapley and A.J. Doig, "Hydrogen bonding interactions between glutamine

and asparagine in alpha-helical peptides.", *The Journal of Molecular Biology* **272**(3), pp. 465–473 (1997).

[153] N.G. Faux, S.P. Bottomley, A.M. Lesk, J.A. Irving, J.R. Morrison, M.G. de la Banda, and J.C. Whisstock, "Functional insights from the distribution and role of homopeptide repeat-containing proteins.", *Genome Research* **15**(4), pp. 5375–51 (2005).

[154] M.F. Perutz, "Glutamine repeats and neurodegenerative diseases: molecular aspects", *Trends in Biochemical Sciences* **24**(2), pp. 58–63 (1999).

[155] T. Takahashi, S. Katada, and O. Onodera, "Polyglutamine Diseases: Where Does Toxicity Come From? What Is Toxicity? Where Are We Going?", *The Journal of Molecular Cell Biology* **2**(4), pp. 1801–91 (2010).

[156] R.H. Lathrop, M. Casale, D.J. Tobias, J.L. Marsh, and L.M. Thompson, "Modeling protein homopolymeric repeats: possible polyglutamine structural motifs for Huntington's disease.", *Proceedings of the International Conference on Intelligent Systems for Molecular Biology* **6**, pp. 105–114 (1998).

[157] M.F. Perutz, T. Johnson, M. Suzuki, and J.T. Finch, "Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases.", *Proceedings of the National Academy of Sciences of the United States of America* **91**(12), pp. 5355–5358 (1994).

[158] A.T. Petkova, G. Buntkowsky, F. Dyda, R.D. Leapman, W-M. Yau, and R. Tycko, "Solid state NMR reveals a pH-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide.", *The Journal of Molecular Biology* **335**(1), pp. 247–260 (2004).

[159] E.L. Altschuler, N.V. Hud, J.A. Mazrimas, and B. Rupp, "Random coil conformation for extended polyglutamine stretches in aqueous soluble monomeric peptides", *The Journal of Peptide Research* **50**(1), pp. 73–75 (2009).

[160] M. Tanaka, I. Morishima, T. Akagi, T. Hashikawa, and N. Nukina, "Intra- and intermolecular beta-pleated sheet formation in glutamine-repeat inserted myoglobin

as a model for polyglutamine diseases.", *The Journal of Biological Chemistry* **276**(48), pp. 45470–45775 (2001).

[161] A K Thakur and R Wetzel, "Mutational analysis of the structural organization of polyglutamine aggregates", *Proceedings of the National Academy of Sciences of the United States of America* **99**(26), pp. 17014–17019 (2002).

[162] M.F. Perutz, J.T. Finch, J. Berriman, and A. Lesk, "Amyloid fibers are water-filled nanotubes.", *Proceedings of the National Academy of Sciences of the United States of America* **99**(8), pp. 5591–5595 (2002).

[163] N.F. Bence, R.M. Sampat, and R.R. Kopito, "Impairment of the ubiquitin-proteasome system by protein aggregation.", *Science* **292**(5521), pp. 1552–1555 (2001).

[164] T.E. Williamson, A. Vitalis, S.L. Crick, and R.V. Pappu, "Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin.", *The Journal of Molecular bBology* **396**(5), pp. 1295–1309 (2010).

[165] A. von Mikecz, "PolyQ fibrillation in the cell nucleus: who's bad?", *Trends in Cell Biology* **19**(12), pp. 685–691 (2009).

[166] L Dougan, J Y Li, C L Badilla, B J Berne, and J M Fernandez, "Single homopolypeptide chains collapse into mechanically rigid conformations", *Proceedings of the National Academy of Sciences of the United States of America* **106**(31), pp. 12605–12610 (2009).

[167] C.B. Stanley, T. Perevozchikova, and V. Berthelier, "Structural Formation of Huntingtin Exon 1 Aggregates Probed by Small-Angle Neutron Scattering", *Biophysical Journal* **100**(10), pp. 2504–2512 (2011).

[168] S.L. Crick, M. Jayaraman, C. Frieden, R. Wetzel, and R.V. Pappu, "Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions", *Proceedings of the National Academy of Sciences of the United States of America* **103**(45), pp. 16764–16769 (2006).

[169] Q. Yu, X. Ma, and L. Xu, "Solubility, dissolution enthalpy and entropy of l-glutamine in mixed solvents of ethanol+water and acetone+water", *Thermochimica Acta* **558**, pp. 6–9 (2013).

[170] A. Busick, *Synthesis and Analysis of the Antimicrobial Peptoid N-LfB6*, PhD thesis University of Arkansas (2011).

[171] G.J. Boks, *Structure and Conformational Behaviour of Peptoid Peptidomimetics: Towards an Understanding of the Mechanism of Action of Substance P and NK1 Antagonists*, PhD thesis Universiteit Utrecht (1997).

[172] S.L. Miller, H.C. Urey, and J. Oró, "Origin of organic compounds on the primitive earth and in meteorites", *Journal of Molecular Evolution* **9**(1), pp. 59–72 (1976).

[173] H.J. Issaq and T.D. Veenstra, "Is sarcosine a biomarker for prostate cancer?", *Journal of separation science* **34**(24), pp. 3619–3621 (2011).

[174] S. de Vogel, A. Ulvik, K. Meyer, P.M. Ueland, O. Nygård, S.E. Vollset, G.S. Tell, J.F. Gregory, S. Tretli, and T. Bjørge, "Sarcosine and other metabolites along the choline oxidation pathway in relation to prostate cancer - a large nested case-control study within the JANUS cohort in Norway.", *International Journal of Cancer* **134**(1), pp. 197–206 (2014).

[175] L. Pradeep and J.B. Udgaonkar, "Osmolytes induce structure in an early intermediate on the folding pathway of barstar.", *The Journal of Biological Chemistry* **279**(39), pp. 40303–40313 (2004).

[176] A.M. Rosales, *Nanostructure Control of Biologically Inspired Polymers*, PhD thesis University of California Berkeley (2013).

[177] H.C. Ohanian, *Physics*, W. W. Norton and Company New York, NY 2nd expand edition (1999).

[178] W.H. Bragg and W.L. Bragg, "The Reflection of X-rays by Crystals", *Proceedings of the Royal Society of London Series A-Containing Papers of a Mathematical and Physical Character* **88**(604), pp. 428 (1913).

[179] E. Hecht, *Optics*, Addison-Wesley San Francisco 4th edition (2002).

[180] L. de Broglie, "Waves and Quanta", *Nature* **112**, pp. 540 (1923).

[181] P.P. Ewald, "Introduction to the dynamical theory of X-ray diffraction", *Acta Crystallographica Section A* **25**(1), pp. 103–108 (1969).

[182] T. Young, *A Course of Lectures on Natural Philosophy and the Mechanical Arts* volume 2, Johnson (1807).

[183] J Chadwick, "Possible Existence of a Neutron", *Nature* **129**, pp. 312 (1932).

[184] J. Chadwick, "The Existence of a Neutron", *Proceedings of the Royal Society of London Series A-Containing Papers of a Mathematical and Physical Character* **136**(830), pp. 692–708 (1932).

[185] W H McMaster, N Kerr Del Grande, J H Mallett, and J H Hubbell, *Compilation of X-ray Cross Sections*, U.S. Dept. of Commerce (1969).

[186] V.F. Sears, "Neutron scattering lengths and cross sections", *Neutron News* **3**, pp. 26–37 (1992).

[187] C. Benmore and A.K. Soper, "The SANDALS manual: A Guide to performing experiments on the small angle neutron diffractometer for amorphous and liquid samples at ISIS", *RAL Report* **RAL-TR-98-**, pp. 39 (1998).

[188] J.J. Towey, *A Structural Approach to Reveal the Croprotective Action of Glycerol*, PhD thesis University of Leeds (2013).

[189] A.K. Soper, "GudrunN and GudrunX: Programs for Correcting Raw Neutron and X-ray Diffraction Data to Differential Scattering Cross Section", *RAL Report* **RAL-TR-201** (2011).

[190] A.K. Soper, S. Howells, and A.C. Hannon, "ATLAS - Analysis of Time-of-Flight Diffraction Data from Liquid and Amorphous Samples", *RAL Report* **RAL-89-046** (1989).

[191] A.K. Soper, "Multiple-Scattering from an Infinite-Plane Slab", *Nuclear Instruments and Methods in Physics Research* **212**(1-3), pp. 337–347 (1983).

[192] A.K. Soper, "Empirical potential Monte Carlo simulation of fluid structure", *Chemical Physics* **202**(2-3), pp. 295–306 (1996).

[193] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press San Diego 2nd edition (2002).

[194] G. Hummer, D.M. Soumpasis, and M. Neumann, "Computer-Simulation of Aqueous Na-Cl Electrolytes", *Journal of Physics: Condensed Matter* **6**, pp. A141 – A144 (1994).

[195] C. G. Gray and Keith E. Gubbins, *Theory of Molecular Fluids: Fundamentals*, Oxford University Press (1984).

[196] A.K. Soper, "Empirical Potential Structure Refinement - EPSRshell: A User's Guide" (2011).

[197] I.M. Svishchev and P.G. Kusalik, "Roto-translational motion in liquid water and its structural implication", *Chemical Physics Letters* **215**(6), pp. 596–600 (1993).

[198] R. Pynn, "Neutron Scattering: A Non-destructive Microscope for Seeing Inside Matter", In L. Liang, R. Rinaldi, and H. Schober, editors, *Neutron Applications in Earth, Energy and Environmental Sciences* Neutron Scattering Applications and Techniques chapter 2, , pp. 15–36. Springer US Boston, MA (2009).

[199] S.M. King, "Small-Angle Neutron Scattering", In J.V. Pethrick, R.A and Dawkins, editor, *Modern Techniques for Polymer Characterisation* chapter 7. Wiley Salt Lake City (1999).

[200] H.C. van de Hulst, *Light Scattering by Small Particles*, John Wiley and Sons Ltd. New York (1957).

[201] O. Kratky and G. Porod, "Diffuse small-angle scattering of X-rays in colloid systems", *Journal of Colloid Science* **4**(1), pp. 35–70 (1949).

[202] A. Guinier and G. Fournet, *Small-angle scattering of X-rays*, Wiley New York (1955).

[203] J.S. Higgins and H. Benoît, *Polymers and Neutron Scattering*, Clarendon Press Oxford (1994).

[204] J.S. Pedersen and P. Schurtenberger, "Scattering Functions of Semiflexible Polymers with and without Excluded Volume Effects", *Macromolecules* **29**(23), pp. 7602–7612 (1996).

[205] S.R. Kline, "Reduction and analysis of SANS and USANS data using IGOR Pro", *Journal of Applied Crystallography* **39**(6), pp. 895–900 (2006).

[206] W-R. Chen, P.D. Butler, and L.J. Magid, "Incorporating intermicellar interactions in the fitting of SANS data from cationic wormlike micelles.", *Langmuir* **22**(15), pp. 6539–6548 (2006).

[207] R. Heenan, "THE FISH REFERENCE MANUAL (DATA FITTING PROGRAM FOR SMALL-ANGLE DIFFRACTION etc.)", *RAL Report* **RAL-89-129** (2005).

[208] P. Alina, G., Butler, P., Cho, J. , Doucet, M., and Kienzle, "SANS Analysis Software", *Unsubmitted Manuscript* (2015).

[209] G.F. Knoll, *Radiation Detection and Measurement*, Wiley and Sons Ltd. University of Michigan 3rd edition (2000).

[210] "The LOQ Beamline". http://www.isis.stfc.ac.uk/instruments/loq/documents/the-loq-beamline9961.jpg, *Accessed 01/02/2015.*

[211] R. Borsali and R. Pecora, editors, *Soft Matter Characterization*, Springer Netherlands Dordrecht (2008).

[212] R.W. Heenan, S.M. King, R. Osborn, and H.B. Stanley, "COLETTE USERS GUIDE", *RAL Report* **RAL-89-128** (1989).

[213] S.M. King and R. Heenan, "Using COLETTE", *RAL Report* **RAL-95-005** (2005).

[214] G.D. Wignall and F.S. Bates, "Absolute calibration of small-angle neutron scattering data", *Journal of Applied Crystallography* **20**(1), pp. 28–40 (1987).

[215] J. Janin, "Surface and inside volumes in globular proteins.", *Nature* **277**(5696), pp. 491–492 (1979).

[216] J. Kyte and R.F. Doolittle, "A simple method for displaying the hydropathic character of a protein", *The Journal of Molecular Biology* **157**(1), pp. 105–132 (1982).

[217] D. Eisenberg, "Three-dimensional structure of membrane and surface proteins.", *Annual Review of Biochemistry* **53**, pp. 595–623 (1984).

[218] G.D. Rose and R. Wolfenden, "Hydrogen bonding, hydrophobicity, packing, and protein folding.", *Annual Review of Biophysics and Biomolecular Structure* **22**, pp. 381–415 (1993).

[219] A. Natalello, A.M. Frana, A. Relini, A. Apicella, G. Invernizzi, C. Casari, A. Gliozzi, S.M. Doglia, P. Tortora, and M.E. Regonesi, "A Major Role for Side-Chain Polyglutamine Hydrogen Bonding in Irreversible Ataxin-3 Aggregation", *Plos One* **6**(4), pp. 1–10 (2011).

[220] L. Esposito, A. Paladino, C. Pedone, and L. Vitagliano, "Insights into structure, stability, and toxicity of monomeric and aggregated polyglutamine models from molecular dynamics simulations", *Biophysical Journal* **94**(10), pp. 4031–4040 (2008).

[221] T.F. Koetzle, M.N. Frey, M.S. Lehmann, and W.C. Hamilton, "Precision neutron diffraction structure determination of protein and nucleic acid components. XIII. Molecular and crystal structure of the amino acid L-glutamine", *Acta Crystallographica Section B* **29**(11), pp. 2571–2575 (1973).

[222] A. Wlodawer, W. Minor, Z. Dauter, and M. Jaskolski, "Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures.", *The Federation of European Biochemical Societies Journal* **275**(1), pp. 1–21 (2008).

[223] A K Soper, "The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure?", *ISRN Physical Chemisty* **2013**(279463) (2013).

[224] L.P. McIntosh, E. Brun, and L.E. Kay, "Stereospecific assignment of the NH2 resonances from the primary amides of asparagine and glutamine side chains in

isotopically labeled proteins.", *Journal of Biomolecular NMR* **9**(3), pp. 306–312 (1997).

[225] J. Stewart, "MOPAC", Technical report Fujitsu Limited Tokyo (1993).

[226] K.M. Tych, T. Hoffmann, M. Batchelor, M.L. Hughes, K.E. Kendrick, D.L. Walsh, M. Wilson, D.J. Brockwell, and L. Dougan, "Life in extreme environments: single molecule force spectroscopy as a tool to explore proteins from extremophilic organisms.", *Biochemical Society Transactions* **43**(2), pp. 179–85 (2015).

[227] NE Robinson and A Robinson, *Molecular clocks: deamidation of asparaginyl and glutaminyl residues in peptides and proteins*, Althouse Press London, ON (2004).

[228] L J Root and F H Stillinger, "SHORT-RANGE ORDER IN GLYCEROL - A MOLECULAR-DYNAMICS STUDY", *Journal of Chemical Physics* **90**(2), pp. 1200–1208 (1989).

[229] I. Brovchenko, A. Krukau, N. Smolin, A. Oleinikova, A. Geiger, and R. Winter, "Thermal breaking of spanning water networks in the hydration shell of proteins.", *The Journal of Chemical Physicshemical physics* **123**(22), pp. 224905 (2005).

[230] K. Xiong, D. Punihaole, and S. A. Asher, "UV resonance Raman spectroscopy monitors polyglutamine backbone and side chain hydrogen bonding and fibrillization.", *Biochemistry* **51**(29), pp. 5822–30 (2012).

[231] J.A. Plumley and J.J. Dannenberg, "The importance of hydrogen bonding between the glutamine side chains to the formation of amyloid VQIVYK parallel beta-sheets: an ONIOM DFT/AM1 study.", *Journal of the American Chemical Society* **132**(6), pp. 1758–1759 (2010).

[232] Y-U. Kwon and T. Kodadek, "Quantitative evaluation of the relative cell permeability of peptoids and peptides.", *Journal of the American Chemical Society* **129**(6), pp. 1508–1509 (2007).

[233] L.L. McKinney, E.H. Uhing, E.A. Setzkorn, and J.C. Cowan, "Cyanoethylation of Alpha Amino Acids. I. Monocyanoethyl Derivatives2", *Journal of the American Chemical Society* **72**(6), pp. 2599–2603 (1950).

[234] F.H.C. Stewart, "Use of N-2-cyanoethylglycine derivatives in the synthesis of peptides of N-2-carboxamidoethylglycine, an isomer of glutamine", *Australian Journal of Chemistry* **24**(6), pp. 1267–1275 (1971).

[235] W.L. Jorgensen and J. Gao, "Monte Carlo simulations of the hydration of ammonium and carboxylate ions", *The Journal of Physical Chemistry* **90**(10), pp. 2174–2182 (1986).

[236] L. Sherwood, H. Klandorf, and P. Yancey, *Animal Physiology: From Genes to Organisms*, Cengage Learning Boston, MA (2012).

[237] B. Ibarra-Molero, I.M. Plaza del Pino, B. Souhail, H.O. Hammou, and J.M. Sanchez-Ruiz, "The sarcosine effect on protein stability: a case of nonadditivity?", *Protein Science* **9**(4), pp. 820–826 (2000).

[238] L.M.F Holthauzen and D.W. Bolen, "Mixed osmolytes: the degree to which one osmolyte affects the protein stabilizing ability of another.", *Protein Science* **16**(2), pp. 293–298 (2007).

[239] H. Marshall, M. Venkat, N. Seng, J. Cahn, and D. H. Juers, "The use of trimethylamine N-oxide as a primary precipitating agent and related methylamine osmolytes as cryoprotective agents for macromolecular crystallography.", *Acta crystallographica. Section D, Biological crystallography* **68**(Pt 1), pp. 69–81 (2012).

[240] P. Venkatesu, M-J. Lee, and H-M. Lin, "Osmolyte Counteracts Urea-Induced Denaturation of α-Chymotrypsin", *The Journal of Physical Chemistry B* **113**(15), pp. 5327–5338 (2009).

[241] A. Mostad and S. Natarajan, "Crystal and molecular structure of sarcosine.", *Acta Chemica Scandinavica* **43**(10), pp. 1004–1006 (1989).

[242] A. Chaudhari, P.K. Sahu, and S-L. Lee, "Hydrogen bonding interaction in sarcosine-water complex using ab initio and DFT method", *International Journal of Quantum Chemistry* **101**(1), pp. 97–103 (2005).

[243] W. Klyne and V. Prelog, "Description of steric relationships across single bonds", *Experientia* **16**(12), pp. 521–523 (1960).

[244] G.A. Elgavish and J. Reuben, "Aqueous lanthanide shift reagents. 10. Proton and carbon-13 studies of the interaction of the aquoions with amino acids", *Journal of Magnetic Resonance* **42**(2), pp. 242–254 (1981).

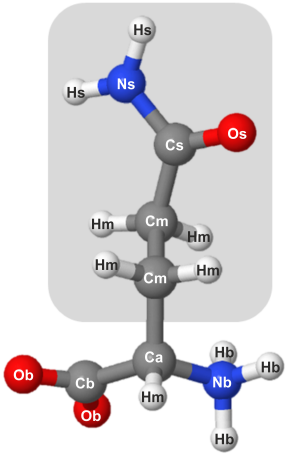
[245] A. Pertsemlidis, A.M. Saxena, A.K. Soper, T. Head-Gordon, and R.M. Glaeser, "Direct evidence for modified solvent structure within the hydration shell of a hydrophobic amino acid.", *Proceedings of the National Academy of Sciences of the United States of America* **93**(20), pp. 10769–10774 (1996).

[246] M.E. Johnson, C. Malardier-Jugroot, and T. Head-Gordon, "Effects of co-solvents on peptide hydration water structure and dynamics.", *Physical Chemistry Chemical Physics* **12**(2), pp. 393–405 (2010).

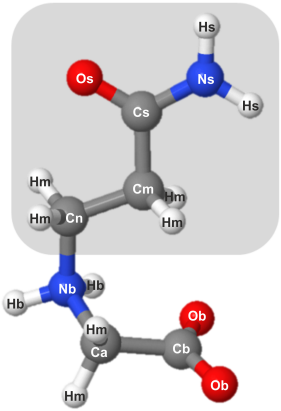
[247] C. Malardier-Jugroot, D.T. Bowron, A.K. Soper, M.E. Johnson, and T. Head-Gordon, "Structure and water dynamics of aqueous peptide solutions in the presence of co-solvents.", *Physical Chemistry Chemical Physics* **12**(2), pp. 382–392 (2010).

[248] A K Soper, "Joint structure refinement of x-ray and neutron diffraction data on disordered materials: application to liquid water", *Journal of Physics: Condensed Matter* **19**(33), pp. 335206 (2007).

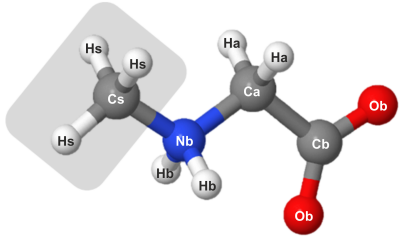
[249] I. Jelesarov and H.R. Bosshard, "Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition.", *Journal of Molecular Recognition* **12**(1), pp. 3–18 (1999).



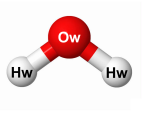
L-glutamine (Gln)



Glutamine imino acid (Ngln)



Sarcosine (Sar)



Water