Intermolecular structure and dynamics of aqueous *N*-methylacetamide

Susan Allison

Thesis submitted for the degree of Doctor of Philosophy

School of Physics
University of Edinburgh

2007

Abstract

The twin questions of how and why protein molecules fold into the specific topologies which enable them to fulfil their biological function have been the subject of continuous scientific investigation since the early twentieth century. Interactions between biological macromolecules and water are obviously crucial to both folding and function but attempts to gain understanding are impeded by the size and complexity of these systems. A useful approach is to consider much simpler model systems which capture some essential element of real biological systems but are experimentally and theoretically tractable.

N-methylacetamide (NMA) is a minimal model of the peptide linkage which forms the backbone of protein molecules. Its behaviour in aqueous solution therefore captures the important competition between peptide - peptide and peptide - water hydrogen bonds which arises in protein hydration. In this thesis aqueous NMA solutions are studied across the full concentration range using classical molecular dynamics simulation. This gives access to the complete spectrum of behaviour between the two important limiting cases of dilute NMA in water and, conversely, dilute water in NMA.

Water is now known to be an active player in biological interactions and the simple system studied here displays significant disruption of the structure and dynamics of pure water with the addition of only a small proportion of peptide groups. At dilute NMA concentrations water molecules continue to form system-size hydrogen bonded networks. Water molecules appear to optimise their local tetrahedral order by forming hydrogen bonds with a combination of NMA and water neighbours, rather than solely with members of their own species.

NMA molecules hydrogen bond through the amide and carbonyl groups to form linear and branched chains in both the pure liquid and in the aqueous solutions. In the NMA-rich region water molecules preferentially donate both hydrogens to chain-end or mid-chain carbonyl oxygens, forming bridges between NMA chains which resemble buried water configurations found in protein cavities. These bridge structures are thought to contribute to the observed slowing of the system dynamics at these concentrations.

The investigation of dynamics by classical simulation is complemented by a quasielastic neutron scattering study of NMA in its liquid and aqueous phases.

Declaration

I declare that this thesis has been composed by myself and has not been submitted in any previous application for a degree. The work reported within was carried out by myself unless otherwise stated.

Susan Allison
6th November 2007

In memory of Brian Duchart 1945 - 2007

Oh tell me fit was on yer road ye roarin Norland wind?
As ye cam blawin frae the land that's niver frae ma mind.
Ma feet they traivel England but I'm deein for the North
Ma man, I saw the siller tides rin up the Firth o' Forth.

Aye wind, I ken them weel eneuch an fine they fa and rise,
And fain I'd feel the creepin mist on yonder shore that lies.
But tell me as ye pass them by fit saw ye on the way?
Ma man, I rocked the rovin gulls that sail abin the Tay.

Bit saw ye naethin leein wind afore ye come tae Fife?
For there's muckle lyin 'yont the Tay that's mair tae me nor life.
Ma man, I swept the Angus braes that ye hivna trod for years.
Oh wind, forgie a hameless loon that canna see for tears.

And far abin the Angus straths I saw the wild geese flee,
A lang, lang skein o beatin wings wi their heids toward the sea,
And aye their cryin voices trailed ahint them on the air.
Oh wind, hae mercy, haud your wheesht for I daurna listen mair.

Violet Jacob, 1915

Acknowledgements

First and foremost I'd like to thank my supervisor Simon Bates for the impressive degree of support and enthusiasm he has given me through all the ups and downs of my PhD.

My second supervisor Jason Crain, Glenn Martyna and Troy Whitfield of IBM, as well as Raph Troitzsch have all been very generous with their advice and expertise. Without the help of Felix Fernandez-Alonso of ISIS the QENS chapter could never have been attempted far less written. His advice and support have been invaluable. Thanks must also go to Hugh Vass who carried out the sample preparation for the QENS experiment.

I couldn't have had a better bunch of officemates, so heartfelt thanks to all occupants of Room 4305/6, past and present. The last few years would have been much less fun without all my friends in the Physics department. Special thanks are due to Rowan for his company on our shared MPhys and PhD journey, Ben for his constant friendship since the days of second year Mathematical Physics, Eva for all her support and girlie chats, Marco for his empathetic emails during the final stages of thesis writing, Rik for sustaining my links with the yummicks and Lucio & Fede for letting me stay in their spare room when my thesis submission date over-ran!

Despite the PhD I've managed to spend most weekends in the Highlands: the biggest thanks are due to Jamie for all the ideas, adventures and for keeping me sane despite myself. Finally, there's a small but significant group of people who have been there all along, picked up the pieces time after time, and somehow still love me: I couldn't have a better family, thank you!

Contents

Chapter 1

Introduction

Proteins, quite simply, make us who we are. Protein functions include acting as structural building blocks, catalysing chemical reactions or defending our bodies from attack by alien molecules. These functionalities are impaired if the shape of the protein molecule is changed by heating, cooling or dehydration. An example of this is the everyday task of cooking an egg: at room temperature the long protein chains of raw egg are curled up into small balls. Heating agitates the molecules, breaking up the weak associations holding the protein balls in shape. As individual protein chains uncurl they come into contact with other chains and inter-chain bonds form creating a network with the rubbery texture of fried egg. Simply changing the shape of the egg protein molecule results in a substance which is completely different in appearance and behaviour to the original raw egg. Sometimes the proteins in our bodies do not fold into the correct shape and consequently function differently, causing diseases such as Alzheimers or Creutzfeld Jakob disease. Understanding why and how protein molecules fold into the shapes that they do is therefore of prime importance in progressing towards treatment and prevention of these currently incurable diseases.

A molecular-level description of protein-protein and protein-water interactions is crucial for understanding protein folding and function. The most exact first principles or *ab initio* techniques use quantum mechanics to calculate the electronic structure around atomic nuclei. This is computationally very expensive and so can only be used to treat small systems on the order of tens or exceptionally hundreds of atoms. In classical or molecular mechanics simulations the force on a single atom is the sum of its pairwise interactions with other atoms. The parameters describing these interactions are empirical: determined either from experiment or from more exact *ab initio* calculations. Classical simulations are much less computationally expensive and can easily be used to simulate systems containing thousands of atoms on a desktop computer. Hybrid quantum mechanics/molecular mechanics (QM/MM) schemes are sometimes used to

try and capture the detail of *ab initio* simulation whilst retaining some of the speed and large system capacity of classical simulation. While current supercomputers struggle to calculate many body interactions for tens of thousands of atoms, nature does not find it such a problem: the smallest protein molecules may contain only a few hundred atoms but many are several orders of magnitude larger. When one considers the number of atoms required for full simulations of macromolecules interacting in their biological environment the system size runs into hundreds of thousands or more likely millions of atoms. By stretching modern computational techniques to their utmost limit a system containing one million atoms was simulated in 2006[1]. This was the first simulation of a complete functioning organism - a satellite virus which is one of the smallest possible reproducing units found in nature. Researchers often circumvent the size problem by grouping atoms together in clusters to be modelled as a single particle or by treating solvents implicitly. While these techniques make it possible to both ask and answer many fundamental questions there is obviously some loss of realism. An alternative method is to pick a model system which mimics some aspect of the behaviour of large biomolecules but is much smaller and thus poses less of a challenge to current computing capabilities.

The molecule *N*-methylacetamide (NMA) is the minimal model of the peptide linkage in proteins and its behaviour captures the essential elements of protein - protein and protein - water interactions whilst its structural simplicity renders it experimentally, computationally and theoretically tractable. In this thesis the structure and dynamics of aqueous NMA solutions at a biologically relevant temperature of 308K (35°C) are investigated using all-atom classical molecular dynamics simulation across the full concentration range.

Molecular mechanics simulations such as those presented in this thesis explore atomic and molecular length scales for time periods on the order of pico or nanoseconds. These length and time scales are the same as can be accessed by neutron scattering experiments. Considering the same system from two such different yet complementary perspectives gives a deeper level of understanding than can be obtained from either individual technique. Quasielastic neutron scattering (QENS) is a method which can be used to investigate the stochastic dynamics of atoms and molecules. In this thesis QENS results for liquid NMA are reported and the extracted self diffusion coefficients are compared with simulation data. The simulated self diffusion coefficient of water in aqueous NMA is also compared with the result of a preliminary experiment on aqueous deuterated NMA at a concentration of $X_{dNMA} = 0.2$ at 308K.

1.1 Thesis Layout

Chapter 2 begins with an introduction to water, hydrogen bonding and protein structure. Current understanding of the aqueous NMA model system from experimental and simulation techniques is reviewed.

In Chapter 3 the classical molecular dynamics simulation method is described, followed by a description of the forcefields used to model NMA and water. The methods used to analyse simulation trajectories are also explained in this chapter.

The results of water simulations are presented in Chapter 4. The TIP3P, SPC/E and F3C models are compared and their intermolecular structure and dynamics determined so as to enable comparison with the results for the aqueous solutions which follow in later chapters.

Chapter 5 presents results of simulations of liquid NMA at 308K, 373K and 458K.

The results of simulations of aqueous NMA are presented in Chapter 6 and compared with those of the pure components previously reported in Chapters 4 and 5.

Quasielastic neutron scattering (QENS) experiments on NMA were undertaken at ISIS, the UK neutron scattering facility in Oxfordshire and Chapter 7 gives a brief introduction to this experimental technique followed by the relevant theory, experimental method, results and discussion.

Chapters 4, 5, 6 and 7 both present and discuss the computational or experimental results. In Chapter 8, key observations from each of the preceeding chapters are drawn together and avenues for future work are suggested.

Chapter 2

Background

Water cannot be regarded as an inert blank canvas for biological processes, instead it is an active player and may be key to understanding the connection between the structure and function of proteins. This chapter begins with an introduction to this "twenty-first amino acid"[2, 3]. True amino acids are the chemical building blocks of proteins whose formation and interactions are also discussed. This motivates the introduction of the model system approach and the chapter concludes with a review of the current understanding of a particular model system: aqueous *N*-methylacetamide.

2.1 Water

Water is ubiquitous: a three atom molecule, its apparent simplicity belied by its complex and unusual behaviour. Consequently it is probably the most studied molecule on the planet, yet despite this there remain many aspects of its behaviour which are poorly understood. What other molecule merits a biography[4], a seven volume treatise[5] and even a continually updated website[6]? An individual water molecule is made up of two hydrogen atoms, each covalently bonded to an oxygen atom as shown in Figure 2.1. A covalent bond is formed when two atoms share a pair of electrons such that the shared electrons are localised between the two nuclei rather than remaining associated with one nuclei or the other. In water, the electrons participating in the covalent bonds are not shared equally between the bonded nuclei due to the large disparity between the charge of the oxygen nucleus (+8e) and that of the hydrogen nucleus (+1e). The shared electrons are attracted towards the oxygen nucleus and it becomes slightly negatively charged by about -0.8e, leaving the hydrogen atoms correspondingly positively charged. Hence although its net charge is zero, water has an uneven distribution of charge density giving it an electric dipole moment of 1.85D. Molecules which have a permanent dipole moment are known as polar molecules.

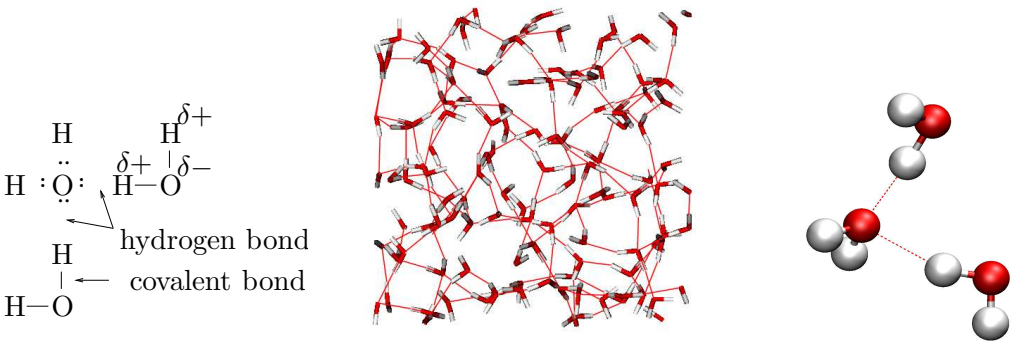

Figure 2.1: The left hand picture includes a Lewis diagram where electrons are represented as dots. An electron pair is shared between hydrogen and the oxygen atoms to form each covalent bond. The oxygen atom pulls these shared electrons towards itself becoming slightly negatively charged. Positively charged hydrogen atoms in adjacent molecules form hydrogen bonds with the lone pairs of electrons on the oxygen atom. In liquid water these hydrogen bonds form a three-dimensional network as illustrated in the simulation snapshot. Hydrogen bonds between water molecules are shown as red dashed lines, oxygen atoms are coloured red, hydrogens white.

2.2 Polar and non-polar liquids

Polar molecules interact via electrostatic dipole-dipole interactions whereby the negative and postively charged regions of adjacent molecules attract one another. These attractions are relatively strong and for this reason the melting and boiling points of water and other polar molecules are elevated well above what would otherwise be expected for molecules of their size. Non-polar molecules such as methane have a uniform charge distribution and no electric dipole. However, even these species have momentary fluctuations in their charge distributions which cause instantaneous attraction between neighbouring molecules. These momentary forces are known as London dispersion or van der Waals interactions and are the reason why even the most non-polar molecules or noble gas atoms can condense from the gas phase to form a liquid or solid.

2.3 Hydrogen bonds

Due to its small size and correspondingly small charge, the hydrogen nucleus is particularly poor at retaining its fair share of electrons, especially when covalently bonded to large, electronegative atoms such as oxygen or nitrogen. The ensuing partial charges give rise to particularly strong dipole-dipole interactions. In water, positively charged hydrogens are attracted to the negatively charged oxygens of neighbouring molecules and these interactions have been described as “hydrogen bonds” since the early twentieth century. Despite widespread use of the term the hydrogen bond

interaction itself has remained a mysterious and somewhat intangible object which defies any single definition. Its very nature has been the subject of disagreement since 1935 when Linus Pauling proposed that it is partly covalent[7]. Since most of its properties can be explained on the basis of electrostatic interactions this idea remained controversial until 1999 when an X-ray scattering study of ice at the European Synchrotron Radiation Facility (ESRF) confirmed Pauling's original suspicions[8]. Mo *et al* subsequently calculated that the total hydrogen bonding energy of a water dimer was -3.84 kcal mol^{-1} of which 65% was due to pairwise additive electrostatic interactions and the remainder to non pairwise additive interactions: 24% to polarization (intramolecular electron delocalisation) and 11% to charge-transfer (intermolecular electron delocalisation)[9]. A 2002 review of the subject by Thomas Steiner explains that the modern concept of a hydrogen bond is much broader than the classical picture of a predominantly electrostatic interaction[10]. Hydrogen bonds exist which are almost entirely covalent in their properties while others are so weak as to be barely distinguishable from van der Waals interactions. Thus the range of bond dissociation energies theoretically calculated for gas phase dimers covers two orders of magnitude, from 0.2 to 40 kcal mol^{-1}. For comparison, normal covalent bonds have bond energies between 60 - 100 kcal mol^{-1} while van der Waals interactions are 0.1 - 0.2 kcal mol^{-1}[11].

2.3.1 A hydrogen bond definition?

Whatever its origin, for practical purposes it is necessary to have a definition of precisely when an interaction is and is not a hydrogen bond. Unfortunately, each "practical purpose" seems to have bred a different definition. For example, geometric criteria are usually applied in methods which obtain atomic coordinates such as diffraction experiments or computer simulations. On the other hand particular features of IR absorption spectra also "define" hydrogen bonds and it is not clear that these two definitions are consistent[10]. In general a hydrogen bond is a local interaction in which an $D-H$ chemical group acts as a proton donor to an acceptor atom A as illustrated in Figure 2.2. The key difference between van der Waals interactions is that in a hydrogen bond interaction the angle $D-H\cdots A$ is preferentially linear whereas van der Waals interactions are isotropic over the range 120 to 180°[12]. Interactions with lower angles are sterically constrained. In computer simulations either geometric or energetic criteria are used to define hydrogen bonds. The energy of a particular configuration can be calculated from the atomic positions and a configuration is considered to be hydrogen bonded if its energy is less than some cut-off value. This is often taken to be -10 kJ mol^{-1} for water[13]. Alternatively, geometric criteria impose constraints on some or all of the following: the donor - acceptor distance, the hydrogen - acceptor distance and

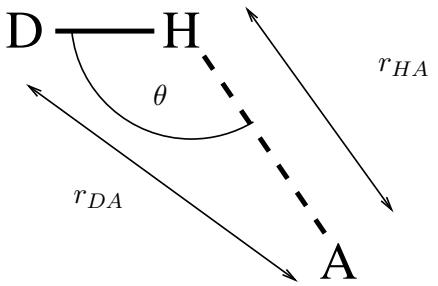

Figure 2.2: Hydrogen bond geometry: the donor group $D-H$ donates a proton to an electronegative acceptor atom A. The donor - acceptor r_{DA} and hydrogen - acceptor r_{HA} distances as well as the hydrogen bond angle θ are used to decide whether or not a particular configuration meets the requirements for a hydrogen bond interaction.

the hydrogen bond angle. It is common to use the first minimum of the relevant radial distribution function as the distance cutoff but the choice of an angular cutoff is more problematic. A recent paper on hydrogen bonding in water discusses the problems associated with different types of definition[14]. The acceptor - donor - hydrogen angle is also often used in geometric definitions and for water $O\cdots O-H < 30^\circ$ is a common choice[13].

2.3.2 Hydrogen bond cooperativity

In condensed matter hydrogen bonds are strongly influenced by their environment and bond strengths vary considerably between gas, liquid, solid and solution phases. This variation is partly due to the fact that hydrogen bond formation is cooperative: the strength of the primary hydrogen bond changes when a second hydrogen bond is formed between either the original donor or acceptor and a third hydrogen bonding group. The formation of the second hydrogen bond can either strengthen (positive cooperativity) or weaken (negative cooperativity) the original hydrogen bond. Water is a classic example of a hydrogen bonding fluid in which positively cooperative bonds between molecules form a network[15]. As hydrogen bond cooperativity is a many body effect the results of experimental gas phase measurements or *ab initio* computation of single molecules or small clusters cannot necessarily be extrapolated to apply to the liquid state. The very properties which make such systems complex and interesting in their behaviour also present a significant challenge to their study.

2.4 Proteins

The human body manufactures over 50,000 different types of proteins. Some of these such as keratin or collagen have a structural purpose, some are enzymes which catalyze chemical reactions in the body and others form part of the immune system

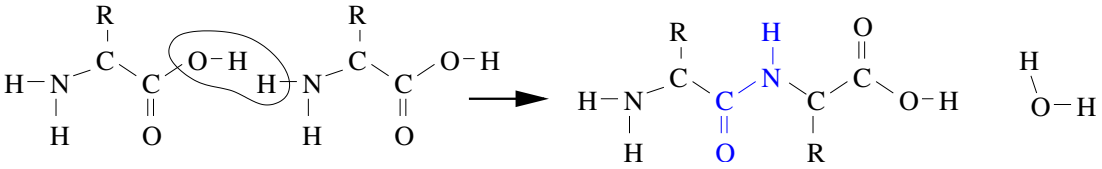

Figure 2.3: Two amino acids join together in a condensation reaction to form a peptide bond (highlighted in blue) with the elimination of water.

as antibodies. All are formed by the polymerisation of amino acids. There are twenty naturally occuring amino acids and the order in which these subunits are connected is fundamental to the subsequent shape and function of the protein molecule. Amino acids contain both amine ($-NH_2$) and carboxyl ($-COOH$) functional groups and these groups combine in a condensation reaction (so-called because it results in the elimination of a single water molecule) to form a peptide bond ($-C(=O)N(H)-$) as illustrated in Figure 2.3. The organic product of the condensation reaction is called a peptide and it too contains both the amine and carboxyl groups thus enabling its participation in further condensation reactions. The chain length increases with polymerisation and long chains are called polypeptides or proteins. Proteins vary greatly in length from several tens to tens of thousands of amino acid residues. The specific sequence of residues forming the protein chain is known as its primary structure. The peptide bond is the basic repeating unit of the protein "backbone" and each such bond contains the partially charged functional groups, $-C^{\delta+} = O^{\delta-}$ and $-N^{\delta-} - H^{\delta+}$. These groups can form hydrogen bonds and this leads to the occurence of characteristic local hydrogen bonded patterns or secondary structure elements. The most prevalent secondary structures are alpha helices and beta sheets, illustrated in Figure 2.4. Alpha helices are created when hydrogen bonds connect backbone $-NH$ and carbonyl groups separated by four residues. Beta sheets occur when hydrogen bonds form between parallel sections of protein backbone. The full three-dimensional structure of a protein molecule is called its tertiary structure, native or folded state. The tertiary structure is stabilised by a variety of interactions, including intramolecular hydrogen bonds, electrostatic interactions between charged side-chains, salt bridges, hydrophobic and hydrophilic interactions. The shape of a protein can be altered from that of its native state, or denatured, by changes in temperature, pressure, pH or dehydration. Denatured proteins lose the ability to perform their biological function and therefore it is clear that the form and function of proteins are interdependent. The loss of biological function upon dehydration indicates that although intra-molecular interactions between various regions of the protein chain are important they are not the full story: interactions between proteins and that ubiquitous solvent water also play a key role.

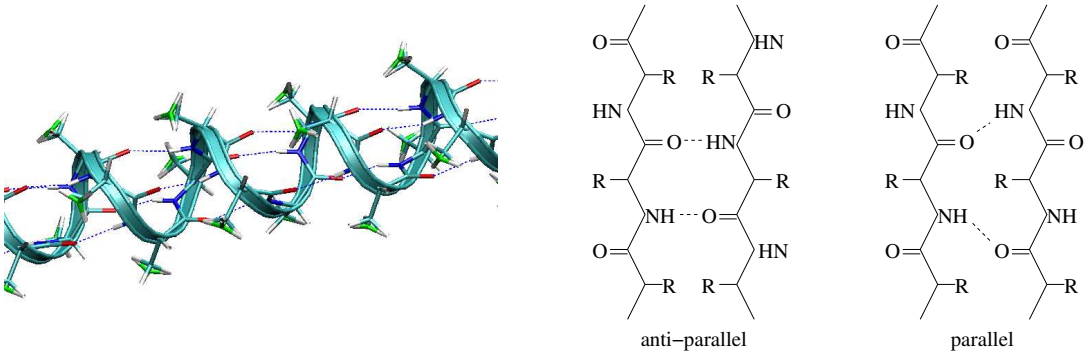

Figure 2.4: A section of an idealised alpha helix showing hydrogen bonds (dotted blue lines) joining adjacent twists of a single polypeptide backbone. On the right are schematics for the formation of both parallel and anti-parallel beta sheets where hydrogen bonds (dashed lines) connect adjacent sections of two separate polypeptide backbones.

2.4.1 Protein folding

The unfolded state was originally assumed to be a completely disordered random coil but is increasingly thought to contain elements of ordered structure[3]. The process by which a protein makes the transition from this state to its biologically active native state is by no means well understood despite the best efforts of several generations of research scientists. One of the first problems relating to protein folding was raised in the late 1960's by Levinthal who recognised that the size of individual protein molecules is such that were the folding process to proceed by checking all possible configurations it would in fact take longer than the age of the universe. He therefore proposed that local amino acid sequences rapidly form stable interactions which "speed and guide" the molecule towards its correct folded shape[16]. Current opinion agrees with this picture and suggests that the unfolded state is much less random than originally thought, containing local conformations which act as nucleation sites for folding which then proceeds via (often) well-defined intermediate states and by the use of molecular "chaperones"[17, 18]. The history of the scientific investigation of protein folding is an interesting and complex subject in itself. In 1936 Mirsky and Pauling were the first to suggest that the unique native structure of proteins was held together by hydrogen bonds[19]. This led to investigations of model compounds which found that inter-peptide hydrogen bonds were enthalpically favoured over peptide-water hydrogen bonds and thus concluded that inter-peptide hydrogen bonds contribute significantly to protein stability and drive protein folding[20]. However, this view was overturned in 1959 by Kauzmann who proposed that the driving force for protein folding was in fact hydrophobicity[21]. Supporting evidence for this new interpretation came from the experimental work of Klotz and Franzen who discovered that the dimerisation energy of a small model peptide in water was immeasurably small thus

casting a serious degree of doubt on the importance of inter-peptide hydrogen bonds to protein folding[22]. The amino acid residues making up the protein chain have varying degrees of solubility resulting in hydrophobic and hydrophilic regions of the protein chain. According to Kauzmann's model the protein chain folds so as to bury its hydrophobic regions in the centre of the structure while hydrophilic regions form the outer surface exposed to the aqueous environment. The dominance of hydrophobicity in protein folding was the prevailing viewpoint of the latter half of the twentieth century while the role of inter-peptide hydrogen bonding was questioned even as to whether such interactions destabilised rather than stabilised protein structure[23, 24]. Towards the end of the twentieth century the role of hydrogen bonding was re-evaluated with many researchers concluding that each intramolecular bond contributes a net stabilisation of 1 or 2 kcal mol^{-1} and thus hydrogen bonding must be of comparable importance to the hydrophobic effect[24, 25]. This view however is not universal[26] and consequently there has been continuing discussion as to whether hydrogen bonds promote or oppose protein folding. Recent experimental work suggests that the rapid collapse of polypeptide chains is at least in part due to intramolecular hydrogen bonds[27] and therefore it appears that the current view has now come full circle back to that of Pauling. In the hydrophobic paradigm folding is driven by the hydrophobicity of sidechains belonging to amino acid residues. A new paradigm has been proposed recently which inverts this view suggesting instead that hydrogen bonding of the protein backbone plays the dominant role[28]. It is clear that this subject is far from closed.

2.4.2 Protein hydration

Although the presence of water is obviously crucial to the hydrophobic effect almost all of the research effort over the latter half of the twentieth century concentrated on understanding protein-protein interactions. However a growing body of evidence suggests that water is not merely a background solvent but an integral player in the stability and interactions of biological macromolecules[2, 3]. Many proteins contain structural water molecules known as cavity or buried waters which can be identified crystallographically. The picture of protein hydration in the hydrophobic paradigm is of a "dry" hydrophobic interior stabilised by intra-molecular hydrogen bonds and sidechain packing with a few cavities containing bound water molecules. The exterior of the protein molecule is then made up of hydrophilic groups which hydrogen bond to water molecules thus creating a monolayer of bound surface or hydration water. Recent spectroscopic investigations of water dynamics have complicated this picture and indicate that the protein interior is much more "wet" than expected and that both the folded and unfolded states contain significant quantities of bound water[3]. Levy and Onuchic list the various contributions water makes to protein interactions:

it is important to structure, stability, dynamics and function; it mediates the collapse of the protein chain and the search for the native topology; it mediates interactions between binding partners in molecular recognition; it contributes to either entropic or enthalpic stabilisation[2]. Simultaneously with this recognition that water plays a key role in protein interactions has come the realisation that water itself is affected by interactions with proteins. While some consider water within the cell to be similar to bulk water, others believe that it is significantly modified by the presence of macromolecules and salts, possibly gel-like, inhomogeneous due to dissolved ions or vapourlike near hydrophobic surfaces[2]. Neutron and X-ray scattering techniques yield structural information on the location of bound water molecules and have shown that the first hydration shell around some proteins is approximately 10% more dense than bulk water[29]. This has been confirmed by computer simulations which indicate that about one third of the density increase arises from modifications of water structure and dynamics[30]. Merzel and Smith also show that local density variations in the first hydration shell can be accounted for by the topographic and electrostatic properties of the protein surface. Constantly evolving experimental techniques are now opening up access to the dynamics of these systems. Over 95% of the hydration layer has been shown to be no more than two-fold rotationally retarded compared to bulk water[31] although the rotational relaxation time of buried water can be 30 to 60ps, much longer than that of bulk water at approximately 1ps[3]. The residence time for water molecules belonging to the first hydration shell is on the order of 500ps while that of bound waters buried in protein cavities varies from nano to milli-seconds[2]. On the other hand it has been argued that applying the residence time concept to surface hydration water is problematic[32] and it is evident from the volume of publications that new experimental techniques are subject to interpretational controversy. In a 2002 review Gabel *et al* state that the major conclusion of neutron scattering studies of protein dynamics over the last decade has been that there exists a strong coupling between internal motions and the macromolecular environment[33]. Regardless of the exact interpretational details it is clear that a complete picture of protein folding and stability requires an explicit understanding of both the structure and dynamics of water - protein interactions.

2.5 Model systems

Protein molecules consisting of many hundreds or thousands of amino acid residues have an extremely large number of complex intra-molecular interactions, not to mention the many, many more inter-molecular interactions with their biological environment. Consequently a common approach has been to choose a model system which epitomises some element of these interactions but is much more tractable experimentally, computationally and theoretically. The peptide - peptide hydrogen

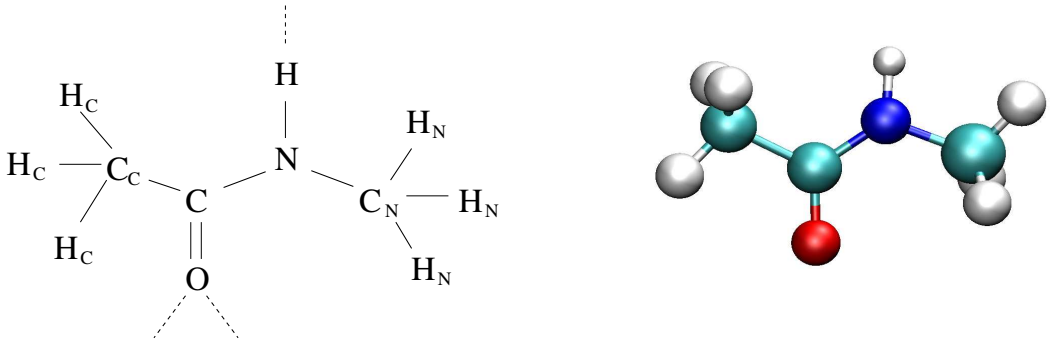

Figure 2.5: *N*-methylacetamide (NMA) : a peptide bond capped on either end by a methyl group. Atoms belonging to the methyl group bonded to the carbonyl carbon are labelled with the subscript *C* while those bonded to the nitrogen are labelled *N*. The three potential hydrogen bonding sites are marked with dashed lines. On the right is a visualisation of an NMA molecule from one of the simulations. The colour scheme is as follows: hydrogen, white; carbon, green; oxygen, red; nitrogen, blue.

bond which forms the secondary structure elements of alpha helices and beta sheets is one important element. Another is the competition between peptide - peptide and peptide - water hydrogen bonds which arises when proteins are in their biological (aqueous) environment. The molecule *N*-methylacetamide (NMA) is formed by capping either end of a peptide bond with a methyl group as illustrated in Figure 2.5. NMA molecules therefore form peptide - peptide hydrogen bonds and in aqueous solution a competition arises between NMA - NMA and NMA - water hydrogen bonds. The aqueous NMA system thus captures two crucial elements of protein interactions. Consequently this system has been the subject of many experimental and theoretical investigations and has played an important historical role in the elucidation of protein interactions.

The study of aqueous NMA is a logical extension of previous investigations into simple liquid mixtures by the Condensed Matter and Optics groups in the School of Physics at the University of Edinburgh. Over the past few years a combination of experimental and computational techniques have revealed unexpectedly complex behaviour in binary mixtures of water and methanol, a prototype amphiphile. These mixtures display molecular-scale segregation of the two components which was shown to explain the anomalous negative excess entropy of mixing characteristic of aqueous lower alcohols without resorting to water restructuring arguments which are unsupported by experiment[34, 35, 36, 37].

2.5.1 Key historical role of the aqueous NMA model system

Investigation of the aqueous NMA model system was key to the rejection of the inter-peptide hydrogen bond driven theory of protein folding in favour of Kauzmann's theory of hydrophobic hydration. Klotz and Franzen reported damningly that "in aqueous solution the interamide hydrogen bond of this model for a peptide has essentially no intrinsic strength ... other forces must be primarily responsible, therefore, for the maintenance of protein configuration in aqueous solution"[22]. More recently theoretical calculations of the free energy of forming an NMA dimer in water and in alkanes by Honig and co-workers indicate that the formation and burial within the protein interior of such hydrogen bonds opposes folding[26]. This result directly contradicts that of other researchers who find that intra-molecular hydrogen bonds stabilise protein structures[28]. Thus the aqueous NMA model system continues to provide relevant and thought-provoking insights into the interactions of real biological systems.

2.6 Current understanding of NMA and aqueous NMA

2.6.1 Crystal structure by neutron and x-ray diffraction

The crystal structure of NMA has been investigated by x-ray diffraction[38, 39] and single crystal neutron scattering[40]. NMA melts at 301K but a solid-solid phase

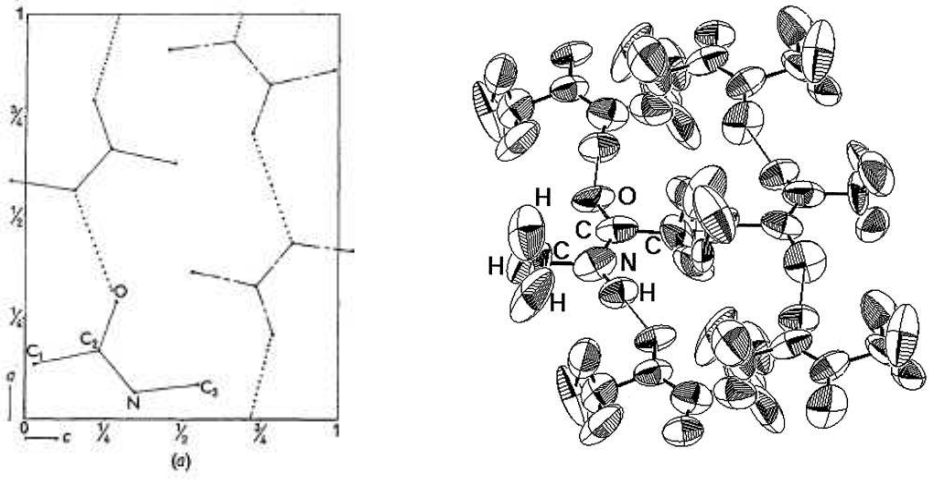

Figure 2.6: Diagram taken from Katz and Post's paper on the x-ray diffraction of crystalline NMA showing the linear chain structure below the solid-solid transition[38]. The right-hand figure is a thermal ellipsoid plot of the NMA unit cell at T= 250 K drawn at 40% probability from the single crystal neutron study by Eckert *et al*[40].

transition has been identified at 283K[38] or 274K[40]. Considerable dynamic disorder

has been reported from above this transition down to about 200K[40]. The $C-N-C-C$ backbone atoms are coplanar to within 0.015Å and in the crystal molecules form layers parallel to this plane with a separation of 3.26Å. Within the layers molecules are connected by hydrogen bonds between the NH and $C=O$ groups ($N\cdots O$ separation 2.825Å) to form linear chains as illustrated in Figure 2.6.

2.6.2 Theoretical and experimental studies of the vibrational spectra

The infrared (IR) vibrational spectra of NMA and other monosubstituted amides in their solid, liquid, gaseous and solution phases were first investigated by Miyazawa in 1956[41]. Subsequent investigations have employed inelastic neutron scattering (INS) [40, 42, 43], quasielastic neutron scattering (QENS) [44, 45], nuclear magnetic resonance (NMR) [46, 47, 48, 49], Raman spectroscopy [50, 47, 51, 52, 53, 54] and infrared spectroscopy [46, 55, 56, 47, 57, 58, 59, 60] including 2D correlation infrared spectroscopy [61, 62, 63, 64] and optically heterodyne-detected optical Kerr-effect spectroscopy[65]. The motivation for this sustained interest is that the vibrational spectra, particularly of liquid and solvated NMA, have been subject to almost continual re-evaluation[43]. An example of this is the work of Liu, Noda and Ozaki whose assignations of certain new spectral features to specific lengths of NMA oligomer[63, 64, 51] were later refuted by Huang as temperature induced shifts of the main NMA bands[62]. This thesis is not concerned with the analysis of vibrational spectra so a complete review of the many experimental and theoretical investigations in this field is outwith the scope of this chapter. However, a few pertinent conclusions can be drawn from this extensive body of work.

2.6.3 Strong intermolecular forces between NMA and water molecules

Firstly, the difficulty in interpretation of the vibrational spectra is proof of the fact that strong inter-molecular forces (hydrogen bonds) act between NMA molecules rendering simple theory inadequate in its explanation of the experimental results[43]. Chen, Schweitzer-Stenner *et al* have shown that NMA and its hydrogen bonded water molecules are vibrationally coupled and thus the problem is exacerbated when NMA is dissolved in water[50].

2.6.4 Cooperativity leads to chain formation

Secondly, hydrogen bonding in NMA has been found to be both positively and negatively cooperative. The formation of a hydrogen bond to the NH group increases the strength of an existing bond to the $C=O$ group and vice versa (positive cooperativity) [46, 66, 67, 57, 53] while the acceptance of a second hydrogen bond from water by the $C=O$ acceptor reduces the strength of its original hydrogen

bond (negative cooperativity)[67]. The interplay of these negative and positive cooperativities make it likely that NMA molecules form linear chains with both the $C = O$ and the NH groups forming a single hydrogen bond[67, 68]. Due to hydrogen bonding NMA molecules show strong self-aggregation in non-interacting solvents such as carbon tetrachloride [61, 56, 49]. In aqueous solution water has been shown to prevent NMA aggregation at low NMA concentrations but at higher NMA concentrations the positive cooperativity of NMA-NMA hydrogen bonds favours the existence of larger NMA oligomers over NMA-water dimers[57, 53]. Most studies have investigated the water-rich regime and there is relatively little information available on the NMA-rich region. A recent IR study by Czarnecki and Haufa finds that NMA self-association is little affected by the addition of small amounts of water[61].

2.6.5 Attempts to quantitatively define NMA self-association

Due to the complexity of the intermolecular interactions it is increasingly common for complementary simulation and experimental work to be undertaken and published together, particularly for the liquid and solution phases where interpretation of the experimental results is much more difficult. *Ab initio* Density Functional Theory (DFT) tends to be the theory of choice[57, 69]. Unfortunately this method is severely limited in the number of molecules it is possible to consider. Trabelsi found good agreement with X-ray diffraction data on liquid NMA when monomers, dimers and trimers were considered[69]. Schweitzer-Stenner assigned specific spectral bands to monomers, dimers and "longer" oligomers[53] although these assignments were later questioned by Huang who see no reason not to assume a broad distribution of oligomeric species[62]. Similarly, Ludwig and co-workers found better agreement with experiment when their calculations included all sizes of oligomer up to nonomers rather than the pentamers of their previous publication[48]. In their more recent work Huelsekopf and Ludwig report that hexamers and heptamers are the dominant species at low temperature being supplanted by smaller oligomers as the temperature is increased[70]. Thermodynamic calculations lead them to claim that oligomers longer than nine molecules (the maximum length included in their calculations) are not present. Other authors do not argue that such oligomers are not present but further DFT calculations indicate that the frequency shift induced by cooperative hydrogen bonding converges to a constant before the number of molecules in the oligomer reaches ten[47]. This may place a limit on the level of quantitative information which can be obtained from the vibrational spectra regarding the length of NMA oligomers and their population distribution. A calculation of the hydration enthalpy of NMA considers a much wider distribution of cluster sizes[66]. To summarise, although there have been many attempts made to obtain quantitative results, the publications to date do not present a unanimous

picture of the self-association of NMA in its liquid and aqueous phases.

2.6.6 Hydrogen bonding of methyl groups?

The situation is also unclear regarding whether or not the methyl groups participate in hydrogen bonding. In contrast to the amide modes, methyl group vibrational modes are essentially temperature-independent which is taken to indicate that they do not participate in hydrogen bonds[47, 48]. Conversely, frequency shifts of the methyl bands have been observed in the IR and NMR spectra of aqueous NMA on changing concentration, in particular, the methyl group bonded to the carbonyl group was found to be favoured in forming weak or improper hydrogen bonds[71, 72]. Additionally, *ab initio* calculations find that the lowest energy NMA dimer configuration involves a methyl hydrogen bond donated by this same group[73]. Classical and *ab initio* molecular dynamics simulations of liquid NMA suggest that links between molecules forming NMA chains as well as inter-chain associations reminiscent of beta sheets are stablised by improper hydrogen bonds between methyl groups[54, 74].

2.6.7 Relative strengths of intra- and inter-species hydrogen bonds

The aqueous NMA system has four types of strong hydrogen bond: NMA donating to NMA, NMA donating to water, water donating to NMA and water donating to water. Their relative strengths are accepted to be amide-amide > amide-water > water-water[68] although perturbation theory calculations concluded that the amide-amide and amide-water hydrogen bonds are approximately iso-energetic[75]. The two amide-water hydrogen bonds are different and the hydrogen bond formed by water molecules donating to the NMA carbonyl oxygen is calculated by DFT to be 50% stronger than that formed by water molecules accepting from the NH group[57].

2.6.8 Other theoretical studies

In addition to the results described above, NMA and aqueous NMA have been studied by a wide range of computational and theoretical techniques. As a model peptide NMA has been used in the optimisation of the CHARMM[76] and OPLS[68] forcefields. The majority of theoretical studies investigate small clusters of molecules in vacuum using *ab initio* techniques: isolated NMA [77], NMA dimers [78, 79, 73], short oligomers [70, 48, 80] and NMA-$n$$H_2O$ complexes [67, 81, 57, 82, 80]. Combined QM/MM techniques have also been used to model NMA-$n$$H_2O$ complexes in vacuum, implicit and explicit solvent [83, 84, 85]. The cis-trans isomerisation pathway has been studied using both *ab initio* and empirical potentials in vacuum and in explicit water solvent[86, 87]. Classical molecular dynamics simulations have concentrated on the pure liquid[54, 74] or the infinite dilution limit[88] although Zhang *et al* have recently

considered the full concentration range of the aqueous solution using the OPLS-AA and SPC forcefields[89].

2.6.9 Thermodynamic properties

The physical properties of aqueous NMA were studied in 1967 by Assarsson and Eirich[90] and more recently by Victor and Hazra[91]. There is good agreement between

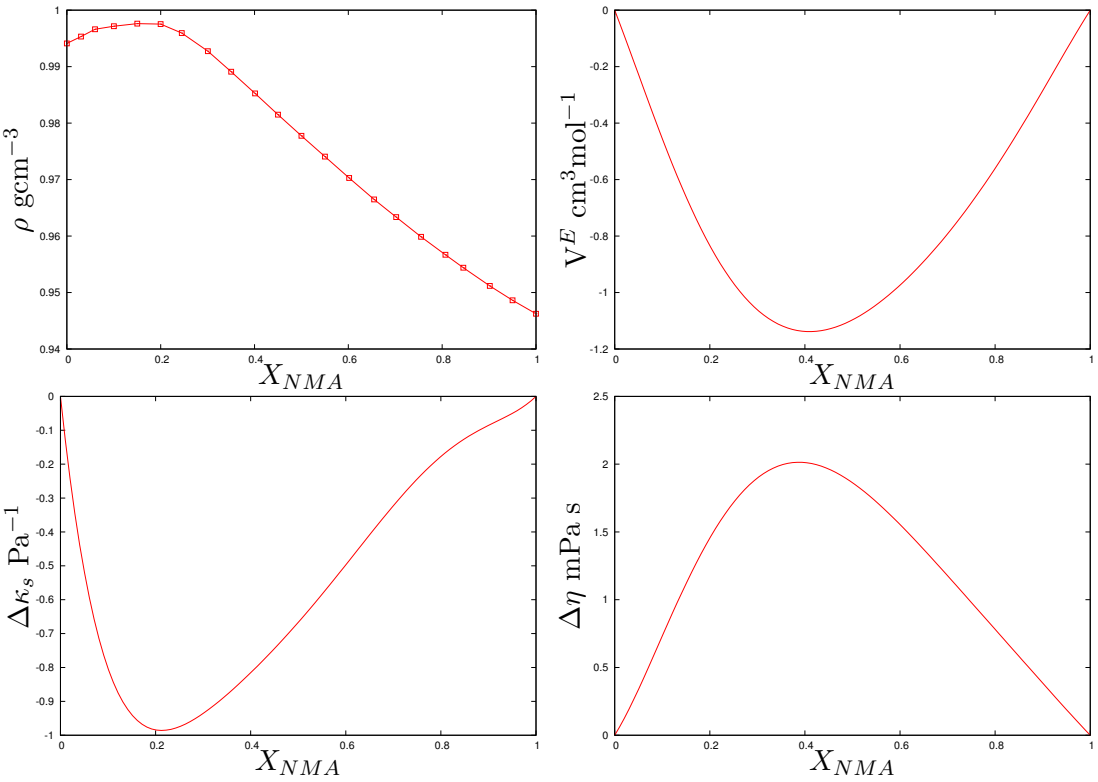

Figure 2.7: Densities (ρ), excess molar volumes (V^E), isentropic compressibility changes ($\Delta\kappa_s$) and viscosity deviations ($\Delta\eta$) for NMA-water mixtures at 308.15K[91].

the two sets of viscosity and density data. The first authors studied a variety of amides and found that viscosity maxima occur at specific mole fractions for the different homologous series. The mole fraction is one method of specifying the concentration of a mixture: X_A of species A in a mixture is defined as the number of molecules of that species, N_A, divided by the total number of molecules in the mixture, N.

$$X_A = \frac{N_A}{N} \tag{2.1}$$

For the N-monosubstituted amides, including NMA, the mole fraction of the viscosity maximum was $X_{amide} = 0.5$ which Assarsson and Eirich interpreted as the insertion of water molecules into existing amide hydrogen bonded associations at the spare lone electron pair of the carbonyl oxygen. The second authors measured densities,

viscosities and ultrasonic velocities across the full concentration range and calculated excess functions from their results. Excess functions are a measure of the deviation of a system's properties from those expected of an ideal mixture. An ideal mixture is defined as one in which enthalpic contributions to inter-species interactions are negligible, consequently the inter-species interaction $A - B$ between two species A and B is the same as the average of the $A - A$ and $B - B$ interactions. Therefore the driving force for mixing is the increase in entropy as the two species mingle and move towards a homogeneous solution state. The excess volume V^E, viscosity deviations $\Delta\eta$ and isentropic compressibility changes $\Delta\kappa_s$ calculated by Victor and Hazra as a function of NMA mole fraction (X_{NMA}) and plotted in Figure 2.7 are defined by

$$V^E = V - (V_{NMA}X_{NMA} + V_{H_2O}X_{H_2O}) \tag{2.2}$$

$$\Delta\eta = \eta - (\eta_{NMA}X_{NMA} + \eta_{H_2O}X_{H_2O}) \tag{2.3}$$

$$\Delta\kappa_s = \kappa_s - (\kappa_{sNMA}X_{NMA} + \kappa_{sH_2O}X_{H_2O}) \tag{2.4}$$

While the measured viscosity, η, of aqueous NMA has a maximum at approximately $X_{NMA} = 0.5$, $\Delta\eta$ has its maximum at $X_{NMA} = 0.4$, as does the excess volume. The density maximum occurs at a much lower mole fraction of $X_{NMA} = 0.15$, close to the minimum of the isentropic compressibility at just over $X_{NMA} = 0.2$. The NMA and aqueous NMA systems investigated in Chapters 5 and 6 were created at the experimental densities measured by Victor and Hazra, and were run at constant volume. Excess volumes, viscosities and isentropic compressibilities have not been calculated as part of this work on simulated aqueous NMA.

2.7 Summary

The excess functions of aqueous N-methylacetamide show that this solution deviates from the behaviour expected of an ideal mixture. The deviation is due to the formation of hydrogen bonds between water and the peptide group of the NMA molecule. Hydrogen bonds between water and peptide groups are prevalent in nature as over half of the hydration waters surrounding a protein molecule are hydrogen bonded to peptide groups belonging to the protein backbone[2]. The competition between peptide groups hydrogen bonding to other peptide groups, peptide groups hydrogen bonding to waters or water molecules hydrogen bonding to other waters is fundamental to protein stability, dynamics and thereby biological function. As proteins are extremely large and complex molecules it is common to study model systems to gain understanding of specific interactions. The aqueous NMA model system is ideally suited for elucidation of the complexities of hydrogen bonding between water and peptide groups since structurally

it is as simple as possible whilst retaining the peptide - water behaviour inherent in protein - water interactions. NMA molecules are capable of accepting two hydrogen bonds at the carbonyl oxygen and donating a single hydrogen bond from the $N-H$ group. If the carbonyl group accepts a single hydrogen bond then the formation of a second hydrogen bond by donation of the $N-H$ group is positively cooperative whilst the acceptance of a second hydrogen bond by the carbonyl group is negatively cooperative. This leads to the formation of linear hydrogen bonded chains in the solid and liquid phases. The population distribution of these chains in the liquid phase has not as yet been conclusively quantified. Most *ab initio* investigations of NMA and aqueous NMA only consider monomers or dimers and thus do not take into account the effect of hydrogen bond cooperativity. Relatively few investigations of the NMA-rich region of the aqueous phase have been undertaken although spectroscopic results show that the NMA structure remains dominant with water molecules incorporated into NMA chains by hydrogen bonding to carbonyl groups[61]. Water is increasingly recognised as an active player in protein folding and molecular recognition and the medium range structural ordering of water molecules around biological molecules is now known to be different from that of bulk water. Although aqueous NMA is a much studied system the focus has lain on NMA-NMA and NMA-water interactions and there is only indirect information available on the effect of NMA on water structure or dynamics. Since water is not merely a blank canvas it could be interesting to know how bulk water structure is perturbed by the presence of peptide groups in this important model system for protein-water interactions.

Chapter 3

Methodology

In this chapter the technical methods used to simulate mixtures of water and NMA are introduced. Firstly, the Classical Molecular Dynamics simulation method is outlined. This is followed by a more detailed description of the empirical forcefields used in the work carried out for this thesis. Finally the methods used to analyse the static and dynamic properties of the simulated systems are described.

3.1 Classical Molecular Dynamics

As the name suggests, Molecular Dynamics (MD) techniques simulate the movement of individual particles in a deterministic fashion. This is in contrast to Monte Carlo (MC) techniques where successive states are generated stochastically and accepted probabablistically. While both MD and MC methods yield information about instantaneous structure only MD can be used to investigate system dynamics. This section outlines technical aspects of the classical molecular dynamics or molecular mechanics method of computer simulation which are relevant to this thesis. It is not intended as a complete theoretical description and further information can be found from many other sources including references [92, 93, 94, 95].

During a classical molecular dynamics simulation a set of particles is allowed to evolve in time according to the laws of classical mechanics. Each individual particle trajectory is not particularly informative but the strength of the technique lies in the fact that the complete set of trajectories is distributed according to a statistical ensemble and thus the rules of statistical physics can be applied to obtain macroscopic thermodynamic averages from the microscopic behaviour of the simulated particles. Simulations for this thesis have been run using the DL_POLY package developed at Daresbury Laboratory by W. Smith, T.R. Forester and I.T. Todorov[96].

3.1.1 Statistical Ensembles

The principal tenet of statistical physics is that a macroscopic observable, A, can be calculated as the weighted average of the corresponding microscopic variables a_i

$$A = \langle a \rangle = \frac{1}{N} \sum_{i}^{N} p_i a_i \tag{3.1}$$

where p_i is the probability of the system being in microstate i of an ensemble of N microstates. In both molecular dynamics simulations and real life experiments mean values are usually obtained by averaging properties of the same system at different times rather than by averaging over the properties of an ensemble of systems. The assumption that time averaging is equivalent to ensemble averaging is known as the ergodic hypothesis. This statement holds for systems such as aqueous NMA but there are many examples of non-ergodic systems such as glasses.

The microcanonical ensemble: constant N, V, E

The time integration of Newton's equations of motion for a set of particles generates a succession of states for which particle number N, volume V, and energy E are conserved. This corresponds to the microcanonical ensemble and consequently this is often called the "natural" ensemble of Molecular Dynamics simulation.

The canonical ensemble: constant N, V, T

There are practically no real life experiments carried out under constant N, V, E conditions. It is much more common to hold temperature T or pressure P constant rather than energy. Consequently methods have been developed to allow simulation in ensembles other than the microcanonical. The method developed by Nosé and Hoover generates deterministic trajectories according to the canonical ensemble. This is achieved by modifying the equations of motion to introduce a frictional coefficient whose time evolution is driven by the difference between the instantaneous and desired kinetic energy[97, 98]. The time evolution is controlled by a specified time constant, usually in the range 0.5-2ps[96]. This is one of the thermostats implemented by DL_POLY and is used in this work for NVT simulations.

The isobaric - isothermal ensemble: constant N, P, T

In the isobaric-isothermal ensemble particle number N, pressure P and temperature T are conserved. Berendsen's method of controlling temperature or pressure can be likened to coupling the system to an external bath at constant temperature or pressure. The strength of the coupling in each case is governed by a specified relaxation

time constant. Static properties are adequately reproduced by coupling constants greater than 0.01ps but reliable information on the dynamics can only be achieved for coupling constants greater than 0.1ps[99]. The temperature or pressure of the system then evolves in time driven by the difference between its instantaneous and desired values. In the temperature case this is achieved by rescaling particle velocities at each timestep. Pressure is altered by isotropic rescaling of the size of the simulation box. One consequence of the rescaling process is that particle trajectories are no longer deterministic. The Berendsen thermostat and barostats are implemented in DL_POLY and used in this work for NPT simulations.

3.1.2 Molecular Dynamics Algorithm

A generic algorithm for a molecular dynamics simulation is as follows

1. define control parameters specifying the thermodynamic state point, length of simulation, sampling frequency etc.
2. select initial positions and velocities for all particles in the system
3. compute forces on all particles
4. integrate the equations of motion
5. compute system properties

Steps 3, 4 and 5 are repeated until the required simulation time specified by the control parameters has been reached. At the end of the simulation system properties calculated "on the fly" are averaged and printed. During the simulation information on particle positions or velocities and system properties may be written to an output file at specified time intervals and analysed after the simulation has completed. This is known as post-processing. Steps 3 and 4 are the crucial steps in any Molecular Dynamics simulation and will be considered in more detail in subsequent sections.

3.1.3 Limitations of Classical Molecular Dynamics

Before exploring the intricacies of the molecular dynamics algorithm in more detail let us first take a step back and consider the limitations of the technique. Most obviously, this is a classical method using Newtonian mechanics, however, since the early 20^{th} century we have known that we live in a quantum world. Molecules are composed of atoms, which in turn are composed of electrons and nuclei. In quantum mechanics these particles are described by wave functions whose evolution in time and space is governed by the time dependent Schrodinger equation. Electrons are very light and display the characteristic behaviour of both particles and waves. Nuclei are much heavier and move

more slowly. This difference motivates the Born-Oppenheimer Approximation in which the Schrodinger equation is separated into two parts: one describing the electronic wavefunction for a fixed configuration of nuclei, and the other describing the nuclear wave function.

Electronic structure calculations aim to solve the Schrodinger equation for the electronic wavefunction, however, the system sizes which can be treated in this way are limited to several tens or hundreds of atoms. In these calculations the dynamics of the nuclei are calculated by classical mechanics (Newton's Laws) using forces derived from the calculation of the electronic structure. Treating nuclei classically is in general a good assumption, the only possible exceptions being very light atoms such as hydrogen or helium where quantum effects may not be negligible.

In classical molecular dynamics chemical bonds between atoms tend to either be of fixed length or described as a classical harmonic oscillator. Rigid bonds are appropriate if the thermal energy of the system is much less than the separation between the energy levels of the quantum harmonic oscillator. If this is the case the system will be restricted to the ground state. The classical oscillator description is suitable if the thermal energy is much greater than the energy level separation thus allowing the system to effectively access a continuum of energy states. The quantised energy levels of a harmonic oscillator are separated by $\hbar\omega$ where $\hbar$ is the reduced Planck constant and ω is the vibrational frequency of the oscillator. At 308K the thermal energy $k_BT = 2.56\text{kJmol}^{-1}$, while the highest frequency vibrational mode (water OH asymmetric stretch at 3600cm^{-1}[6]) corresponds to an energy of 43.09kJmol^{-1}. As the energy level separation is about 17 times larger than the thermal energy to a very good approximation the system will remain in its ground state and be well described by a rigid bond.

Treating neither the electrons, nor the nuclei quantum mechanically allows for the simulation of much larger systems, up to tens or hundreds of thousands of atoms. Therefore classical molecular dynamics is the most appropriate technique for studying the system sizes and timescales under investigation in this work. Nevertheless, an empirical description where molecules are considered to be atomic balls connected by chemical bonds which can be rod- or spring-like does not allow for natural processes such as electron transfer and the breakage or formation of covalent bonds in chemical reactions. This leads to the question of how realistic classical molecular dynamics simulations actually are.

Particles in the simulations move in response to forces exerted on them by other

particles and therefore it is the specification of the intermolecular potentials, known as the forcefield, which determines whether or not simulation results agree with experimental data. Most forcefields are a summation of bonded terms corresponding to chemical bonds, bond angles or dihedral angles and non-bonded terms representing van der Waals and electrostatic interactions. These terms are empirical and thus tuned to reproduce experimental data such as density or diffusion coefficients. Naturally this leads to a rash of models, some of which reproduce the properties of the physical system better than others. The limitations of current computing resources restrict the size and length of simulations which may also affect the realism of their results. For example, most simulated systems contain hundreds or thousands of particles, numbers which are far removed from the thermodynamic limit whose properties they purport to study. Several computational tricks such as the implementation of periodic boundary conditions are used to overcome these constraints.

3.1.4 Periodic boundary conditions

Each aqueous NMA system simulated for this thesis contains on the order of 3000 atoms. They occupy a volume of approximately $2.7\times10^{-26}\mathrm{m}^3$ corresponding to a cubic simulation box with side length 30Å. Periodic boundary conditions are imposed on this box in order to simulate the bulk phase. If a particle moves "out" of one side of the box it immediately moves "in" from the opposite side. Each particle interacts not only with all the other particles in the simulation box but also with all the periodic images (including its own) illustrated in Figure 3.1. In this way the boundary conditions mimic the presence of an infinite bulk phase surrounding the central simulation box. However, the introduction of periodic boundaries entails the calculation of an infinite number of interactions between particles and all their periodic neighbours, rather than between a finite number of particles in a single simulation box.

3.1.5 Short vs long range interactions

The problem is simplified in that some of the non-bonded interactions are short-ranged, in this context "short-ranged" implies that the effect of the potential is negligible outwith a certain cutoff distance, r_{cut}. If that cutoff happens to be less than half of the width of the simulation box then the number of calculations is much reduced. In this case the only interactions which need be considered are those of a particle with the nearest periodic image of all the other particles. This is illustrated in Figure 3.1. If the potential is not zero by definition at distances larger than the cutoff then an error is introduced by ignoring interactions between particles separated by distances greater than r_{cut}. This error can be reduced by increasing the value of r_{cut}. It is also possible to account for the effect of truncation at r_{cut} by adding a long range correction term

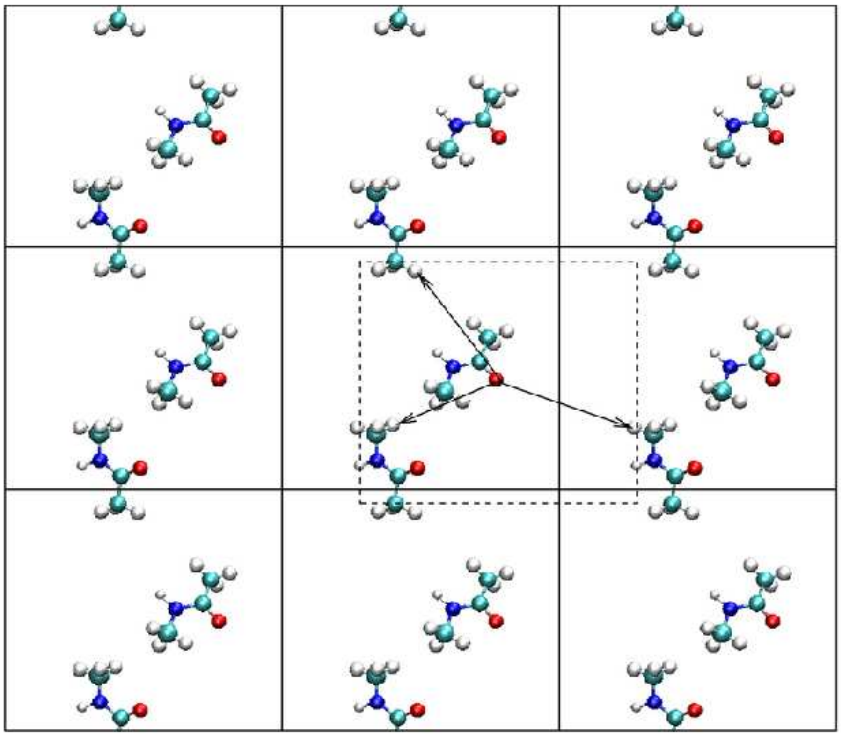

Figure 3.1: The implementation of periodic boundary conditions can be visualised as surrounding the central simulation box with an (infinite) number of copies of itself. As atoms reach one edge of the box they immediately appear at the opposite side resulting in molecules appearing to be split across the box boundaries. In fact, because of the periodicity of the system the exact coordinates of the origin of the simulation box are wholly irrelevant. The dashed box and arrows indicate the interactions of a particle with its nearest periodic neighbours which are not necessarily within the same periodic image.

to the calculation of the total potential energy,

$$U^{tot} = \sum_{i<j} u_{cut}(r_{ij}) + \frac{N\rho}{2} \int_{r_{cut}}^{\infty} dr\, u(r) 4\pi r^2 \tag{3.2}$$

where u_c represents the truncated potential energy function for pairs of particles i and j, $u(r)$ is the actual potential energy function at distance r, N is the total number of particles and ρ is the average number density. This expression implicitly assumes that outside the cutoff the probability of finding a particle is the same everywhere. Therefore in order to minimise errors the value chosen for the cutoff should be no less than the value at which the radial distribution functions described in Section 3.4.1 tend to 1. From Equation 3.2 it can be seen that the long range correction will be infinite unless $u(r)$ decays to zero as r^{-3} or more rapidly. The non-bonded interactions considered in simulations of aqueous NMA are van der Waals and electrostatic interactions. Van der

Waals interactions are modelled using a simple 12-6 pair potential,

$$U(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \tag{3.3}$$

where A and B are constants. Since both terms of this potential decay to zero faster than r^{-3} this is a short-range interaction and can be treated using a spherical cutoff and long range correction term. On the other hand the Coulomb potential for electrostatic interactions is proportional to r^{-1}, decays to zero slower than r^{-3} and consequently cannot be treated by truncation and tail correction. Various techniques have been developed to adequately account for long range electrostatic interactions in simulations of which the Ewald summation method is one of the best and most widely used[92]. In simple terms the Ewald summation replaces an infinite sum in real space with two finite sums, one in real space and one in reciprocal space. Further details on the Ewald method are given in Appendix A.

3.1.6 Integration

Once the forces have been calculated in step 3 of the molecular dynamics algorithm it remains to integrate the equations of motion and step the system forward in time. The DL_POLY package uses integration algorithms based on the Verlet leapfrog scheme. In this method forces and positions are known at integer timesteps while velocities are evaluated at half-integer timesteps. If forces and positions are known at a certain timestep t, velocities are then calculated half a timestep ahead at $t + \frac{1}{2}\Delta t$

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + \Delta t \frac{f(t)}{m} \tag{3.4}$$

The positions at the next timestep are then calculated using this velocity

$$r\left(t + \Delta t\right) = r\left(t\right) + v\left(t + \frac{1}{2}\Delta t\right)\Delta t \tag{3.5}$$

This algorithm is fast and very economical in that it is only necessary to store one set of positions and one set of velocities for the system. In practise it is often useful to know the velocity at integer, rather than half-integer timesteps. In DL_POLY this is achieved by storing a extra set of velocities and averaging over the values at $t - \frac{1}{2}$ and $t + \frac{1}{2}$. One disadvantage of the Verlet scheme is that it is not particularly accurate as it is derived by truncating the Taylor expansion of the particle coordinates around time t. Consequently, the positions are accurate to order Δt^4 while the velocities are only accurate to Δt^2 and therefore it is necessary to choose a reasonably small timestep. Importantly, the Verlet alorithm shows little long term energy drift as it is time-reversible and preserves area in phase space as discussed by Frenkel and Smit in

reference [94].

3.1.7 Rigid molecules

Step 3 is the rate limiting step of the molecular dynamics algorithm and one method of conserving computation time and energy is to consider molecules as rigid bodies and thus avoid calculating forces due to bonds, bond angles, dihedral angles and so on. One means of doing this is to define additional contraints between non-bonded atoms. For example, in the water molecule one would define a constraint between the two hydrogen atoms. In DL_POLY the contraint is implemented by a routine called SHAKE[100] during step 4 of the molecular dynamics algorithm, the integration step described in the previous subsection. The SHAKE algorithm is quite simple,

1. move all atoms according to the Verlet algorithm described above
2. move atoms again according to an additional force proportional to the discrepancy between their new position and the constraint
3. check all bondlengths, if any exceed a given tolerance then repeat the previous step until the tolerance is met

In principle one iteration ought to be sufficient but in the case of molecular systems the correction of one bondlength may cause distortion of another and so a number of iterations may be required before all constraints are satisfied. On occasion the number of iterations can become prohibitive and in these cases it may be preferable to define the rigid body using its centre of mass and a well-defined orientation, then solve the subsequent rotational as well as translational equations of motion.

3.1.8 Summary

The basic mechanics of a molecular dynamics simulation have been described in this section, from its theoretical roots in statistical physics to one or two of the more esoteric computational tricks required to make the problems tackled tractable. Molecular dynamics is a toolkit which can be applied to theoretical models in order both to test the theory and to make experimental predictions. In the following section the theoretical models under investigation in this thesis are described.

3.2 Forcefields

In this section the forcefields used for the simulation of water and *N*-methylacetamide (NMA) are described. There are a wide variety of different water forcefields in current use from which three have been selected for this study. The rigid TIP3P model is

a logical choice as the CHARMM22 forcefield used for the simulation of NMA is parameterised to this forcefield. However, it is well recognised that the TIP3P forcefield does not adequately describe many features of liquid water. Therefore two other water models have also been investigated, another rigid model, SPC/E, which gives results much closer to that of real water and a flexible model known as F3C which was developed specifically to investigate proteins in aqueous solution. Rather than clutter the text with tables, the full set of forcefield parameters used to perform simulations of aqueous NMA are listed in Appendix B.

3.2.1 The TIP3P(CHARMM) forcefield

The TIP3P model as implemented in CHARMM22[76] is a modified version of the original TIP3P water model[101] and has the geometry illustrated in Figure 3.2.

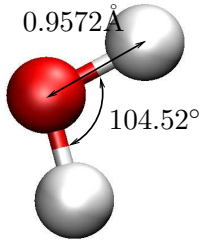

Figure 3.2: Geometry of the TIP3P water molecule: the oxygen (O) - hydrogen (H) bond length is fixed at 0.9572Å and the HH separation is also fixed to give an $H\hat{O}H$ angle of 104.52°.

Intramolecular Interactions

The TIP3P model is rigid and therefore the chemical bonds between the oxygen and hydrogen atoms as well as the non-bonded distance between the two hydrogens are constrained as described in Section 3.1.7. The constrained bond lengths are listed in Table B.2.

Non-bonded interactions: *van der Waals interactions*

The CHARMM potential form for the non-bonded van der Waals interaction between atoms i and j is given in Equation 3.6,

$$U(r_{ij}) = \epsilon_{ij}\left[\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{6}\right] \tag{3.6}$$

where ϵ_{ij} is the well depth of the potential and R_{ij}^{min} is the intermolecular distance at which this minimum occurs. It is common practice to calculate the values of these

parameters for unlike atoms, $i \neq j$, by combining the parameters for interactions between pairs of identical molecules, ii and jj. Values of ϵ and R^{min} for interactions between like atoms are listed in Table B.3. Lorentz - Berthelot mixing rules are implemented in the CHARMM forcefield in which case the cross terms are calculated using the geometric combination rule for ϵ, Equation 3.7 and the arithmetic combination rule for R^{min}, Equation 3.8.

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{3.7}$$

$$R^{min}_{ij} = \frac{R^{min}_{ii} + R^{min}_{jj}}{2} \tag{3.8}$$

In order to minimise calculation time these parameters were converted into the coefficients A and B of the DL_POLY 12-6 potential function given in Equation 3.9.

$$U(r_{ij}) = \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \tag{3.9}$$

By comparison of Equations 3.6 and 3.9 it can be seen that

$$A = \epsilon_{ij}(R^{min}_{ij})^{12} \tag{3.10}$$

$$B = 2\epsilon_{ij}(R^{min}_{ij})^{6} \tag{3.11}$$

These coefficients are listed in Table B.4.

Non-bonded interactions: *Coulomb interactions*

Long range electrostatic interactions between charged atoms are calculated using the Ewald summation method and the charges listed in Table B.1. For the hydrogen bonded interaction between TIP3P oxygens and hydrogens the electrostatic contribution (-60 kcal mol^{-1})to the potential is roughly 800 times larger than the van der Waals contribution (-0.07kcal mol^{-1}). The ratio of these two contributions is similar for the other two water models when calculated using the relevant forcefield parameters and average hydrogen bond length.

3.2.2 The SPC/E forcefield

The SPC/E[102] model was introduced by Berendsen and co-workers in 1987 as an improvement on their original Single Point Charge (SPC)[103] model of water. The SPC/E model reproduces the tetrahedrality of the water hydrogen bonded network much more accurately than the TIP3P models. This was achieved by setting the $H\hat{O}H$ angle to the tetrahedral angle and altering the charges of the earlier SPC model to add a self-energy correction term intended to mimic the effect of polarisation which is ignored by simpler models. Unfortunately as Alex MacKerell points out [76] this

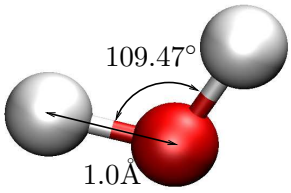

Figure 3.3: Geometry of the SPC/E water molecule: the OH bond length is fixed at exactly 1Å and the $H\hat{O}H$ angle set at 109.47°. This is the tetrahedral angle. In contrast, the TIP3P model uses the OH bond length and $H\hat{O}H$ angle found for the equilibrium structure of an isolated water molecule[104] but there is evidence that in liquid form the intramolecular OH bond lengthens and the $H\hat{O}H$ angle widens[105].

"correction" may cause problems when dealing with two-phase mixtures since the interaction potentials for the second component will be unaware of the energy correction and so there will be an imbalance in the relative strengths of the water-water and water-solute interactions. Despite this note of caution the CHARMM and SPC/E forcefields have been combined successfully, for example by Tarek and Tobias[106], who found that not only was there negligible structural difference between proteins hydrated with TIP3P or SPC/E water but that the dynamical properties of systems simulated using the SPC/E model provided much closer agreement with neutron scattering data than did the TIP3P simulations.

Intramolecular Interactions

SPC/E water is rigid and this contraint was applied as described in Section 3.1.7.

Non-bonded interactions: *van der Waals interactions*

The SPC/E model includes a Lennard-Jones interaction, Equation 3.12 between the oxygen atoms only, which is the same potential form as the DL_POLY 12-6 potential, Equation 3.9.

$$U(r_{ij}) = \frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^{6}} \tag{3.12}$$

Berendsen *et al* give parameters for $A^{1/12}$ and $B^{1/6}$ in (kJ mol^{-1})$^{1/2}$ nm which have been converted into A and B in units of kcal mol^{-1} Å and listed in Table B.6.

Non-bonded interactions: *Coulomb interactions*

As before, the long range electrostatic interactions between charged atoms are calculated using the Ewald summation method and the charges listed in Table B.5.

3.2.3 The F3C forcefield

The F3C forcefield[107] was developed in the late 1990s by Levitt and co-workers specifically for use in simulations of biological macromolecules in solution. The model is simple with each atom represented by a single point charge and all bonds and angles fully flexible. The equilibrium geometry of the F3C water molecule is the same as that of the SPC/E model.

Intramolecular Interactions: *bonds*

Covalent chemical bonds are described using a harmonic potential for the bond length, r

$$U(r) = \frac{1}{2}k(r - r_o)^2 \tag{3.13}$$

Bond parameters k and r_o are listed in Table B.8.

Intramolecular Interactions: *angles*

The HOH angle, θ, is also modelled using a harmonic potential, Equation 3.14.

$$U(\theta) = \frac{1}{2}k_\theta(\theta - \theta_o)^2 \tag{3.14}$$

Angle parameters k_θ and θ_o are listed in Table B.9.

Non-bonded interactions: *van der Waals interactions*

The potential forms of both the bonded and non-bonded interactions are the same in the F3C model as they are in the CHARMM model for NMA which will be described in Section 3.2.4. The equivalence of the potential forms used in the two models is reassuring as it implies that it ought to be possible to combine the two without incurring systematic errors. On the other hand, at this point it should be repeated that the CHARMM authors state that using the CHARMM parameter set with water models other than TIP3P may cause inconsistencies since the cross-species and solute-solute interactions may not be balanced[76]. Although this is most likely to occur as a result of the self-energy correction term inherent in the SPC/E model it is also possible that there may be some inconsistencies when combining F3C potentials with the CHARMM forcefield.

The F3C model values of the well depth ϵ and radial coordinate of the potential minimum, R^{min} for interactions between like atoms are listed in Table B.10. The cross terms are calculated using the geometric combination rule, Equation 3.7, for both ϵ and R^{min}. In order to minimise calculation time these parameters were converted

into the coefficients A and B of the 12-6 potential function given in Equation 3.9 which are listed in Table B.11.

Non-bonded interactions: *Coulomb interactions*

As for both the TIP3P and SPC/E models long range electrostatic interactions are calculated using the Ewald summation method and the charges listed in Table B.7.

3.2.4 The CHARMM forcefield for NMA

The CHARMM22[76] forcefield was chosen as it is a widely used biomolecular forcefield and was already in use by collaborators at IBM. Some forcefields include hydrogen atoms implicitly, for example, the OPLS forcefield includes methyl CH_3 groups as a single "bead". This method restricts the scope of an investigation as such a model cannot display weak hydrogen bonding as the methyl hydrogens are not modelled explicitly. In contrast, use of the CHARMM22 forcefield enables the explicit modelling of all twelve atoms in the NMA molecule, including the methyl hydrogens. Many forcefields constrain the bondlengths and angles rigidly to save computing time but here a more realistic approach is used in which all the intra-molecular bonds, angles and dihedrals are flexible. Table B.12 lists the atomic masses and charges used in the parametrisation of NMA.

Intramolecular Interactions: *bonds*

The 12 atoms in an NMA molecule are connected by 11 chemical bonds which are modelled using a harmonic potential:

$$U(r) = \frac{1}{2}k(r - r_o)^2 \tag{3.15}$$

Bond parameters k and r_o are listed in Table B.13.

Intramolecular Interactions: *angles*

The 18 angles formed by the chemical bonds in NMA are also modelled using harmonic potentials:

$$U(\theta) = \frac{1}{2}k_\theta(\theta - \theta_o)^2 \tag{3.16}$$

Angle parameters k_θ and θ_o are listed in Table B.14.

Intramolecular Interactions: *Urey-Bradley angles*

The angular flex in the NMA model is further constrained by the use of a Urey-Bradley potential which prescribes a distance constraint between atoms 1 and 3 of angle triplets.

These interactions were added by the CHARMM authors to improve agreement of the forcefield with experimental vibrational data.

$$U(s) = \frac{1}{2}k_{UB}(s - s_o)^2 \tag{3.17}$$

Bond parameters k_{UB} and s_o are listed in Table B.15.

Intramolecular Interactions: ***dihedrals, improper dihedrals and 1-4 interactions***

A dihedral angle is formed when two planes intersect as illustrated in Figure 3.4. In

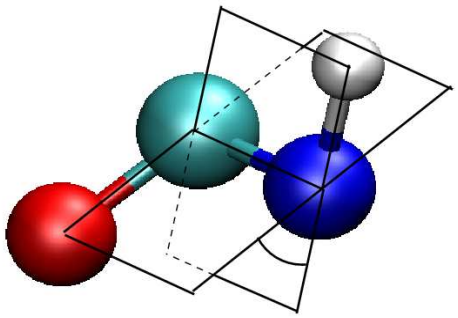

Figure 3.4: Diagram showing the dihedral angle between two planes formed by covalent bonds between the $O = C - N - H$ atoms in the NMA molecule. The atoms have been artificially moved from their natural positions in order to emphasise the angle.

this case we can consider the first plane to be defined by the O-C chemical bond and the C-N bond. The second plane is similarly defined by the C-N and N-H bonds. In this way a dihedral angle can be defined for all groups of four atoms connected by three chemical bonds in this arrangement. In classical molecular dynamics dihedral potentials are used to describe the interactions arising from torsional forces within the molecule. In the CHARMM model the form of the potential is:

$$U(\chi) = k_\chi(1 + cos(n\chi - \delta)) \tag{3.18}$$

Bond parameters k_χ, n and δ are listed in Table B.17. The last two columns of Table B.17 list scaling factors for 1-4 electrostatic and van der Waals interactions. The reason for these scaling factors is as follows. In DL_POLY non-bonded interactions are calculated for all atom pairs except those for which a different interaction type is specified. Therefore, non-bonded interactions are not calculated for any atom pairs which are defined as bonded, such as the first and third atoms in an angle triplet or as the first and fourth atoms in a dihedral. This last is termed a 1-4 interaction.

The DL_POLY parameter specification allows the user to include these interactions multiplied by a scaling factor between 0 (excluded 1-4 interactions) and 1.0 (fully included). 1-4 nonbonded interactions should be fully included in the implementation of the CHARMM model so the scaling factors for both electrostatic and van der Waals interactions ought to be set to 1.0. However, in a few cases the non-bonded interaction is already partially included in the dihedral parameters and therefore applying a scaling factor of less than 1.0 helps to avoid double-counting. The CHARMM model specifies modified van der Waals interactions which are not simply scaled versions of the full non-bonded interactions. In order to implement these modified interactions in DL_POLY I was forced to choose scaling factors which give modified forces which are reasonably close to but not exactly the same as the CHARMM forces.

Improper dihedrals are merely an artificial construct to constrain the molecule in order to better reproduce the correct vibrational modes and need bear no relation to physical bonding patterns. DL_POLY has no separate routines to calculate improper dihedrals and so these are included under the dihedral directive. A harmonic improper dihedral potential is used in the CHARMM model:

$$U(\psi) = k_\psi(\psi - \psi_o)^2 \tag{3.19}$$

Bond parameters k_ψ and ψ_o are listed in Table B.18. The last two columns of Table B.18 list scaling factors for 1-4 electrostatic and van der Waals interactions. In fact the two atom groups (atoms bonded to N and atoms bonded to C) described by the improper dihedral term have no 1-4 interactions so these columns are irrelevant.

Non-bonded interactions: *van der Waals interactions*

The CHARMM potential form for non-bonded dispersion interactions is given in Equation 3.20.

$$U(r_{ij}) = \epsilon_{ij}\left[\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{6}\right] \tag{3.20}$$

ϵ_{ij} and R_{ij}^{min} represent the well depth and the radial coordinate of the minimum of the potential well for interactions between atom types i and j, measured in units of kcal mol^{-1} and Å respectively. Parameters ϵ and R^{min} for interactions between like atoms are listed in Table B.16. Parameters ϵ_{ij} and R_{ij}^{min} for interactions between atom types i and j are calculated from the parameters in Table B.16 using the geometric combination rule for ϵ, Equation 3.7, and the arithmetic combination rule for R^{min}, Equation 3.8. In order to minimise calculation time these parameters were converted into the coefficients A and B of the 12-6 potential function given in Equation 3.9. These

coefficients are listed in Table B.20.

Non-bonded interactions: *Coulomb interactions*

Long range electrostatic interactions are treated using the Ewald summation method described in Section 3.1.5. For the hydrogen bonded interaction between NMA oxygens and hydrogens the electrostatic contribution to the potential energy is approximately 400 times larger than the van der Waals contribution.

3.2.5 Combining forcefields for NMA and water

The specification of van der Waals forces is the only part of the forcefield in which the combination of parameters for different molecules has to be explicitly considered.

TIP3P

Since the TIP3P water model has been parameterised specifically for the CHARMM forcefield there are no theoretical concerns which arise when combining parameters for water and NMA. All that is necessary is the calculation of ϵ_{ij} and R_{ij}^{min} for all the possible atom type pairs using the usual geometric rule for ϵ, Equation 3.7, and the arithmetic combination rule for R^{min}, Equation 3.8. Values of ϵ and R^{min} for the NMA molecule are listed in Table B.16 and in Table B.3 for the TIP3P water model. Applying the combining rules and calculating parameters A and B of the 12-6 potential described by Equation 3.9 gives the values listed in Table B.21.

SPC/E

Combining the NMA CHARMM potentials with SPC/E water is more problematic, even ignoring the obvious question of whether or not naively combining different forcefields in this way is valid. Although both use a 12-6 potential form for the interatomic interaction, Berendsen gives A and B parameters for the water oxygen-water oxygen interaction rather than values of ϵ and R^{min} which are required to work out the interactions between different atom types. Fortunately it is a simple matter to calculate these for the SPC/E model from $A = \epsilon_{ij}(R_{ij}^{min})^{12}$ and $B = 2\epsilon_{ij}(R_{ij}^{min})^{6}$. Since the SPC/E water hydrogens do not participate in Van der Waals interactions it only remains to calculate parameters for interactions between the water oxygens and the NMA atom types using the geometric combination rule for ϵ and the arithmetic rule for R^{min}. These are converted into values of A and B and listed in Table B.19.

F3C

The F3C forcefield uses the geometric combination rule for both ϵ and R^{min} unlike the CHARMM model which uses the arithmetic rule for R^{min}. For heavier atom types there is little difference between parameters calculated by one rule or the other. However, for interactions involving hydrogen atoms there is a significant difference between parameters calculated using the geometric rule for R^{min} and those calculated using the arithmetic rule. This difference is apparent in the simulation results as the maxima and minima of the radial distributions occur at differing values of r. In this case it is therefore important to follow the guidelines of the F3C forcefield and use the geometric combination rule to calculate cross-species values of R^{min}. When this is done A and B can be calculated in the normal way and their values are given in Table B.22

3.3 Simulation protocol

The most complex part in the proceedings is usually not obtaining a suitable molecular dynamics code or forcefield but setting up the "computer experiment". In Appendix C the control parameters and initialisation procedures used for the simulations run for this thesis are described in a way which would enable their repetition by another researcher.

3.4 Post simulation analysis methods

Once the simulation has been run and a set of trajectories generated the next task is to glean some meaning from this deluge of data. Some system properties such as energy and pressure are calculated "on the fly" as the simulation is in progress but others are obtained by post-processing of the trajectory files. This section contains a decription of the analysis methods used to obtain the results presented in Chapters 4, 5 and 6.

3.4.1 Radial distribution functions, $g(r)$

The radial distribution function is one method of quantifying the average structure of a solid or liquid and is particularly important in that it can be extracted from neutron scattering measurements and derived by theory as well as calculated from simulations. Essentially the radial distribution function says whether or not two particles are likely to be found with a separation distance of r. It can be constructed by considering a central particle (coloured dark blue in Figure 3.5) and asking how many particles there are in an annulus of width dr at a distance of r from the central particle. Varying the distance r generates a histogram of the number of particles as a function of distance, n(r). It is clear that this function increases without bound, therefore it is logical to normalise

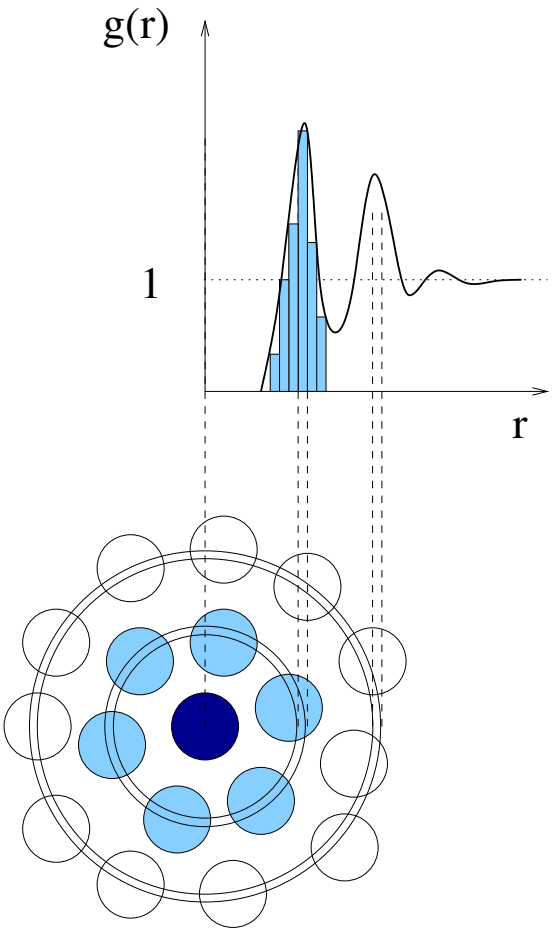

Figure 3.5: The radial distribution function, $g(r)$ and its relation to interparticle structure

by the number of expected neighbours for a system where the particle positions are uncorrelated. This normalised distribution is known as the radial distribution function $g(r)$ and for a liquid has the general form shown in Figure 3.5. First of all it can be seen that there is no contribution at short distances since two particles cannot occupy the same space. There is a peak at a distance corresponding to the first neighbour shell (coloured in light blue). Note that although this peak is sharp it is not a δ-function since liquid particles have some freedom of movement. Naturally after this peak there is a minimum in the distribution as the second neighbour particles cannot occupy the same space as the first neighbours. Since liquid particles are uncorrelated at large separations it is clear that the normalised distribution $g(r)$ for liquids must tend to 1 for large r.

For molecular systems such as the aqueous NMA solutions investigated in this thesis it is insufficient to simply consider the correlations between molecules. Instead the radial distribution functions between specific atom types are calculated as outlined above by

the DL_POLY radial distribution function routine.

3.4.2 2D radial angular correlations

The radial distribution function is not only averaged over all particles of the specified type but also over all interparticle orientations as all contributions within an annulus of width dr are summed. Choosing not to carry out this summation but to instead count contributions according to both their interparticle separation *and* angular dependence gives greater insight into the liquid structure. This can be illustrated by discussing the example in Figure 3.6. In this case the radial angular distribution $g_2(\mathrm{r}_{AA}, \theta_{AAB})$ for a liquid of molecules part of whose intramolecular structure is the covalent bond connecting atoms A and B will be constructed. Given a central atom A (coloured dark blue in Figure 3.6) all atoms within an annulus dr at radius r are identified as in the construction of the radial distribution function, $g(r)$. Then the angle formed by the vector connecting the central atom A to atom A at separation r and the vector connecting atom A at separation r to its intramolecular atom B is calculated. In Figure 3.6 an arrow indicates the intermolecular separation r_{AA} for the dark blue and green coloured atoms of type A. A second arrow indicates the angle formed by this vector and the covalent bond (A-B) in the second molecule. Calculating r_{AA} and θ_{AAB} for all molecular pairs and collecting the results in radial bins of width dr and angular bins of width $d\theta$ enables the construction of $g_2(\mathrm{r}_{AA}, \theta_{AAB})$. A contour plot of this distribution is shown in Figure 3.6 where the x-axis represents r_{AA} and the y-axis θ_{AAB}. The correlation between the dark blue and green A type atoms in the schematic contributes to the second peak of the radial distribution function, but the contour plot gives the additional information that $\theta_{AAB} = 120°$. This extra information is particularly useful in cases where there are two (or more) structural motifs contributing to the $g(r)$ peak. This occurs for the first peak of $g(r)$ in Figure 3.6. It is clear from the contour plot that there are two distinct contributions, one at $\theta_{AAB} = 0°$ (red) and one at $\theta_{AAB} = 180°$ (light blue). In this way the calculation of $g_2(\mathrm{r}, \theta)$ rather than $g(r)$ both deepens the level of information extracted from simulation results and simplifies the identification and classification of structural motifs.

This method was devised by collaborators at IBM, Troy Whitfield and Glenn Martyna, and has been described in detail in references [74, 54]. Their code was extended by myself in order to differentiate between the two different methyl groups in the NMA molecule.

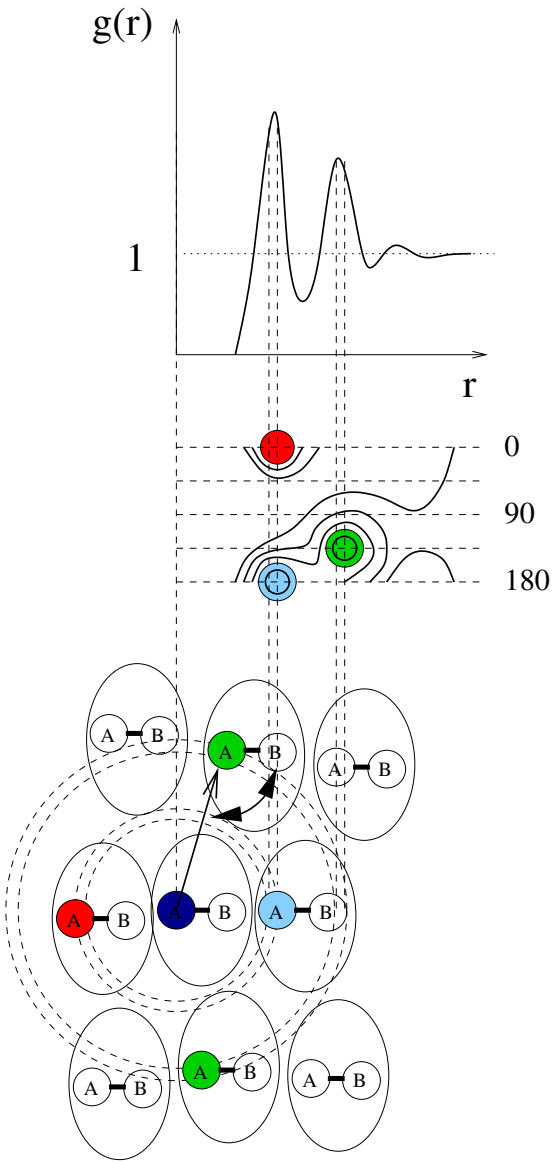

Figure 3.6: The radial angular distribution function, $g_2(\mathrm{r}_{AA},\ \theta_{AAB})$ and its relation to interparticle structure. By summing $g_2(\mathrm{r}_{AA},\ \theta_{AAB})$ over the full angular range for a specific value of r, the one dimensional radial distribution $g(r)$ can be recovered.

3.4.3 Hydrogen bonded cluster analysis

Both NMA and water molecules form intra- and inter-species hydrogen bonds which determine their inter-molecular structure. Therefore a technique to statistically analyse hydrogen bonding behaviour is a prerequisite for characterising the liquid and solution structure of these molecules. The first question is simple: "What is a hydrogen bond?" The phenomenon of hydrogen bonding was introduced in Section 2.3 where the parameters relevant to a geometrical definition were illustrated in Figure 2.2. Here the term is simply defined as it has been applied to the results detailed in subsequent chapters of this thesis.

There are four possible acceptor - donor combinations in aqueous NMA solutions and their acceptor - hydrogen radial distribution functions are plotted in Figure 3.7. Although the functions are similar overall there are sufficient differences that it makes

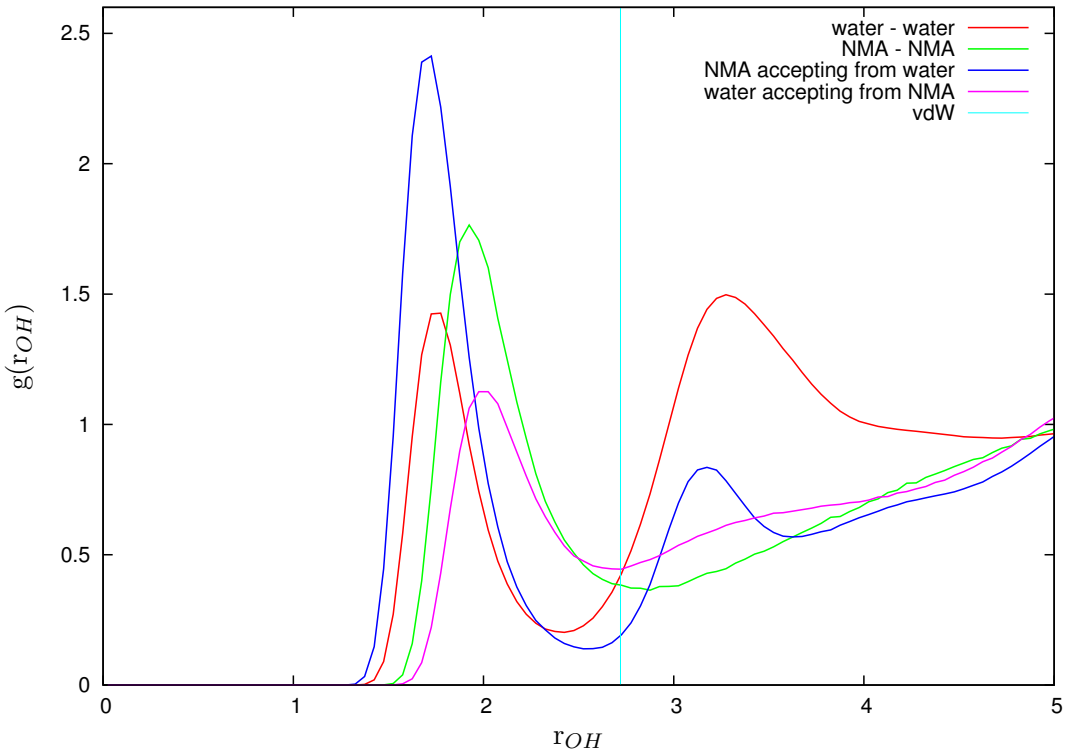

Figure 3.7: Oxygen - hydrogen radial distribution functions for the four different types of oxygen - hydrogen interactions in aqueous NMA.

more sense to consider each as a different type of hydrogen bonded interaction with different mean separations and different strengths. The vertical line represents the sum of the van der Waals radii for oxygen and hydrogen atoms (1.52Å + 1.20Å) and it is obvious that this choice of cutoff does not differentiate clearly between the different

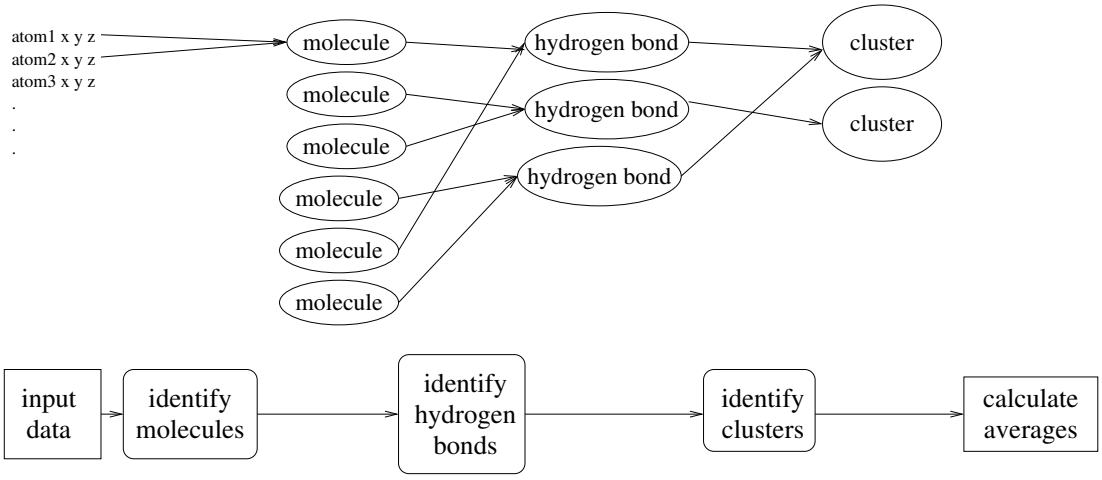

Figure 3.8: A flowchart depicting the object oriented methodology of the hydrogen bond analysis code.

structural configurations contributing to the first and second peaks of $g(r_{OH})$, especially for interactions with water molecules. In general the first peak of $g(r_{OH})$ is known as the "hydrogen bonding peak" and therefore it is logical to take the first minimum of the distribution as the cutoff value. This definition is easy to apply consistently to simulation results after the run has completed and takes into account the differences between the four hydrogen bonded interactions which are so clearly shown in Figure 3.7. Consequently the cutoff distance for a water - water hydrogen bond is different to that for a water - NMA, a NMA - water or a NMA - NMA hydrogen bond and also varies with temperature, concentration and pressure.

The choice of angular criteria is more problematic, with some authors requiring hydrogen bond angles to be within 90°[108] of linear, while others, especially when considering hydrogen bonding in water, set the cutoff much higher, at 30° from linear[13]. A more in depth discussion of the angular cutoff criterion for NMA takes place in Chapter 5. As a result of these considerations the definition of hydrogen bonding used in this thesis is as follows:

- hydrogen - acceptor distance $< r_{min}$ where r_{min} is the position of the 1^{st} minimum in $g(r_{acceptor-hydrogen})$
- donor - hydrogen - acceptor angle $> 140°$

Now that a working definition of a hydrogen bond has been established it is possible to use this to define a cluster as a group of molecules of the same species connected by a continuous network of hydrogen bonds. I wrote Java code to count hydrogen bonds and identify hydrogen bonded clusters following the object-oriented approach outlined in Figure 3.8. The system properties calculated by this code are:

- proportion of molecules accepting or donating specific numbers of hydrogen bonds as a function of donor and acceptor molecular species
- frequency distribution of hydrogen bond angle as a function of donor and acceptor species
- average number of hydrogen bonds donated by species i to species j
- average length of hydrogen bond donated by species i to species j
- average angle of hydrogen bond donated by species i to species j
- average number of hydrogen bonds per molecule as a function of molecular species
- cluster frequency as a function of cluster size
- frequency distribution of number of endpoints as a function of cluster size

3.4.4 Tetrahedral Order Parameter

The two lone pairs of electrons belonging to the water oxygen and the two pairs of electrons shared in the covalent bonds between the oxygen and water hydrogens mutually repel each other with the consequence that they form an approximately tetrahedral arrangement around the nucleus of the water oxygen. Thus, when hydrogen

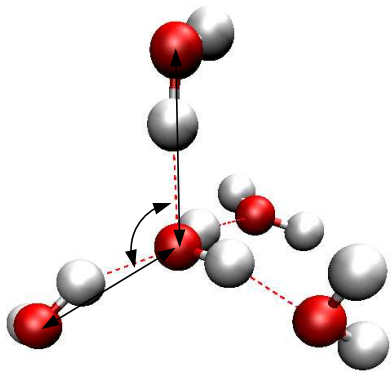

Figure 3.9: A water molecule with four neighbours which form a tetrahedral arrangement due to the mutual repulsion of the negatively charged lone pair and covalent bond electrons. The angles formed between lines connecting the central oxygen with its neighbours are used to calculate the tetrahedral order parameter q. In a perfect tetrahedron the angles between the vertices connecting the oxygen atoms would be 109.47°.

bonds form between water molecules the extended structure takes on a tetrahedral arrangement as shown in the simulation snapshot of Figure 3.9. The extent to which

this arrangement approaches that of a true tetrahedron can be quantified by means of the orientational order parameter q[109] defined in Equation 3.21.

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(cos\psi_{jk} + \frac{1}{3}\right)^2 \tag{3.21}$$

In this equation ψ_{jk} is the angle between the two lines made by joining the central oxygen with the oxygen atoms of its j^{th} and k^{th} neighbours. These lines and the angle between them are indicated in Figure 3.9. If the arrangement is perfectly tetrahedral then $cos\psi_{jk} = -\frac{1}{3}$ for all jk molecule pairs and thus $q = 1$. Conversely, if the arrangement of the four neighbours is completely random then the mean value of q averaged over a large number of molecular configurations will be 0. The possible values of q range from -3 to $+1$ but as will be seen in the results chapters the value of q for real systems tends to lie in the range between 0 and 1. Originally, this method was only applied to pure water, however, in the aqueous NMA solutions water also forms hydrogen bonds with NMA molecules. Regardless of the species of the hydrogen bond partner there will always be mutual repulsion between the lone pairs participating in intermolecular hydrogen bonding and the shared electrons of water's intramolecular covalent bonds. Therefore in order to properly characterise the tetrahedral order of individual water molecules it is necessary to include the hydrogen bonds formed with NMA molecules. The Python code I wrote not only calculates q for water molecules with four water neighbours but also for waters with 1, 2, 3, or 4 NMA neighbours. In the inter-species case an NMA neighbour is defined as an NMA molecule which has either an oxygen (hydrogen bond acceptor) or a nitrogen atom (hydrogen bond donor) within the four closest donors or acceptors to a water oxygen.

3.4.5 Diffusion

Diffusion is the process by which matter is transported in a fluid in spite of the absence of flow and is caused by the thermal motion of individual fluid particles. Particles diffuse due to a concentration gradient, ∇c, in accordance with Fick's Law

$$\mathbf{j} = -D\nabla c \tag{3.22}$$

where $\mathbf{j}$ is the flux of the diffusing species and the constant of proportionality D is known as the diffusion coefficient.

A simple case to consider is the diffusion of a labelled particle through a solution of otherwise identical particles, a process known as self-diffusion. Combining Fick's Law with an expression for the conservation of mass leads to the Einstein diffusion relation

which equates the self diffusion coefficient D with the mean square displacement of the labelled particle[94].

$$D = \lim_{t\to\infty} \frac{1}{6t} \left\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \right\rangle \tag{3.23}$$

While the left hand side of this equation is a macroscopic transport coefficient the right hand side is a microscopic measure of the distance moved by individual particles. This is simple to extract from the trajectories generated by computer simulation. Note that the Einstein relation is only valid in the long time limit. When the mean square displacement is plotted as a function of time there is an initial regime in which it increases quadratically representing ballistic motion (no collisions). After a short time or distance travelled particles begin to experience collisions with other particles and there is a cross-over into the diffusive regime described by Equation 3.23. For this reason the self-diffusion coefficient is calculated from the gradient of the long-time linear portion of the mean square displacement.

Experimentally, self-diffusion coefficients can be measured by quasi-elastic neutron scattering (QENS) as described in Chapter 7. The theoretical equations used to fit the scattering data are derived by assuming that only hydrogen nuclei scatter the incoming neutron beam. This is a safe assumption as the incoherent scattering cross-section of hydrogen nuclei is approximately ten times that of most other elements. However, because of this approximation in the experimental analysis I decided to write my own code to calculate centre of mass mean square displacements as it would then be simple to adapt it to calculate mean square displacements for 1) the centre of mass of hydrogen atoms only and 2) the "centre of mass" according to actual neutron scattering cross-sections. In this way the validity of the hydrogen only assumption for the specific case of the NMA molecule can be checked.

3.4.6 Hydrogen bond dynamics

In Section 3.4.3 hydrogen bonds were defined according to the instantaneous relative geometry of the hydrogen, donor and acceptor atoms. Inherent in the concept of a hydrogen bond is the notion that it is not merely an instantaneous association of three atoms but a favoured structure with some longevity. Consequently it is natural to ask over what period do such complexes persist, in other words, what is the lifetime of a hydrogen bond?

As with the definition of an instantaneous hydrogen bond this question is not as simple as it first appears. First of all, librational motions lead to an apparent breaking and reforming of hydrogen bonds which has led some authors to add a temporal constraint to the existing geometric or energetic hydrogen bond criteria[110]. However, it is

not entirely clear how long this temporal resolution ought to be. Even supposing that this problem can be resolved calculations of average bond lifetimes by direct counting methods are liable to be affected by the system size and length of the simulation. Rapaport introduced the now standard method of determining relaxation times from time dependent hydrogen bond autocorrelation functions[111]. There are two autocorrelations in general use[112, 113, 13]:

- history dependent autocorrelations for molecular pairs bonded *continuously* from time $t_\circ$ to time t
- history independent autocorrelations for molecular pairs bonded at time $t_\circ$ and at time t irrespective of possible breakage between these two times

The second of these has been found to provide more useful results[13] as the first is strongly dependent on the choice of hydrogen bond criteria and consequently it is the second method that I have implemented to investigate the hydrogen bond dynamics of aqueous NMA solutions.

History independent hydrogen bond autocorrelations

First of all, consider a binary operator $h_{ij}(t)$ for each pair of particles i and j

$$h_{ij}(t) = \begin{cases} 1 & \text{if particles i and j are hydrogen bonded at time t} \\ 0 & \text{if i and j are not hydrogen bonded at time t} \end{cases} \tag{3.24}$$

The average number of hydrogen bonds in a system of N molecules is given by the product of the number of possible pairs $\frac{1}{2}N\,(N-1)$ and the mean value of the hydrogen bond operator $\langle h \rangle$. The brackets indicate an average over all possible molecular pairs and times. In an infinite system, $\langle h \rangle = 0$ as the probability of two randomly chosen molecules being hydrogen bonded is zero. This is not true in the simulated system. The finite size effect can be accounted for by autocorrelating the operator $\tilde{h}_{ij}(t) = h_{ij}(t) - \langle h \rangle$ where the trivial correlation has been subtracted from the hydrogen bond operator. The history independent autocorrelation function $c(t)$ is given by

$$c(t) = \frac{\left\langle \tilde{h}(t_\circ)\tilde{h}(t_\circ + t) \right\rangle}{\left\langle \tilde{h}^2 \right\rangle} \tag{3.25}$$

and expresses the probability that two arbitrarily chosen particles are hydrogen bonded at time t, given that the pair were hydrogen bonded at time $t_\circ$, irrespective of whether or not the pair have been hydrogen bonded at times between $t_\circ$ and t.

Following the analysis by Starr[13] a stretched exponential can be used to fit $c(t)$

$$c\left(\frac{t}{\tau_R}\right) = A_\circ \exp^{-\left(\frac{t}{\tau_R}\right)^\beta} \tag{3.26}$$

where the fitting parameter τ_R is the relaxation time. However, this method raises the question of whether to fix $A_\circ$ and β or to leave them as free parameters in the fitting procedure.

An alternative definition of the relaxation time arises from the time integral of the autocorrelation function[113]

$$\tau_R = \frac{\int dt\, tc(t)}{\int dt\, c(t)} \tag{3.27}$$

Although this is a more theoretically satisfying method its accuracy suffers from compounded errors in the long time limit. A third and much simpler method of characterising the correlation function is to identify the time for which it drops to half its initial value[111]. As I am principally interested in probing differences induced by a change of simulation model, temperature or hydration, it is relative changes that are most relevant, for which this simple third method is sufficient.

Chapter 4

Water

Water is an intriguing substance with properties and behaviour quite different to those expected of a simple three atom molecule. In contrast to most other substances the solid form is less dense than the liquid: ice floats on water. Hydrogen bonding is responsible for this phenomenon: in the crystal hydrogen bonds between oxygen and hydrogen atoms of neighbouring molecules form a rigid tetrahedral network. When the solid melts - at a higher temperature than other molecules of this size thanks to the self-same hydrogen bonds - the rigid tetrahedral network becomes disordered and the density increases to a maximum at 4°C. The liquid structure is the subject of ongoing debate with various hypotheses advocating broken hydrogen bonds, bent hydrogen bonds, interstitial molecules or two phases, one tetrahedrally ordered, one disordered.

The biological environment is an aqueous medium in which water is increasingly believed to play an active rather than passive role. Not only does water affect the behaviour of biological molecules but the effect is reciprocated: the presence of such molecules influences the structure and dynamics of water itself. In this chapter the simulation results for three different water models, F3C, SPC/E and TIP3P are analysed using the tools described in Chapter 3 and the liquid structure and dynamics are characterised. The intention is not only to compare and contrast the three models but also to establish a baseline from which to investigate the changes in intermolecular structure and dynamics resulting from the addition of peptide groups in the form of the mono-peptide *N*-methylacetamide.

4.1 Radial distribution functions

Radial distribution functions as described in Section 3.4.1 are a simple yet effective means of characterising the local structure around specific atoms. The two hydrogen atoms in water molecules are equivalent and therefore there are three possible radial

distribution functions: $g(r_{OO})$, $g(r_{OH})$ and $g(r_{HH})$. Radial distribution functions can be obtained experimentally and this has been done for water by neutron diffraction[114]. The process of extracting radial distributions from the experimental data is not without its complications. The two main sources of error in this analysis are firstly the statistical uncertainty in the data and secondly the approximate nature of the estimation of single atom scattering which must be subtracted from the diffraction pattern. Corrections to the raw data also have to be applied to account for effects including background radiation, scattering from the sample container and multiple scattering events taking place in the sample itself. This treatment yields structure factors which can be Fourier transformed to give the radial distribution functions. Transforming each structure factor individually does not guarantee that the obtained set of radial distributions will be self-consistent and correspond to a realistic liquid structure. This problem can be avoided by fitting all the experimental data simultaneously using the Empirical Potential Structure Refinement (EPSR) method developed at ISIS. EPSR is a technique based on reverse Monte Carlo simulation which adjusts an initial reference potential until the simulation results are consistent with the diffraction data and known physical and chemical constraints. The reference potential used for the EPSR fits which are compared with the simulation results in Figures 4.1 and 4.2 was SPC/E water[114]. These fits are the current best guess of the actual radial distribution functions of water under ambient conditions.

4.1.1 1^{st} peak of $g(r_{OH})$: hydrogen bonding

Figure 4.1 compares $g(r_{OH})$ for the three different water models investigated. All are fairly similar and show two main peaks without much variation in the location of these peaks between the three models. The SPC/E and F3C models are particularly close: both over-structure the first peak but crucially it is located in the same place as the neutron result. They also provide much closer agreement with experiment for the first minimum and the second peak than does the TIP3P model. The first peak at 1.8Å arises from hydrogen bonded contacts between water oxygens and hydrogens in neighbouring molecules. Since each water molecule has two hydrogens there will therefore be a contribution from the second hydrogen in the neighbouring molecule and this causes the peak at 3.3Å . Integrating under the curve $g(r)$ yields the number of atoms contributing to the peak which is known as the coordination number. These coordination numbers and the positions of the 1st peaks in $g(r)$ are listed together with the experimental values in Table 4.1. The neutron experiment was carried out at 298K whereas the simulation temperature was set to 308K as this is a more relevant temperature for biological activity. The ten degree temperature difference can be expected to cause some slight differences in the radial distributions as the density

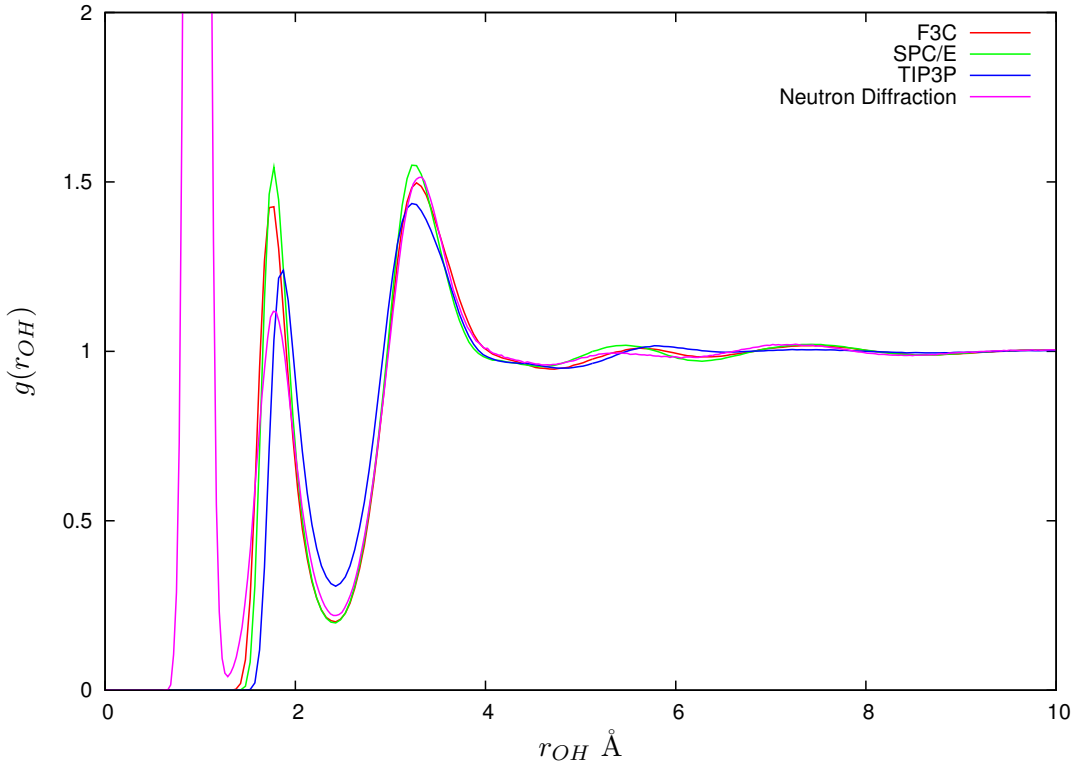

Figure 4.1: Oxygen-hydrogen radial distribution function for the three water models compared with neutron diffraction data taken at 298K[114]. The first large peak in the neutron diffraction data is due to intra-molecular hydrogens but these have been excluded from the simulation results. The bin width of the histograms used to collect radial distribution data was 0.05Å.

of water at 298K is 997.05 kg m^{-3}[6] whereas the simulation density was fixed at the experimental value for 308.15K, 994.11 kg m^{-3}[91]. Peak and minima locations can be expected to move outwards slightly with decreasing density. From the results listed in Table 4.1 it can be seen that the first peak positions of $g(r_{OH})$ for F3C and SPC/E water agree with the neutron result while that of TIP3P is 0.1Å larger. All three models agree well with experiment on the position of the first minimum. The experimental value for the coordination number of the hydrogen bonded first peak in $g(r_{OH})$ at 298K is 1.8±0.06. This agrees with the TIP3P value whereas both the SPC/E and F3C models give a slightly larger value of 1.9. Higher coordination numbers than experiment are to be expected of these forcefields[118]. Given the qualitative and quantitative agreement of the models with experiment it seems that all three models reproduce the water - water hydrogen bond structure with reasonable accuracy. On average each water oxygen is closely associated with just under two hydrogen atoms at separations ranging between 1.3 and 2.4Å.

model	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r_{min})
TIP3P	OH	1.875	2.425	1.8
SPC/E	OH	1.775	2.425	1.9
F3C	OH	1.775	2.425	1.9
expt[114]	OH	1.77	2.4	1.8±0.06
TIP3P	OO	2.775	3.675	6.5
SPC/E	OO	2.775	3.375	4.7
F3C	OO	2.775	3.425	4.9
PBE/PW[115]	OO	2.71	3.30	4.3
BLYP/PW[115]	OO	2.73	3.32	4.4
BLYP/PW[116]	OO	2.74	3.22	4.1
BLYP/DVR[117]	OO	2.77	3.3	4.1
expt[114]	OO	2.73	3.4	4.6±0.1

Table 4.1: Coordination numbers for the 1^{st} peaks in $g_{OO}(r)$ and $g_{OH}(r)$. The experimental values are from neutron diffraction measurements at 298K[34, 114]. Variable results have been obtained using different methods and basis sets for *ab initio* simulations as discussed in references [115, 117]. The results shown are from a comparison of Car-Parrinello molecular dynamics simulations using the Perdew-Burke-Ernzerhof (PBE) and Becke-Lee-Yang-Parr (BLYP) exchange and correlation energy functionals with a plane wave basis set at 300K[115]. Two 2006 publications employ the BLYP functional in conjunction with either a plane wave (PW) basis set (Mantz and Martyna)[116] or a discrete variable representation (DVR) basis set (Lee and Tuckerman)[117] both under ambient conditions.

4.1.2 2^{nd} peak of $g(r_{OO})$: tetrahedral order

The differences between the water models become more apparent in Figure 4.2 which shows $g(r_{OO})$. In particular the TIP3P model produces significantly different results to the other two. The first peak arising from pairs of hydrogen bonded molecules is well reproduced by all three models. Coordination numbers for this peak are listed in Table 4.1 from which it can be seen that all three models predict more dense local structure than that found experimentally. In particular the TIP3P oxygens have nearly two more oxygen atoms within their first neighbour shell which is not nearly as well defined as in the SPC/E or F3C models or the neutron scattering results. The F3C and SPC/E models reproduce the second experimental peak of $g(r_{OO})$ well, while the real drawback of the TIP3P model becomes apparent in that the second peak is wholly absent. This peak arises from second neighbour interactions, specifically molecules which are connected by two hydrogen bonds as illustrated in Figure 4.3. The location of this peak at 4.5Å is particularly significant as this is exactly the distance one obtains by using the cosine rule to calculate the length of the base of an isoceles triangle constructed using the tetrahedral angle, 109.5° and two sides with lengths equal to the sum of the hydrogen bond length, 1.8Å and the length of the OH covalent bond, 1Å. Therefore this peak is important in that it signifies the extended tetrahedral

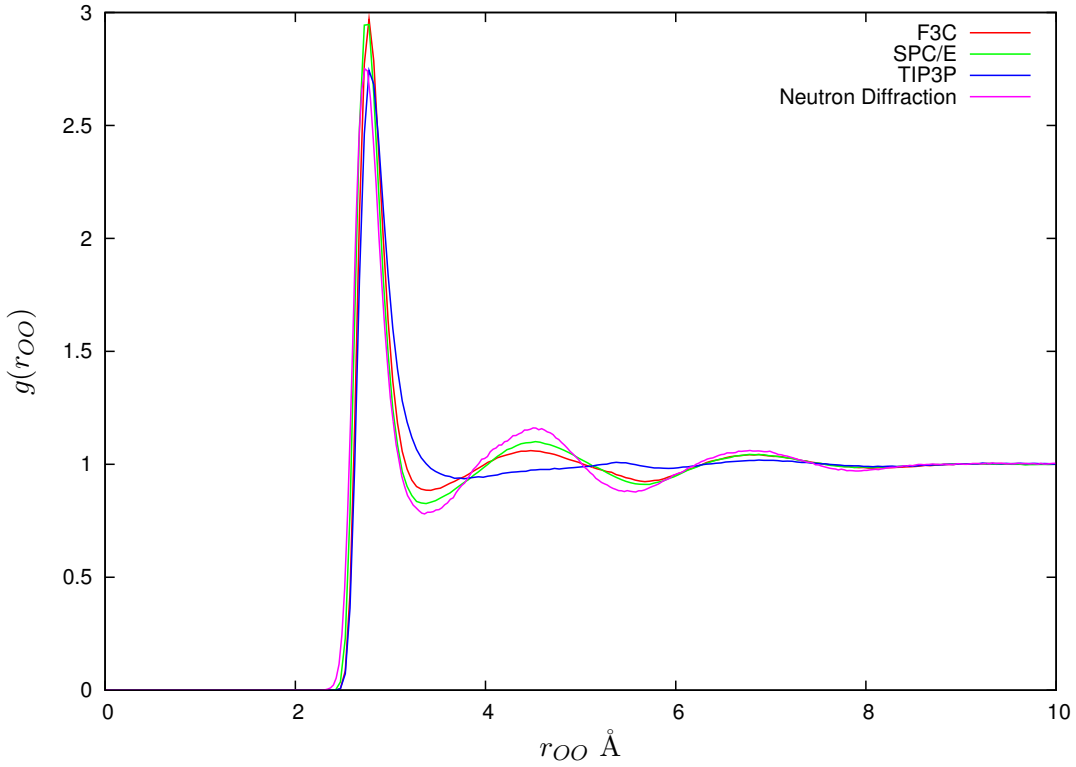

Figure 4.2: Oxygen-oxygen radial distribution function for the three water models compared with neutron diffraction data[114].

intermolecular structure of water. Table 4.2 lists the coordination numbers and 2^{nd} peak positions for the three models. The second minimum in both the F3C and SPC/E

model	2^{nd} peak position (Å)	2^{nd} min position (Å)	n(r_{min})	coord num
TIP3P	5.375	5.925	28.3	21.9
SPC/E	4.525	5.725	25.5	20.8
F3C	4.475	5.675	24.8	19.9
expt	4.5	5.58		18.6±0.06

Table 4.2: Coordination numbers for the 2^{nd} peak in $g_{OO}(r)$ from simulation and experiment[34, 114].

models occurs at slightly larger distances than the experimental result, this can be attributed to the temperature/density difference between simulation and experiment as other comparisons have shown that the F3C, SPC/E and experimental second minima coincide[107]. Both models predict about 20 neighbours in the second shell. As before this is slightly larger than the experimental value of 18.6. (An alternative comparison of various forcefields with the same neutron diffraction data calculated the coordination number of the second peak to be 22.39 for the experiment, 22.58 for the F3C model,

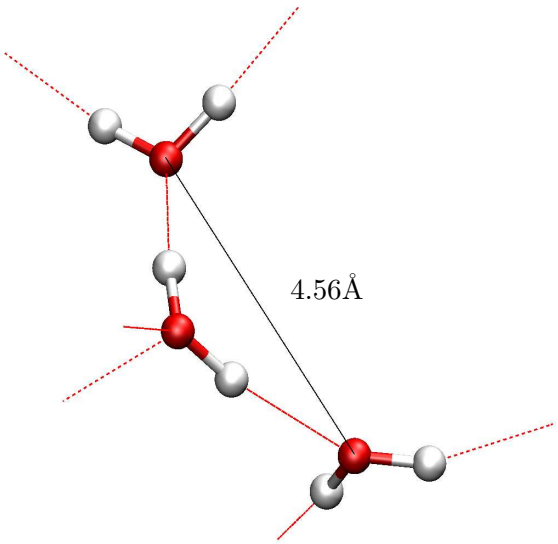

Figure 4.3: Simulation snapshot illustrating the tetrahedral structural motif giving rise to the 4.5Å peak in g(r_{OO}), Figure 4.2. Oxygen atoms are illustrated in red, hydrogens in white and hydrogen bonds identified using geometric criteria are shown as dotted red lines.

22.59 for SPC/E and 22.10 for TIP3P all at 298K and integrated over the range 3.3 to 5.5Å[118].)

4.2 2D radial - angular correlation functions

More information can be obtained by resolving the radial distribution function in an angular direction as described in Section 3.4.2. Figure 4.4 shows the radial angular distributions which result from resolving g(r_{OH}) and g(r_{OO}) in this manner for F3C and SPC/E water. The relevant x and y axis parameters are illustrated schematically in Figure 4.5. Both plots in Figure 4.4 show large angular spreads for most of the peaks as the distributions are necessarily averaged over the two equivalent intramolecular hydrogens. This gives the misleading impression that the hydrogen bond interaction which generates the first peak in Figures 4.4(a) and (b) (labelled A) has a wide angular range since $g_2(r_{HO}, \theta_{HOH}) > 1$ for θ within the range 80° to 180°. In fact, these interactions have a much narrower angular spread as will be shown in Section 4.4.

All of the features at distances between 3 and 4Å in Figures 4.4(a) and (b) can be explained by considering a pair of molecules connected by a single hydrogen bond as illustrated in Figure 4.5. There are two angular regions which contribute to the $g(r_{OH})$ peak at 3.3Å. The broad peak of $g_2(r_{HO}, \theta_{HOH})$ labelled B with an angular

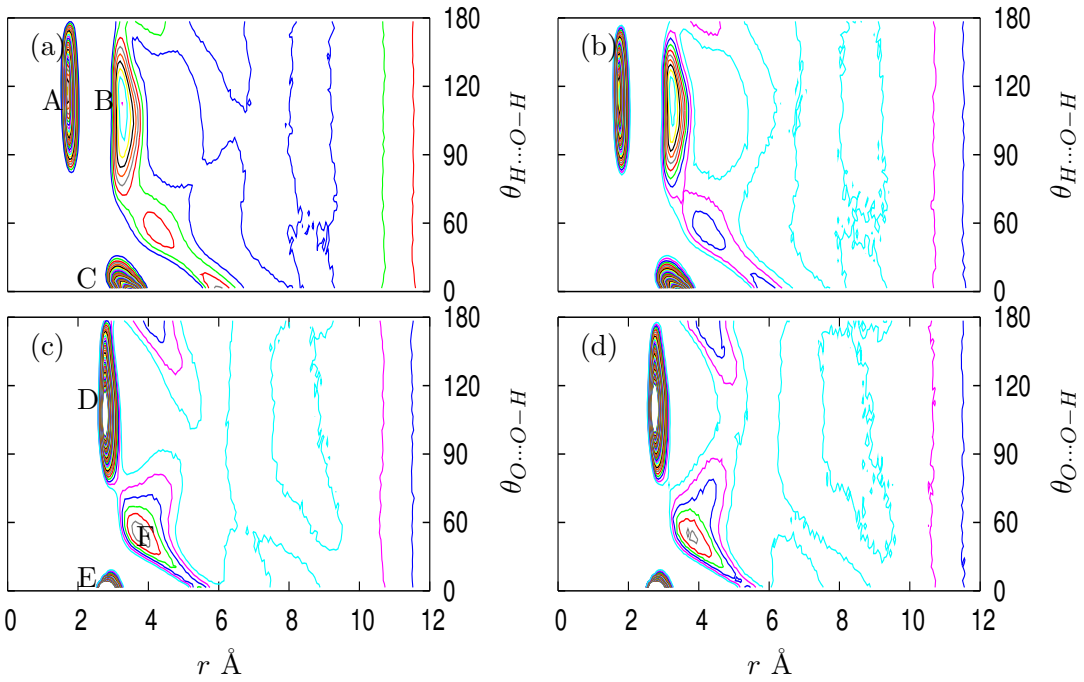

Figure 4.4: Radial angular distribution functions for water-water associations. The upper figures plot $g_2(r_{H\cdots O}, \theta_{H\cdots O-H})$ as a function of $H\cdots O$ separation and $H\cdots O-H$ angle for the (a) F3C and (b) SPC/E models. Figures (c) and (d) show $g_2(r_{O\cdots O}, \theta_{O\cdots O-H})$ which considers the OO separation and $O\cdots O-H$ angle for the F3C and SPC/E models. Contour levels begin at 1 (light blue) and increase in 0.1 increments (except for (a) in which $g_2 = 1$ is dark blue). Labels refer to the interactions described in the text and illustrated in Figure 4.5.

range of 70° to 180° arises from the associations $H_1^1\cdots O_2$-H_2^1 or $H_1^1\cdots O_2$-H_2^2, where the nomenclature refers to the labelling in Figure 4.5. The peak labelled C centred on $\theta = 0$ arises from atoms $H_2^1\cdots O_1$-H_1^2 and $H_2^2\cdots O_1$-H_1^2.

The first set of peaks in $g_2(r_{OO}, \theta_{OOH})$, Figures 4.4(c) and (d) can also be explained with reference to a single pair of hydrogen bonded molecules. The broad peak between 70° to 180° labelled D in Figures 4.4 and 4.5 arises from correlations between atoms $O_1\cdots O_2$ and either H_2^1 or H_2^2.The peak labelled E centred on $\theta = 0$ illustrates the true angular range of the water-water hydrogen bond as this peak arises from configurations equivalent to that between atoms $O_2\cdots O_1$-H_1^2. In this instance there is no misleading angular spread introduced by the indistinguishability of the intra-molecular hydrogens.

A broad band extends across the entire angular range of $g_2(r_{OO}, \theta_{OOH})$ for the SPC/E model, Figure 4.4(d) between 3 and 5Å. In the F3C model this band is broken but otherwise the radial-angular correlations for F3C and SPC/E are very similar. The features within this band do not appear to correspond to those of the one-dimensional

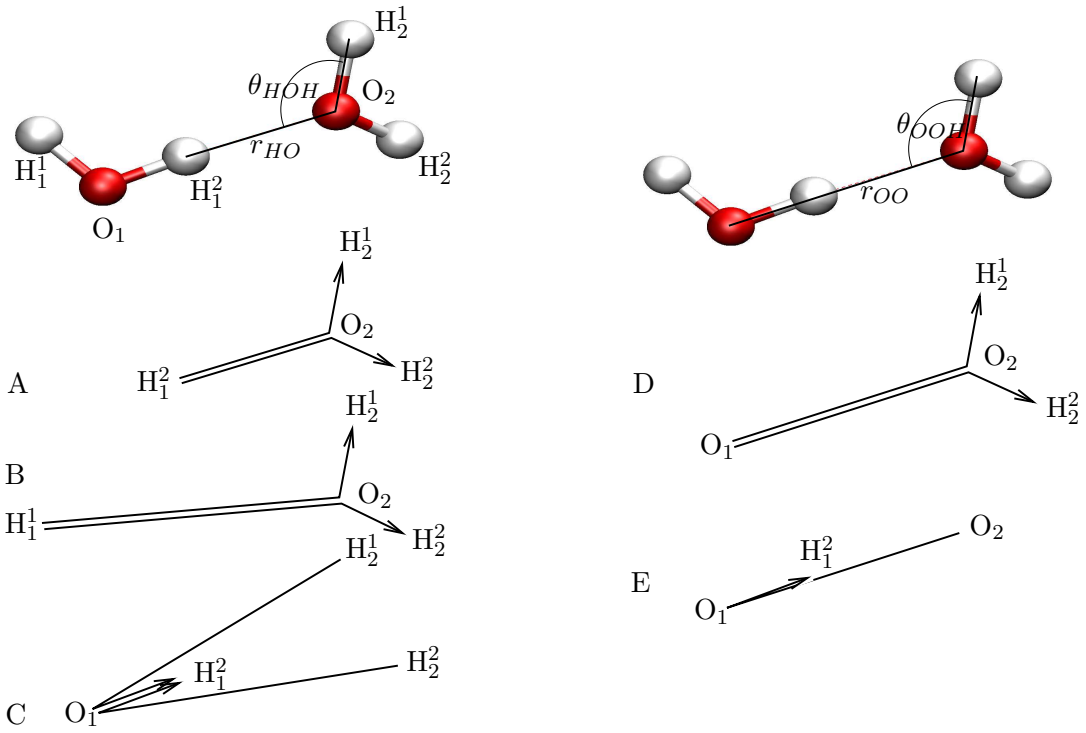

Figure 4.5: The left hand picture defines the oxygen - hydrogen separation and hydrogen - oxygen - hydrogen angle used to define $g_2(r_{HO}, \theta_{HOH})$ plotted in Figure 4.4(a) and (b). The right hand picture defines the oxygen - oxygen separation and oxygen - oxygen - hydrogen angle used to define $g_2(r_{OO}, \theta_{OOH})$ plotted in Figure 4.4(c) and (d). Diagrams A, B, C, D and E correspond to the labelled peaks of the radial angular distributions shown in Figure 4.5.

radial distribution function $g(r_{OO})$ which has a trough centred on $r = 3.4$Å and a maximum at $r = 4.5$Å. The radial-angular distribution clearly shows maxima at $g_2(r_{OO} = 3.8\text{Å}, \theta = 54°)$ and $g_2(r_{OO} = 4.2\text{Å}, \theta = 180°)$. How are these two seemingly contradictory results to be explained? In fact, it has been observed previously by Svishchev and Kusalik that the position of the first minimum of $g(r_{OO})$ coincides with local maxima in $g_2(r_{OO},\Omega)$ for specific solid angles Ω[119]. They propose that two different types of intermolecular ordering exist at what they term "interstitial" separations of 3.2 to 4.4Å. Figures 4.4(c) and (d) clearly show the same feature identified by Svishchev and Kusalik at $g_2(r_{OO} = 3.8\text{Å}, \theta = 54°)$ labelled F. It is difficult to interpret Figure 4.4(b) beyond this confirmation due to the angular spread induced by averaging over both intra-molecular hydrogens in the second molecule. The problem is due to the fact that there are two intra-molecular vectors r_{OH} and a clearer picture would have been obtained by choosing a unique intra-molecular vector such as the dipole moment with which to define the correlation angle.

In fact, it has been known since the 1930s that a four-coordinated tetrahedral network

model of water is insufficient to describe the observed number density at distances less than 4.5Å [120]. This increased number density at low radial separations explains why water can be more dense than ice and may be the origin of many of water's thermodynamic anomalies. Several structural models have been proposed to explain the presence of water molecules at distances smaller than that predicted by a tetrahedral arrangement. A two-state interstitial model whereby one fraction of water molecules forms a three-dimensional ice-like framework while the remainder of the molecules occupy cavities within the framework has been developed by workers such as Samoilov, Frank and Quist[121] and Narten, Danford and Levy[122]. An alternative one-state model proposed originally by Pople[123] accounts for increased number density at small separations by allowing large fluctuations in the lengths and angles of hydrogen bonds. Sciortino, Geiger and Stanley[124] propose a model in which bifurcated hydrogen bonds - in which a single hydrogen acts as a donor to two oxygen atoms - are allowed resulting in a population of five-coordinated water molecules. A conclusive description of liquid water structure continues to evade definition and is an active area of research, particularly in the supercooled regime due to a hypothesised liquid-liquid phase transition in an experimentally unreachable area of the phase diagram between high-density and low-density water: the liquid analogues of high density and low density amorphous ice[125].

4.3 Tetrahedral orientational order parameter, q

Radial and joint radial-angular distribution functions are not the only means of quantifying local structural order in a liquid. The orientational order parameter q introduced in Section 3.4.4 and defined by equation 4.1 provides an alternative method as it measures to what extent the four nearest neighbours around a central water molecule adopt a tetrahedral arrangement[109].

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(cos\psi_{jk} + \frac{1}{3}\right)^2 \tag{4.1}$$

In calculating q only the positions of the water oxygen atoms are considered. In Equation 4.1 the angles $cos\psi_{jk}$ are formed by drawing lines from the oxygen of the central molecule to each pair of its four nearest neighbours. If the arrangement is perfectly tetrahedral then $cos\psi_{jk} = -\frac{1}{3}$ for all jk molecule pairs and thus $q = 1$. Conversely, if the arrangement of the four neighbours is completely random then the mean value of q averaged over a large number of molecular configurations will be 0. The possible values of q range from -3 to +1, but as can be seen from Figure 4.6 the average value for a system ranges between 0 and +1. The lack of a second peak in

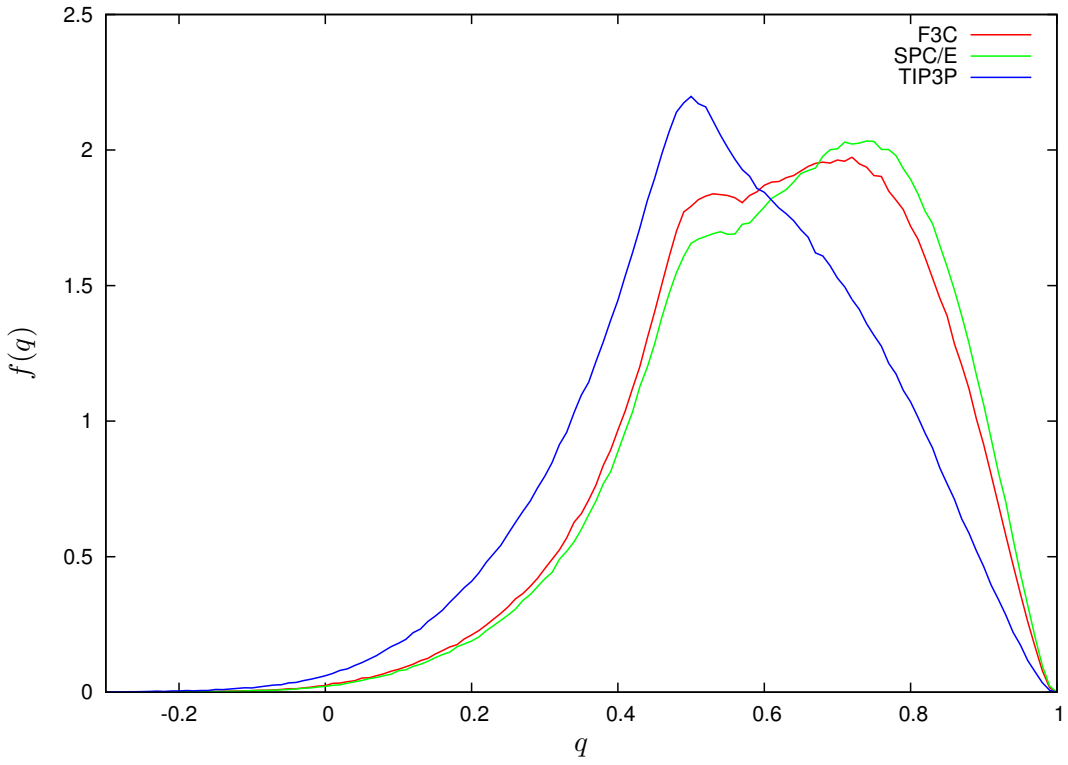

Figure 4.6: Frequency distribution of the orientational order parameter q for the F3C, SPC/E and TIP3P water models. The bin width of the histogram used for collecting frequency data was 0.01 and the area under the each curve sums to 100.

$g(r_{OO})$ for TIP3P water is generally taken to indicate a lack of tetrahedrality which is confirmed by the calculation of q for this model. The frequency distribution of the order parameter peaks at 0.50 and its overall shape is similar to the distribution found for a Lennard-Jones fluid by Errington and Debenedetti[109]. The shape of $f(q)$ for TIP3P water agrees with results in the literature which show that the TIP3P model displays strong temperature dependence and at lower temperatures (240K) resembles the SPC/E distribution[126]. 240K is well below the freezing point of real water but the melting temperatures of water models are well below that of the real liquid. The melting temperature of the TIP3P model has been calculated to be 146K[127] so at 240K it remains a liquid. Most popular water models give a very poor estimation of the melting point: the SPC model melts at 190K and SPC/E only marginally higher at 215K[6]. The shapes of the distributions given by the SPC/E and F3C models are qualitatively similar and display a slight bimodality implying that the instantaneous arrangement of molecules is either well structured (high q peak) or somewhat unstructured (low q peak). This static heterogeneity has been interpreted in support of the hypothesis of a liquid-liquid phase transition as it indicates that there are two types of well-defined local volumes[125]. The equilibrium geometry of the F3C model is the same as that of

the SPC/E with an internal angle of 109.47° equal to the tetrahedral angle. In contrast the TIP3P model uses the $H-O-H$ angle found in gas phase measurements which is 104.52°. The rigid SPC/E model is biased towards a higher degree of tetrahedral order and its high q peak is more pronounced than that of F3C. As the F3C model is fully flexible one can imagine that the O-H covalent bonds may flex and extend to accommodate a wider range of structural configurations than are permitted by the rigid SPC/E model and that the effect of this could be a smearing out of $f(q)$. The average values of q calculated for the F3C, SPC/E and TIP3P models are 0.612, 0.622 and 0.549 respectively.

4.4 Hydrogen bonding

As described in Chapters 2 and 3 a variety of methods can be used to define hydrogen bonding. Geometric criteria are common and one of the most obvious definitions is to

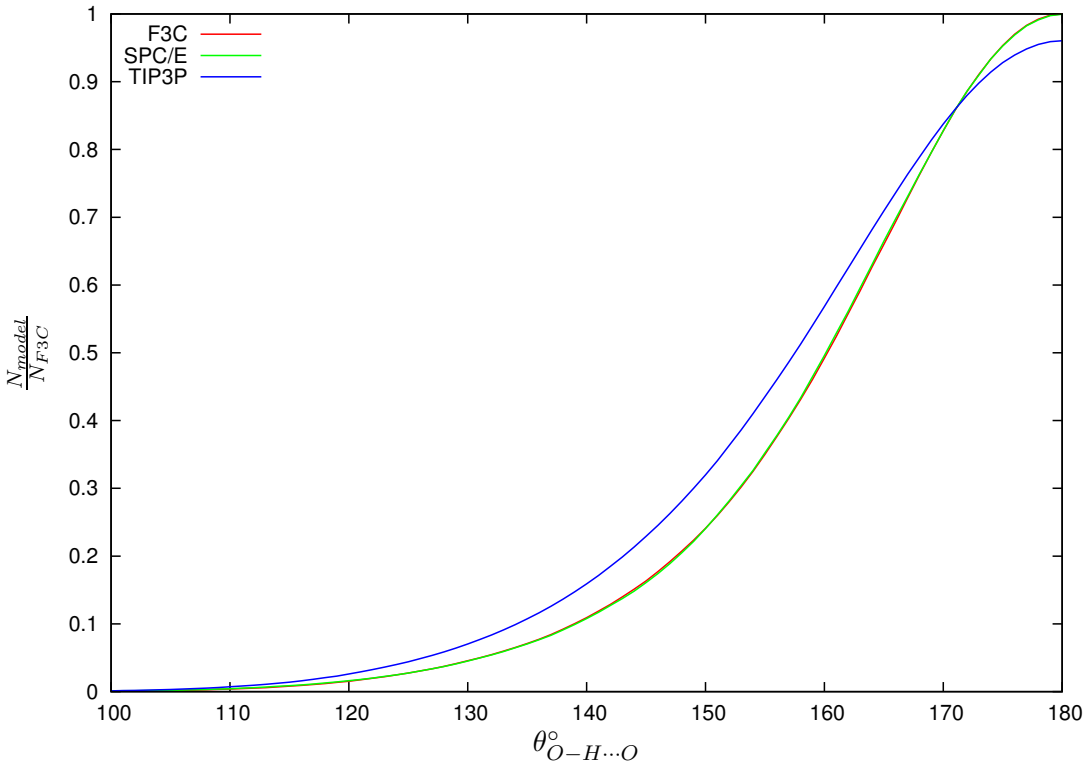

Figure 4.7: Angular dependence of the number of hydrogen bonds in the TIP3P, SPC/E and F3C simulations normalised to the total number identified in the F3C simulation plotted as a cumulative function of the hydrogen bond angle $\theta_{O-H\cdots O}$ defined in Figure 2.2.

consider all pairs of molecules contributing to the first peak of $g(r_{OH})$ as hydrogen bonded. Applying this definition to the simulation trajectories reveals differences between the water models used in the simulations. The F3C and SPC/E models give

almost identical results however the TIP3P model not only has fewer pairs of hydrogen bonded molecules but the hydrogen bonds formed by the TIP3P molecules are less linear. This is illustrated in Figure 4.7 where the number of hydrogen bonds formed by each model is plotted as a cumulative function of the hydrogen bond angle $\theta_{O-H\cdots O}$.

4.4.1 Inclusion of an angular constraint in the hydrogen bond criteria

In general molecules at the tail end of the distribution illustrated in Figure 4.7 are not considered to be hydrogen bonded therefore in addition to the distance constraint it is normal to also impose a restriction on the hydrogen bond angle. The choice of angle is fairly arbitrary but is often taken to be $\theta_{O\cdots O-H} < 30^\circ$ in water[13]. There are a variety of reasonable criteria with which to define bonded and non-bonded regions and the precise definition has little effect upon general behaviour and dynamics beyond an initial transient time[128]. In this thesis a cutoff angle of $140^\circ < \theta_{O-H\cdots O} < 180^\circ$ for water - water hydrogen bonds was decided upon so as to be consistent with the definition used for NMA. (Note that this is not the same angle as $\theta_{O\cdots O-H}$.) The choice of the numerical value of the angular cutoff is discussed in Chapter 5. When the cutoff angle is applied "bent" contacts are excluded and the average length of a "hydrogen bond" decreases showing that the majority of the such contacts are not only bent but also long which further justifies their exclusion. Of the three models F3C has the shortest average hydrogen bond length while that of TIP3P is consistently 0.1Å longer than either the SPC/E or the F3C values regardless of the angular constraint as can be seen from Table 4.3.

constraint	none		$\theta_{O-H\cdots O} > 140^o$		$\theta_{O-H\cdots O} > 150^o$	
	$\langle l \rangle$ (Å)	$\langle \theta \rangle$ (°)	$\langle l \rangle$ (Å)	$\langle \theta \rangle$ (°)	$\langle l \rangle$ (Å)	$\langle \theta \rangle$ (°)
TIP3P	2.00	154.29	1.96	159.30	1.94	162.77
SPC/E	1.91	157.66	1.88	161.02	1.86	163.71
F3C	1.90	157.63	1.86	161.04	1.83	163.75

Table 4.3: Dependence of average hydrogen bond length $\langle l_{OH} \rangle$ and angle $\langle \theta_{O-H\cdots O} \rangle$ on the choice of angular constraint.

4.4.2 General hydrogen bonding behaviour

Imposing an angular constraint significantly alters the distribution of hydrogen bonds per molecule. Figure 4.8(a) shows the population distribution of n-hydrogen bonded molecules when only a distance criterion is used. In the F3C and SPC/E simulation more than half the molecules form four hydrogen bonds and slightly fewer than one third form three. Upon applying the constraint that the hydrogen bond angle must be greater than 140° the fractions of molecules forming three and four hydrogen bonds

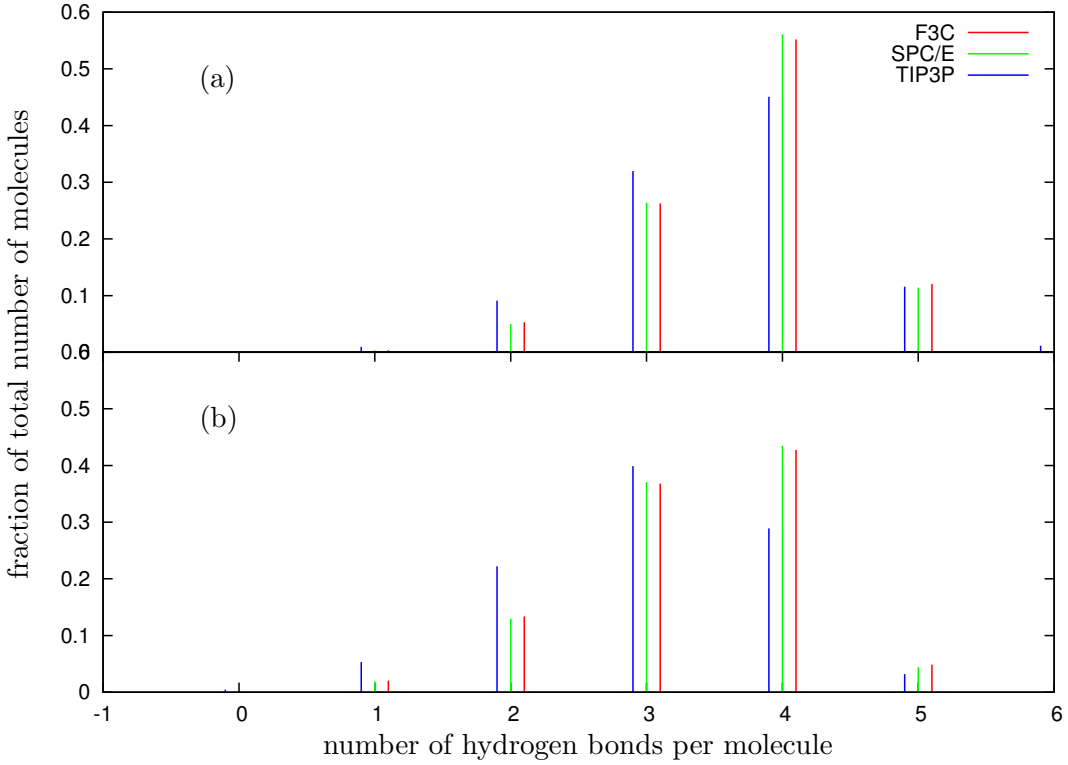

Figure 4.8: The upper figure (a) shows the fraction of water molecules forming 0, 1, 2, 3, 4 or 5 hydrogen bonds without applying an angular constraint while the lower figure (b) shows the fractions which result when an angular constraint of 140° is applied.

become more equal as can be seen by comparing Figures 4.8(a) and (b). Species forming four hydrogen bonds are dominant in the SPC/E and F3C models but the behaviour of TIP3P changes to favouring three. A small fraction of molecules continue to form five hydrogen bonds even when the 140° cutoff is applied, which is also the case for a 150° cutoff (not illustrated). This shows that the concept of a water molecule as consisting of exactly two hydrogen bond donors and two acceptors is too simplistic. A population of five-coordinated water molecules is one explanation which has been proposed to account for the anomalous density maximum of liquid water[124]. One would expect the bifurcated hydrogen bonds which lead to the five-coordinated waters suggested by Sciortino and Stanley to be quite bent and therefore it is striking that they are not eliminated by setting the angular cutoff to 150°. Only a tiny fraction of water molecules do not form any hydrogen bonds at all. As these data are averaged over many static snapshots of a dynamic system this result shows that water molecules are almost continuously connected to a hydrogen bonded network.

4.4.3 Acceptor behaviour

The conventional picture of hydrogen bonding in water is that the oxygen is capable of accepting two hydrogen bonds donated by neighbouring molecules. This is because the

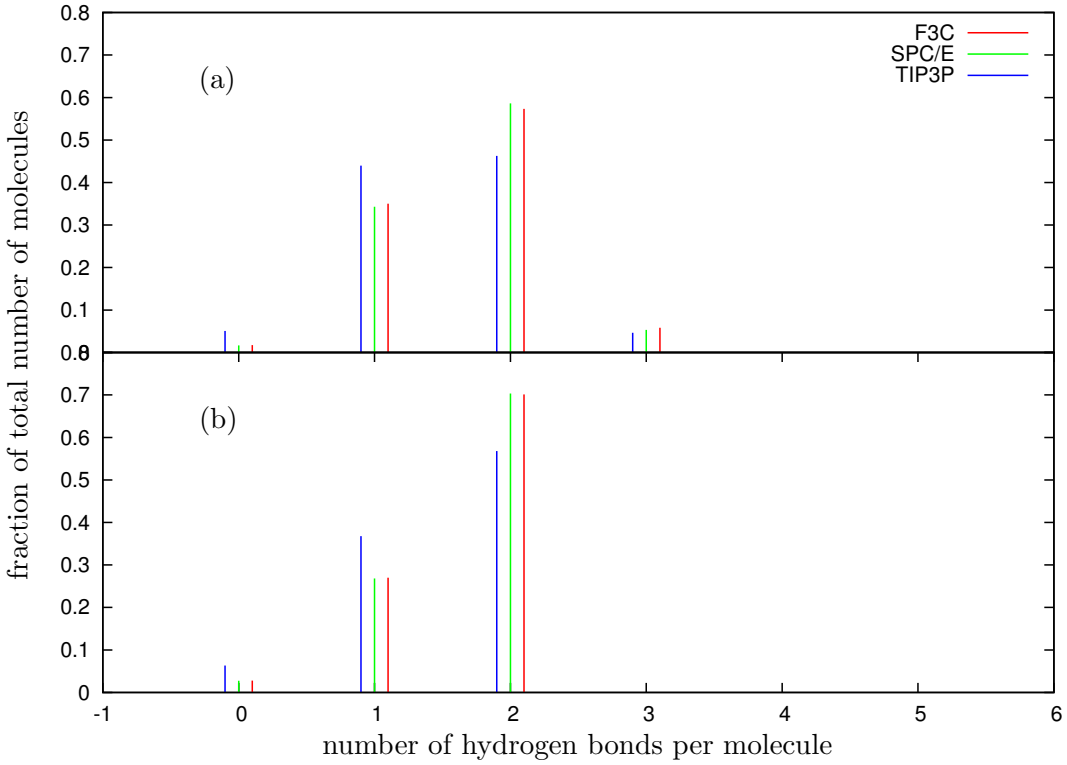

Figure 4.9: The upper figure (a) shows the fraction of water molecules accepting 0, 1, 2 or 3 hydrogen bonds from other water molecules while the lower figure (b) shows the fraction which donate 0, 1, or 2 hydrogen bonds when an angular constraint of $\theta_{O-H\cdots O} > 140°$ is applied.

oxygen atom has 6 electrons in its outer electron shell, two of which form covalent bonds with hydrogen atoms. The remaining four electrons pair up to form what are known as "lone pairs". Each of these negatively charged lone pairs is capable of attracting positively charged species and forming a hydrogen bond. From Figure 4.9(a) it can be seen that although this is in general true, even when using an angular constraint a small fraction (less than 10%) of the water oxygens are "hydrogen bonded" to three hydrogens from neighbouring molecules. The hydrogen bonding propensity of the TIP3P model is different to that of the SPC/E and F3C models in that a larger fraction of TIP3P waters accept zero or one hydrogen bonds whereas a lower fraction accept two or three when a comparison is made between the three models. F3C and SPC/E water are most likely to accept two hydrogen bonds when a 140° cutoff is used to define the hydrogen bond. This is the same behaviour as when no cutoff is used but when a 150° constraint is applied SPC/E and F3C waters are marginally more likely to only accept one hydrogen bond (these comparisons are not illustrated).

4.4.4 Donor behaviour

Figure 4.9(b) illustrates another discrepancy between the TIP3P model on the one hand and the F3C and SPC/E models on the other. TIP3P water molecules are much more likely to not donate any hydrogen bonds at all, or to only donate one, relative to the other models. This lack of double donors easily explains why TIP3P lacks tetrahedral order as found both in the lack of a second $g(r_{OO})$ peak and in the distribution of its tetrahedral order parameter q. Just over 70% of SPC/E and F3C waters donate both hydrogens when the hydrogen bond is defined by a 140° cutoff angle. For a 150° cutoff this proportion is reduced to just over 50% while the TIP3P molecules favour the donation of one rather than two hydrogen bonds.

From these results it is obvious that any discussion of hydrogen bonds is wholly dependent on the convention used to define their existence. The population distribution of n-hydrogen bonded waters calculated here is reasonably consistent with results in the literature using energetic criteria[101]. The main aim of subsequent chapters is to establish how the addition of peptide groups disrupts bulk water structure for which relative rather than absolute values are required. Therefore the exact hydrogen bond definition is not so important as the requirement that once chosen and validated, it is applied consistently.

4.4.5 Hydrogen bonded clustering

Having defined a "hydrogen bond" it is now possible to define a "cluster" as a group of molecules connected by a continuous network of hydrogen bonds. The simulations were carried out using periodic boundary conditions on a cubic box containing 1000 molecules. Figure 4.10 shows that all three models form system size hydrogen bonded networks. This figure plots the cluster mole fraction, X_n, as a function of the cluster size n (number of molecules in the cluster) defined as

$$X_n = \frac{N_n}{N_{tot}} \tag{4.2}$$

where N_n is the number of clusters of size n and N_{tot} is the total number of clusters identified during the simulation. The F3C and SPC/E results are almost identical while the population distribution for TIP3P show that this model forms a more porous system size network. The higher proportion of small clusters can be attributed to the poorer donor and acceptor behaviour identified in previous sections.

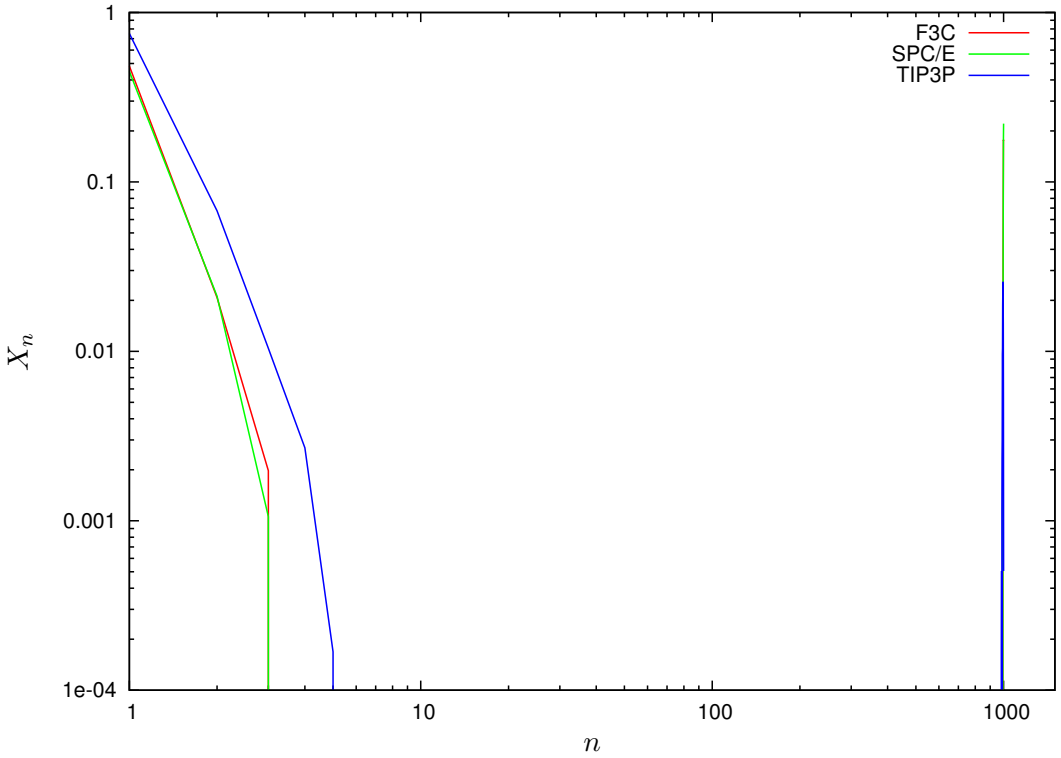

Figure 4.10: Cluster population distribution for the F3C, SPC/E and TIP3P water models.

4.5 Dynamics

The advantage of Molecular Dynamics over Monte Carlo simulation is that as a deterministic method it is possible to investigate the dynamical behaviour of a system in addition to its static properties. The most obvious question to ask concerns the speed with which individual molecules move through the liquid, this process is known as self-diffusion and was described in Section 3.4.5.

4.5.1 Self-diffusion

Self diffusion coefficients are most readily calculated from simulation trajectories by finding the gradient of the linear portion of the mean square displacement plots shown in Figure 4.11. The coefficients for the F3C, SPC/E and TIP3P models calculated in this way are 0.34, 0.31 and 0.64Å^{2}ps^{-1} respectively. All three water models simulated give too high a value of the self-diffusion coefficient when compared to the experimental value which is 0.2907 Å^{2}ps^{-1} (2.907 10^{-9}m^{2}s^{-1}) at 308.15K (35°C)[129]. The SPC/E model gives the closest result at only 0.02Å^{2}ps^{-1} faster than experiment. Allowing for the higher temperature these results are in line with literature values for 298K where the self-diffusion coefficients of SPC/E and TIP3P water are 0.249 and 0.519Å^{2}ps^{-1}

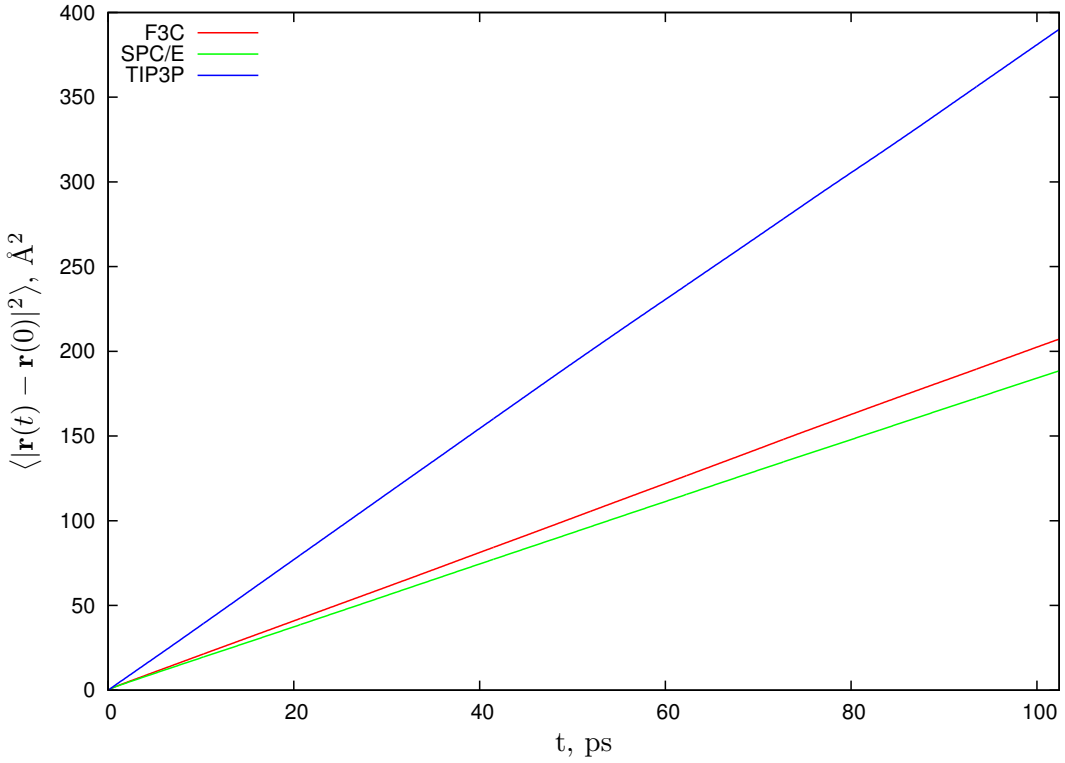

Figure 4.11: Mean square displacement for the F3C, SPC/E and TIP3P water models. The mean square displacement is related to the diffusion coefficient according to Equation 3.23.

respectively compared to an experimental value of 0.23Å^{2}ps^{-1}[6]. Levitt *et al* calculate the self-diffusion coefficient of F3C water to be 0.24 and 0.32Å^{2}ps^{-1} at 298 and 323K respectively[107]. There are discrepancies between the implementation of the F3C model in that paper and in this work which may account for the higher value of 0.34Å^{2}ps^{-1} found in these simulations. Levitt truncates van der Waals interactions at 6Å and compensates using a scaling parameter whereas in this work the cutoff was set at 14.0Å and no scaling parameter was used. Additionally, Levitt's self-diffusion coefficients were calculated over a short 50ps period giving a standard deviation of about 10% (0.023Å^{2}ps^{-1}). For this work mean square displacements were calculated for a 1ns simulation both in its entirety and when divided into four 250ps sections. Correlation arrays with lengths between 50 and 500ps were used to calculate the mean square displacements and it was found that very short and very long arrays do not give accurate values of the self-diffusion coefficient. The error in fitting the mean square displacement results to a straight line are on the order of 1×10^{-5}. This is negligible when compared to the variability between simulation runs. Consideration of the whole suite of simulations yields a conservative error estimate of ±0.01Å^{2}ps^{-1}. While most simulations were carried out in the NVT ensemble, simulations within the NVE ensemble give self diffusion coefficients that are well within this error range and

so it is possible to be confident that the NVT simulations reproduce the correct system dynamics. The SPC/E and F3C models give results which are quite close to experiment whereas the TIP3P self-diffusion coefficient is almost double that of real water. This echoes the findings of Section 4.1 where the TIP3P model failed to reproduce the neutron diffraction radial distributions while both the SPC/E and F3C models gave reasonable agreement. The TIP3P model has been shown to form fewer hydrogen bonds than either of the other two models and therefore it can be concluded that it is the formation of extra hydrogen bonds that causes the self-diffusion of SPC/E, F3C and real water to be much slower than that of TIP3P.

4.5.2 Hydrogen bond dynamics

The use of the word "bond" implies a more than fleeting association between molecules and consequently it is natural to ask for how long hydrogen bonds persist in water and its aqueous solutions. The analysis method for hydrogen bond dynamics was described

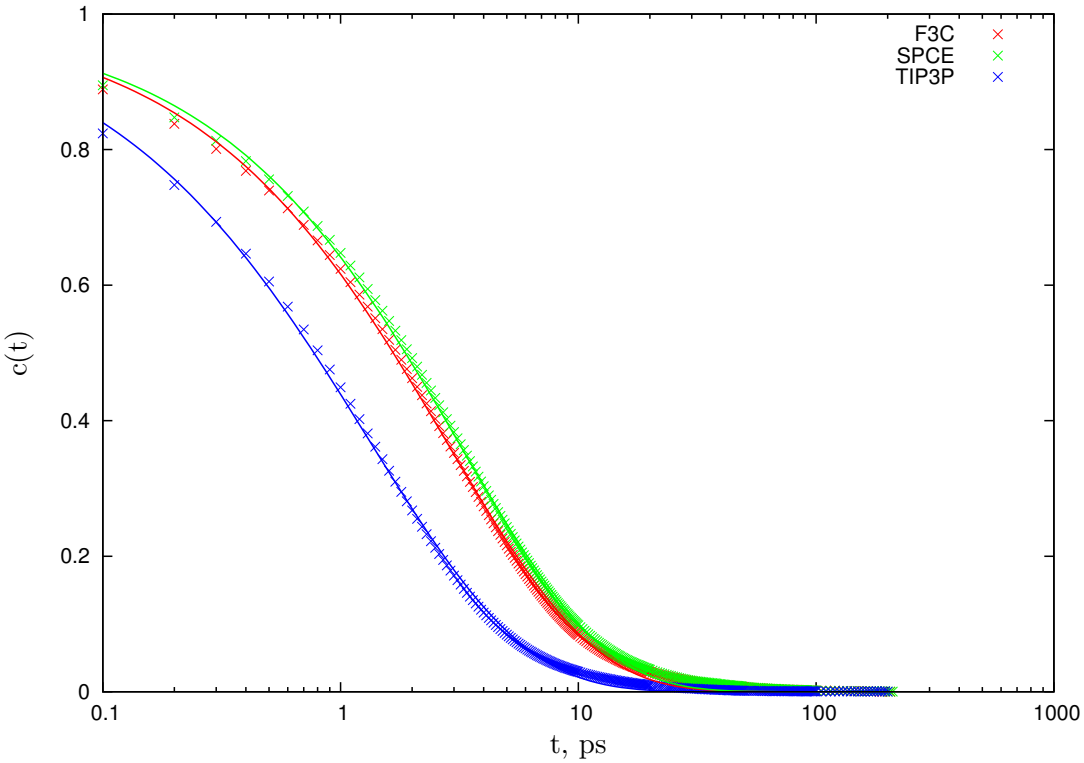

Figure 4.12: History independent hydrogen bond autocorrelations for F3C, SPC/E and TIP3P water models (fits using stretched exponentials).

in Section 3.4.6. Here results are presented for the history independent hydrogen bond autocorrelation $c(t)$, Equation 3.25, for the F3C, SPC/E and TIP3P water models. There are two methods of extracting a quantity with the dimensions of time from

these correlations. Firstly one can define a relaxation time τ by integration of the correlation function as in Equation 3.27 and using this method gives $\tau_{F3C} = 13.0$ps, $\tau_{SPC/E} = 13.4$ps and $\tau_{TIP3P} = 7.8$ps. This is the method used by Raiteri, Laio and Parrinello for simulations of the TIP5P water potential and these results are comparable with their τ_{ω_1} which is approximately 12ps at 304K[113]. The problem with this method is that errors in the long time tail are compounded by integration and consequently long simulations of reasonably large systems are required. For pure water this is not a problem but it becomes an issue when considering the self-diffusion of water in the NMA-rich simulations of aqueous NMA whose results are presented in Chapter 6. An alternative method fits the correlation functions using a stretched exponential.

$$c\left(t\right) = A_{\circ} \exp^{-\left(\frac{t}{\tau_R}\right)^{\beta}} \tag{4.3}$$

Fitting the simulation data using this equation gives the values of $A_{\circ}$, β and τ_R listed in Table 4.4 and both data and fits are plotted in Figure 4.12. Starr *et al*

model	$A_{\circ}$	β	τ_R
F3C	0.996±0.005	0.708±0.005	2.82±0.02
SPC/E	0.992±0.004	0.719±0.004	3.17±0.03
TIP3P	1.004±0.005	0.666±0.004	1.33±0.01

Table 4.4: Fitted parameters and errors in the fit of the simulation data to equation 4.3.

simulated SPC/E water for a range of temperatures and found that both history dependent and independent correlation functions could be scaled to a single curve of the form of Equation 4.3 with constants $A_{\circ} = 0.97 \pm 0.05$ and $\beta = 0.66 \pm 0.06$. When these values are used for fitting the DL_POLY simulation data the relaxation times are not dissimilar to those obtained by fitting with all three parameters : τ_R values of 2.82ps, 3.15ps and 1.38ps are obtained for the F3C, SPC/E and TIP3P models respectively. Starr *et al* found $\tau_R = 2.06$ps at 350K and 5.18ps at 300K using a slightly different geometric definition of hydrogen bonding to that used in this chapter ($r_{OH} < 3.5$Å $\theta_{O\cdots O-H} < 30^{\circ}$)[13]. A simpler method of quantifying the time dependence of the autocorrelations functions is simply to identify the value of t for which the function falls to half its original value[111]. This method gives τ =1.7, 1.9 and 0.8ps for F3C, SPC/E and TIP3P respectively. Dividing these values by $\sqrt[\beta]{ln2}$ (≈ 0.6) derived from Equation 4.3 returns the values obtained from the fit to a stretched exponential.

These relaxation times cannot be interpreted as hydrogen bond lifetimes since no account is kept of how many times the hydrogen bond breaks and reforms between the two times $t_{\circ}$ and t of the autocorrelation function $c(t)$ defined by Equation 3.25

in Chapter 3. Instead, they are a measure of the time it takes for a pair of molecules to diffuse sufficiently far from each other so as to no longer be within what could be termed the hydrogen bonding sphere of influence of the other molecule. In this way these autocorrelations do give a quantitative description of the persistence of hydrogen bonded associations in the liquid. Hydrogen bonded associations are most fleeting in TIP3P water as expected given its high self-diffusion coefficient. F3C water diffuses slightly faster than SPC/E water and similarly its hydrogen bonded associations are marginally less persistent.

4.6 Discussion

In summary, all three water models convey the same broad brush portrait of water as a percolating network of molecules connected by predominantly linear hydrogen bonds. On a deeper level there are significant differences between the TIP3P model on the one hand and the SPC/E and F3C models on the other. These are apparent in the radial distribution functions, epitomised by the lack of a second peak in $g(r_{OO})$ for the TIP3P model. Fewer TIP3P molecule pairs fit the criteria for hydrogen bonding and those which do tend to form longer and more bent hydrogen bonds than those formed by the other two models. TIP3P waters are most likely to form three hydrogen bonds (defined by $\theta_{O-H\cdots O} > 140°$) whereas the F3C and SPC/E models are predominantly four-coordinated. None of the static properties investigated show any significant differences between SPC/E and F3C water. This is interesting as SPC/E is a rigid model while F3C is flexible and thus computationally more expensive. The only slight discrepancy between the results of the SPC/E and F3C models lies in the frequency distribution of the tetrahedral order parameter q. Although their qualitative shape is similar the SPC/E results are more bimodal showing that SPC/E water is unable to adopt as wide a range of configurations as the F3C model, a feature which is probably due to the flexibility of the F3C water model. There are also subtle differences between the dynamical behaviour of the SPC/E and F3C models. Self diffusion coefficients for both models are slightly higher than the experimental value of $0.29 \text{Å}^2\text{ps}^{-1}$ while that of the TIP3P model is more than twice that of experiment. Diffusion and hydrogen bond dynamics are obviously closely related properties and therefore the result that the relaxation time of history independent hydrogen bond autocorrelations in TIP3P water is half that of SPC/E or F3C hydrogen bonds is not unexpected. The similarity of the results for the SPC/E and F3C models is particularly useful as it indicates that if differences between these two models are found in the simulations of aqueous NMA the root cause of these differences must lie in the water-solute interaction potentials.

Chapter 5

N-methylacetamide

N-methylacetamide (NMA) is one of the simplest molecules containing the peptide linkage which is ubiquitous in biological molecules and therefore it has intrinsic value as a model peptide system. It is known from the previous investigations reviewed in Chapter 2 that NMA molecules hydrogen bond to each other through the carbonyl $C = O$ and amide $N - H$ groups to form chain-like structures in the liquid state. In this chapter the structure and dynamics of the pure liquid are characterised so as to establish a baseline from which to determine the changes in the medium-range intermolecular structure and dynamics of NMA that occur upon hydration. Specifically considered are the short range hydrogen bond interactions between peptide groups in neighbouring molecules and the characteristics of clusters defined as groups of molecules connected by continuous networks of hydrogen bonds. Simulations were carried out at three temperatures across the liquid range and therefore in this chapter the changes which take place upon heating are also investigated.

5.1 $g(r_{OH})$ and $g(r_{NO})$: the $N - H \cdots O = C$ hydrogen bond

In Chapter 4 radial distribution functions for the three different water models were compared with neutron diffraction data. Unfortunately this is not possible for liquid NMA as there are no experimentally determined radial distribution functions. The CHARMM22 forcefield was developed for the study of proteins but parameterised to experimental structural, thermodynamic and vibrational data for the model compounds NMA and alanine dipeptide. Consequently, the simulation results for NMA give good agreement with experiment by design[76]. Additionally, the simulated radial distributions of the CHARMM22 forcefield give good agreement with the OPLS combined-atom[68] and all-atom[71] forcefields. A comparison of the CHARMM22 classical forcefield with *ab initio* DFT calculations shows that there are minimal differences between the radial and angular distribution functions for the two models[74].

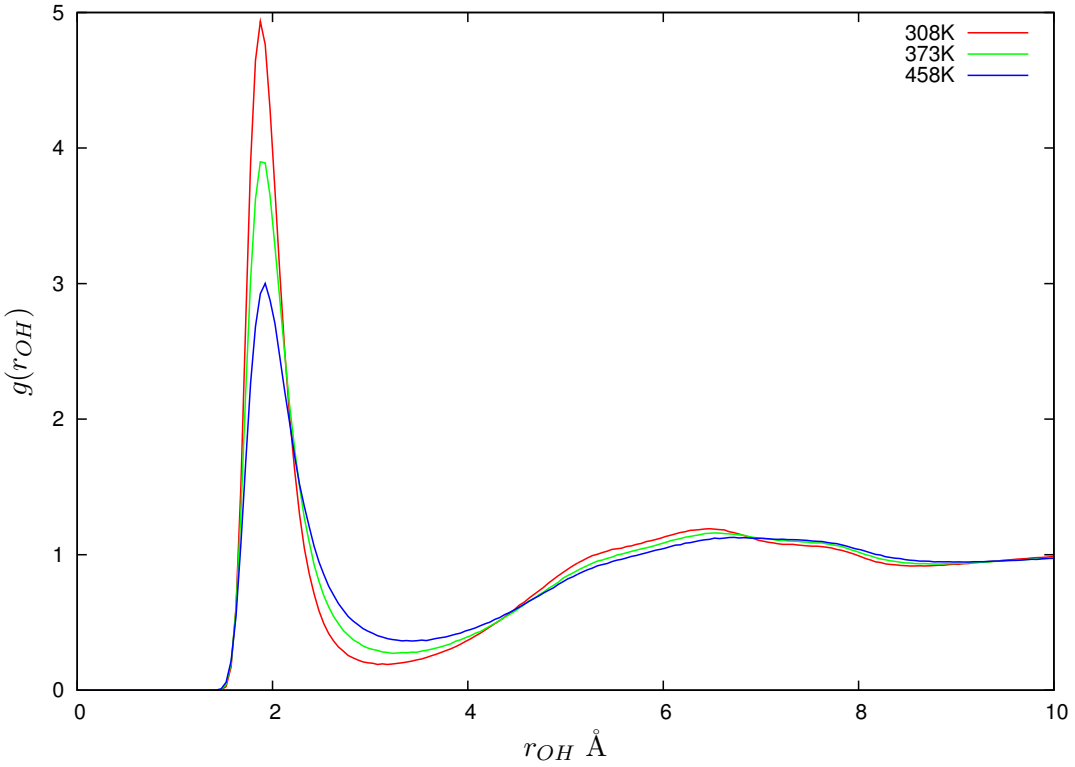

Figure 5.1: $O \cdots H$ radial distribution functions for pure NMA at 308, 373 and 458K

The oxygen - hydrogen radial distribution function $g(r_{OH})$ illustrated in Figure 5.1 clearly displays the characteristic peak of strong, also known as conventional, hydrogen bonding at just under 2Å. When hydrogen bonds form between water molecules as discussed in the previous chapter, the water oxygen acts as both hydrogen bond donor and acceptor. In the NMA molecule the carbonyl oxygen accepts hydrogen bonds while the nitrogen atom belonging to the amide group acts as the hydrogen bond donor. The radial distribution function $g(r_{NO})$ for the donor - acceptor interaction is shown

T (K)	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r_{min})
308.15	NO	2.875	3.675	1.035
373.15	NO	2.875	3.725	0.997
458.15	NO	2.875	3.875	0.999
308.15	OH	1.875	3.075	0.989
373.15	OH	1.875	3.225	0.976
458.15	OH	1.925	3.325	0.939

Table 5.1: Coordination numbers for the 1^{st} peaks in $g_{NO}(r)$ and $g_{OH}(r)$.

in Figure 5.2. Table 5.1 lists the first peak and minima locations and coordination numbers of $g(r_{OH})$ and $g(r_{NO})$ from which it can be seen that each donor and acceptor

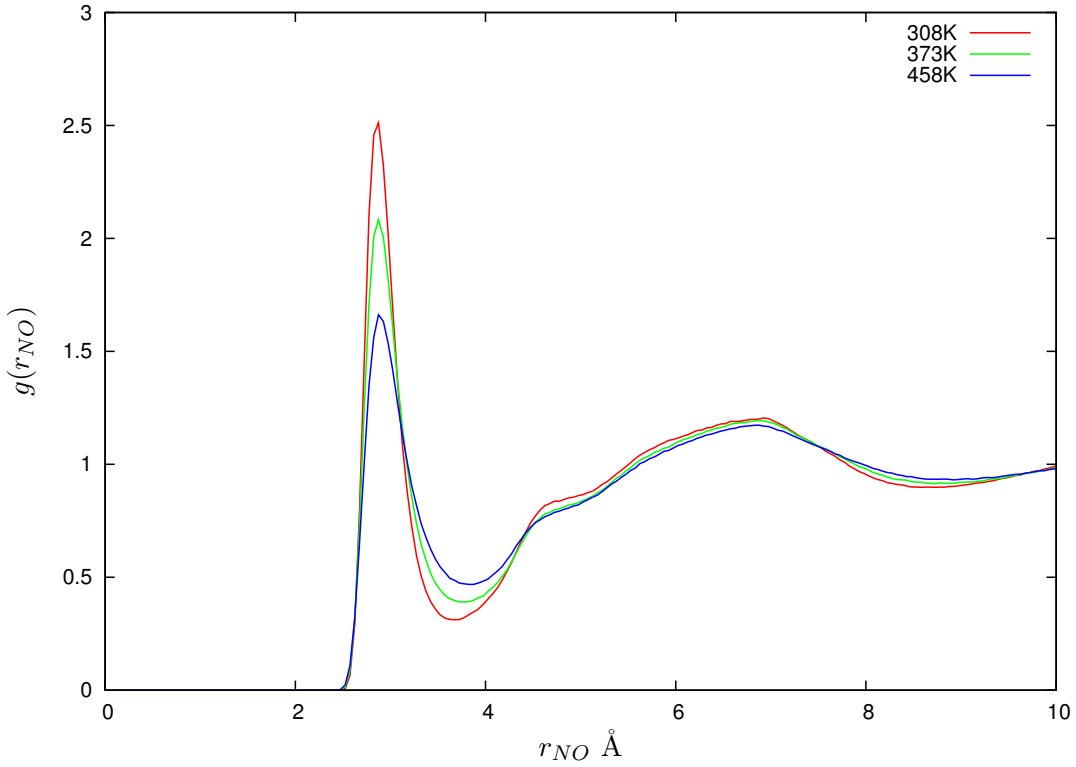

Figure 5.2: $N \cdots O$ radial distribution functions for pure NMA at 308, 373 and 458K

group forms a single hydrogen bond on average and displays a small decrease in coordination number with rising temperature. As the temperature increases there is little change in the positions of the donor-acceptor and acceptor-hydrogen first peaks with only a small outward movement of the OH peak at the highest temperature. This agrees with Czarnecki and Haufa's interpretation of their IR spectroscopic results which suggests that increasing temperature reduces the number of hydrogen bonds but not their strength[61]. However, other authors find temperature-dependent frequency shifts in IR spectra which have been interpreted as the lengthening and weakening of NMA hydrogen bonds as temperature is increased[62]. Additionally, *ab initio* cluster calculations have reported that the NMA-NMA hydrogen bond lengthens as the temperature increases[48]. The positions of the first minimum and the second peak of the simulation results move outward slightly for both distributions which is to be expected given the decreasing density but could also indicate lengthening hydrogen bonds. Further support for the elongation of hydrogen bonds with temperature comes by calculating the mean separation distance of configurations contributing to the first peak in $g(r_{OH})$. Modest increases in length are found as listed in Table 5.2. Average $N - H \cdots O$ angles are also listed in Table 5.2 and show a decrease as the temperature is raised indicating that the $O \cdots H$ separations of NMA-NMA

T (K)	average hydrogen bond length		average hydrogen bond angle	
	no constraint	$\theta_{N-H\cdots O} > 140°$	no constraint	$\theta_{N-H\cdots O} > 140°$
308	2.10	2.03	154.34	159.57
373	2.21	2.09	150.14	158.38
458	2.38	2.18	144.78	157.24

Table 5.2: Mean $O\cdots H$ (acceptor - hydrogen) separation and $N-H\cdots O$ (donor - hydrogen - acceptor) angles for configurations contributing to the first peak of $g(r_{OH})$

configurations contributing to the first peak of $g(r_{OH})$ not only lengthen but also become more bent. Intuitively one expects atoms to move more rapidly and farther from their equilibrium positions as they are given more energy and these observations fit well with such a common sense explanation.

5.2 $g_2(r_{OH}, \theta_{OHN})$: 2D radial angular distribution function

The radial angular distribution, $g_2(r_{OH}, \theta_{OHN})$, illustrated in Figure 5.3 shows a prominent first peak at $r_{OH} = 1.9$Å and $\theta_{OHN} = 180°$. From $g(r_{OH})$ presented

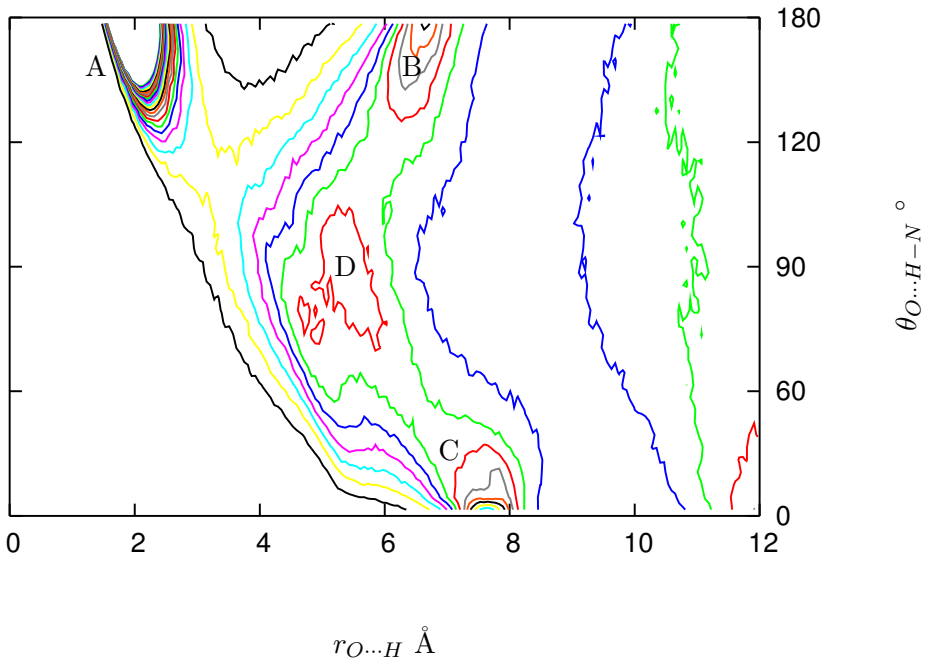

Figure 5.3: Contour plot of $g_2(r_{OH}, \theta_{O\cdots H-N})$, the radial angular distribution function for the acceptor - donor interaction between the carbonyl oxygen and the amide group at 308.15K. Contour levels begin at 0.2 (black) and increase in 0.2 increments up to 5. $g_2(r_{OH}, \theta_{O\cdots H-N}) = 1$ is the dark blue contour. Labels A, B, C and D correspond to the diagrams of Figure 5.4

in Section 5.1 it was clear that the oxygen - hydrogen separation of this first peak lay in the region expected for hydrogen bonding. However, van der Waals interactions

may also satisfy the same distance criteria and for a conclusive definition of hydrogen bonding it is important to show that the interaction is directional. From Figure 5.3 it can be seen that the first peak, labelled A, occupies an angular range beginning at $\theta_{OHN} = 120°$ which is obviously centred on 180° and this confirms the identification of a hydrogen bonded interaction. These bonds are illustrated in Figure 5.4 which shows

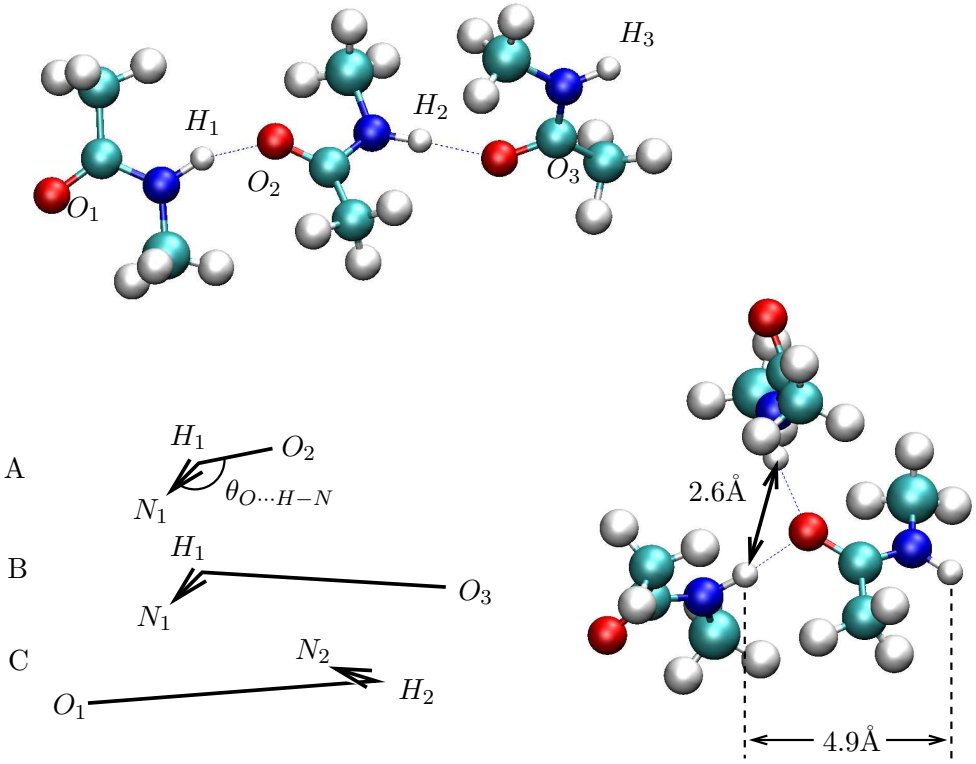

Figure 5.4: Two simulations snapshots illustrate the structural motifs giving rise to the features of $g_2(r_{OH}, \theta_{OHN})$ and $g_2(r_{HH}, \theta_{HHN})$. The left hand linear hydrogen bonded trimer and associated diagrams correspond to the features of Figure 5.3 labelled A, B and C. The snapshot on the right depicts a different type of NMA trimer where one oxygen accepts two hydrogen bonds. This gives rise to feature E of $g_2(r_{HH}, \theta_{HHN})$ shown in Figure 5.5. The separation between equivalent atoms belonging to molecules connected by a hydrogen bond is 4.9Å, thus defining the length of a single "link" in the NMA chain.

two NMA trimers. Interactions such as those between atoms H_1 and O_2 or H_2 and O_3 contribute to the first peaks of g(r_{OH}) and $g_2(r_{OH}, \theta_{OHN})$.

The broad second peak of $g(r_{OH})$ is revealed by Figure 5.3 as being composed of several distinct angular contributions. The second peak in the $\theta_{OHN} = 180°$ direction, labelled B, corresponds to the interaction between "next door but one" neighbours H_1 and O_3 in a hydrogen bonded chain as illustrated in Figure 5.4. The prominent peak labelled C at $\theta_{OHN} = 0°$ arises from the amide and carbonyl groups of a hydrogen

bonded molecular pair which are *not* participating in the hydrogen bond, for example H_2 and O_1. The diffuse lobe labelled D centred on $\theta_{OHN} = 90°$ is perpendicular to the chain direction and therefore gives an indication of the most likely distance between neighbouring chains. However, it is clear that some chains approach each other much more closely that this, with minimum interchain separations starting at just over 3Å. This is closer than the minimum distance observed in x-ray experiments on crystalline NMA where hydrogen bonded NMA chains are stacked in layers with an inter-layer separation of 3.26Å. The distance between adjacent parallel chains within a single layer of crystal is 3.62Å, a short separation which is made possible by interleaving methyl hydrogens belonging to the two chains[38]. The simulated liquid systems are less dense than crystalline NMA: $\rho_{308K} = 0.94624$ kg m^{-3}[91], $\rho_{373K} = 0.894$ kg m^{-3}[68] and $\rho_{458K} = 0.847$ kg m^{-3} compared with $\rho_{283K} = 1.02$ kg m^{-3}[38]. The hydrogen bond only lengthens slightly on melting, from an $N \cdots O$ separation of 2.82Å[38] to 2.88Å judging by these simulation results. Therefore the decrease in density is primarily achieved by an increase in inter-chain separation distances from below 4Å to between 4 and 6Å.

5.3 $g_2(r_{HH}, \theta_{HHN})$: 2D radial angular distribution function

The previous sections have shown that NMA molecules form intermolecular hydrogen bonds donated by the amide N-H group and accepted by the carbonyl oxygen. It is well known that oxygen atoms have two lone pairs of electrons and are therefore capable of accepting two hydrogen bonds. However, the NMA molecule only has one donor group and so the maximum number of hydrogen bonds which can form between amide and carbonyl groups is half that of the number of potential acceptor sites. Do the NMA oxygen atoms therefore restrict themselves to accepting one hydrogen bond each, despite being capable of accepting two? The small peak labelled E in Figure 5.5 centred on $r_{HH} = 2.75$Å and $\theta_{HHN} = 135°$ shows this not to be the case and that some oxygen atoms do accept more than one hydrogen bond. The only structural motif which can account for this peak arises from configurations such as that illustrated in the right hand snapshot shown in Figure 5.4 depicting an oxygen accepting hydrogen bonds from two different molecules. One of the consequences of double acceptance is that the donated hydrogen atoms are brought close together. The coordination number of this small peak is 0.2 implying that a fifth of NMA molecules accept two hydrogen bonds. In reality, this result merely puts an upper limit on the proportion of NMA molecules which accept two hydrogen bonds. The following sections will further explore the concept of hydrogen bonding in liquid NMA. *Ab initio* calculations indicate that the acceptance of a second hydrogen bond by the carbonyl oxygen weakens the original hydrogen bond[67], and even though classical forcefields can only model cooperativity

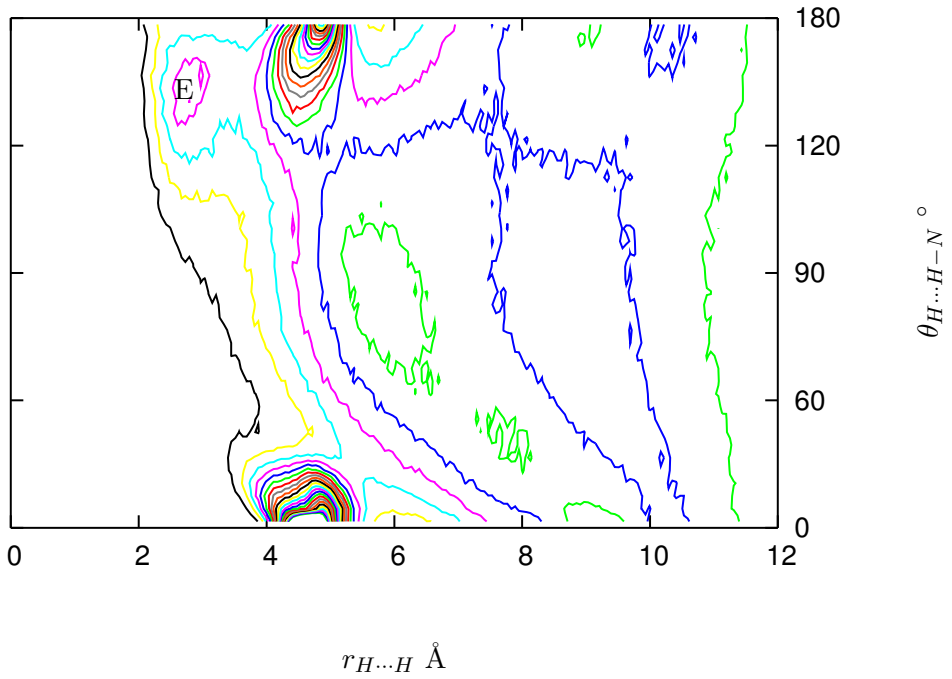

Figure 5.5: Contour plot of 2D radial angular distribution function $g_2(r_{H\cdots H}, \theta_{H\cdots H-N})$ for interactions between amide groups. Contour levels begin at 0.2 (black) and increase in 0.2 increments up to 5. $g_2(r_{H\cdots H}, \theta_{H\cdots H-N}) = 1$ is the dark blue contour.

implicitly through empirical parametrisation, this in conjunction with steric hindrance, explains the low prevalence of doubly accepting oxygens in liquid NMA. On the other hand, NMA is commonly assumed to form purely linear chains in the liquid as it does in the crystal, therefore in a sense it is surprising to find any evidence at all of doubly accepting oxygen atoms. To finish the discussion of Figure 5.5 the peaks with H-H separations of about 4.9Å can be identified as arising from associations between hydrogen bonded neighbours. Their angular dependence is either 0° $(H_2 \cdots H_3 - N_3)$ or 180° $(H_2 \cdots H_1 - N_1)$ corresponding to both chain directions as illustrated in Figure 5.4. The distance of 4.9Å is important in that it is the length of a single unit or link in the NMA chain and thus a characteristic length scale of the system. This length recurs in the analysis of other radial distribution functions and corresponds to a Bragg scattering peak identified in the quasielastic spectra of crystalline NMA.

5.4 Hydrogen Bonding

5.4.1 Geometric hydrogen bond criteria

Just looking at the radial distribution functions g(r_{OH}) and g(r_{NO}) in Figures 5.1 and 5.2 reveals an obvious criterion by which a hydrogen bond between two molecules can be defined. The first peak in both cases is so well defined that it is logical to describe all

molecules contributing to these peaks as hydrogen bonded. Directionality is the primary characteristic distinguishing hydrogen bonds from van der Waals interactions: van der Waals interactions are isotropic in the range 120° to 180° whereas hydrogen bonds show linear directionality. Note that hydrogen bonds will rarely be truly linear because the solid angle is smaller for angles approaching 180°, instead the mean angle will be between 150° and 160° (154.0(4)° [12]). Following this line of reasoning implies that the imposition of an angular criterion is logical and even necessary. Unfortunately the choice of an angular cutoff is less straightforward than the choice of distance constraint as the angular distribution of the $N-H\cdots O$ angle is smoothly varying as shown in Figure 5.6. From this Figure one can see that the most probable angle for contacts

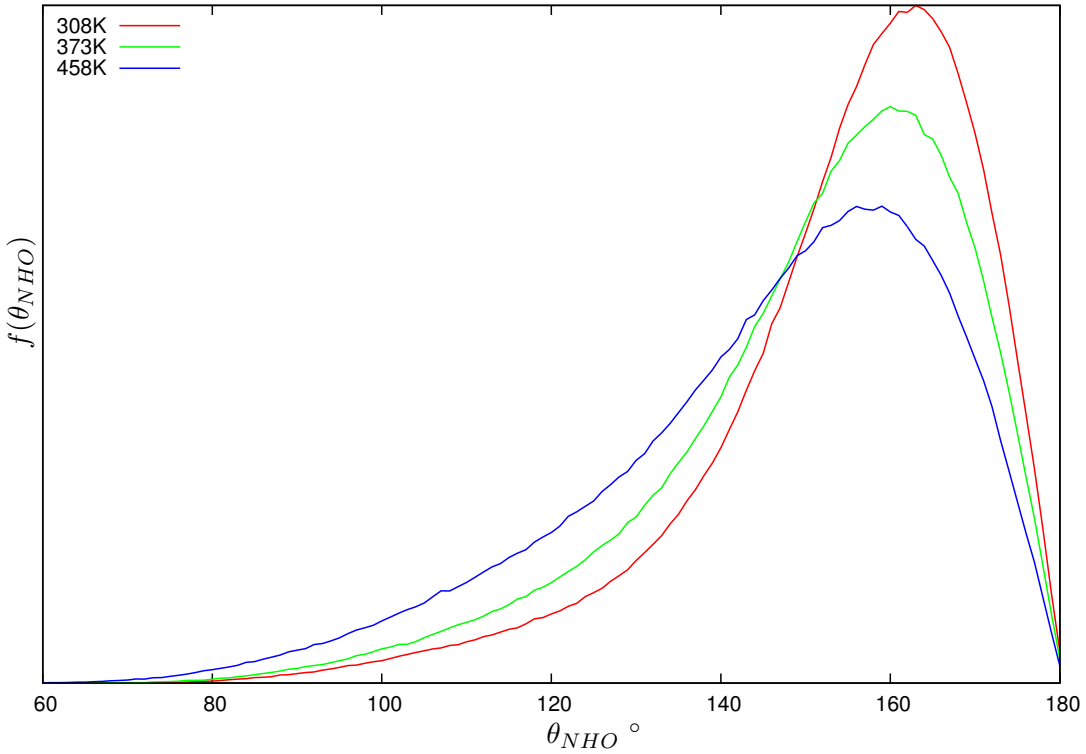

Figure 5.6: Frequency distribution of the hydrogen bond angle $N-H\cdots O$ for NMA molecule pairs which satisfy the hydrogen bond distance criteria at 308.15K, 373.15K and 458.15K

satisfying the distance criteria tends to be around 160°. Taking a range 20° on either side of this excludes close contacts which have very bent (less than 140°) $N-H\cdots O$ angles and are therefore more likely to be van der Waals interactions than hydrogen bonds. As the simulation temperature is increased $\theta_{N-H\cdots O}$ varies over a wider range and becomes less peaked around a single value. This can be interpreted as severe bending or even rupture of hydrogen bonds between neighbouring molecules presumably caused by their faster thermal motion.

5.4.2 Implications of a 140° angular cutoff

The implications of using an angular criterion for the definition of hydrogen bonds was investigated by carrying out calculations for a variety of different cutoff angles. At 308K using a low cutoff such as 130° reduces the number of hydrogen bonds identified to about 90% of the number identified when no angular cutoff is used. A cutoff of 140° reduces the number to about 85% while 150° reduces the number more significantly to around 70%. From the frequency distribution shown in Figure 5.6 it is evident that changing the cutoff angle has more effect at higher temperatures. Almost identical numbers of configurations contribute to the first peak of $g(r_{OH})$ at the three different temperatures investigated. When the constraint $140° < \theta_{N-H\cdots O} < 180°$ is applied NMA molecules in the 373K and 458K simulations form 89.0% and 76.2% respectively of the number of hydrogen bonds found in the 308K simulation. How one interprets this information depends entirely on the definition of the hydrogen bond. If large deviations from a linear configuration are permissible then one could say that as the temperature increases the same number of hydrogen bonds form but these bonds are much more "bent" than at lower temperatures. Conversely, if angular deviations are forbidden then the mean hydrogen bond characteristics (length and angle) remain much the same at higher temperatures but fewer "bonds" are formed. This second interpretation implies that a larger proportion of van der Waals associations are present at higher temperatures which does not seem realistic, therefore it is not sensible to discount bent bonds.

As the angular cutoff is made more stringent the average contact distance for configurations contributing to the first $g(r_{OH})$ peak at 308K decreases from 2.097Å when no cutoff is applied, to 2.051Å (130°), 2.025Å (140°) or 1.990Å (150°) showing that the contacts which are more bent also have a larger separation. Without a cutoff angle a small proportion of the molecules form four or more contacts as shown in Figure 5.7. Applying angular constraints reduces the number of molecules forming two and three hydrogen bonds in favour of those forming one or none at all. When the hydrogen bond angle is constrained to be within the range 140° to 180° most molecules form two hydrogen bonds where the amide group donates and the carbonyl group accepts a single hydrogen bond each. A finite proportion of the molecules have carbonyl groups which accept two hydrogen bonds at the expense of others which accept none. At 308K even without an angular cutoff a higher percentage of molecules (17.2%) accept no hydrogen bonds than accept two (15.7%). When the angular cutoff is raised to 150° the form of the distribution changes as shown in Figure 5.7, suggesting that this cutoff is too restrictive and no longer characterises the intermolecular structure.

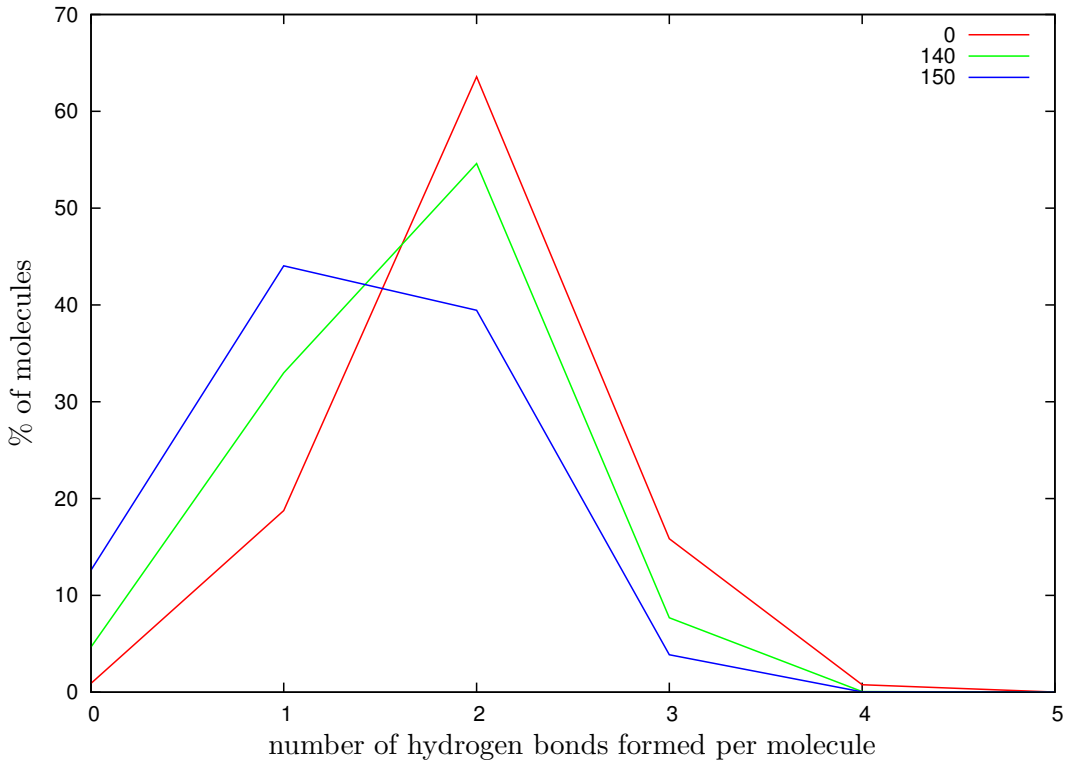

Figure 5.7: Variation in the number of hydrogen bonds formed per molecule at 308K as the cutoff angle used to define a hydrogen bond is changed from $\theta_{N-H\cdots O} = 0°$ to $\theta_{N-H\cdots O} = 150°$

5.4.3 Comparison with the OPLS forcefield

The 373K results can be directly compared to those of the OPLS combined-atom model for which an energetic criterion ($E_{int} < -4.5$ kcal mol^{-1}) as well as a distance cutoff ($r_{OH} < 3.2$Å) was used to define NMA-NMA hydrogen bonds[68]. The distance criterion corresponds to the location of the first minimum in $g(r_{OH})$ for the OPLS model and is the same as that found for the CHARMM22 forcefield. The comparable data listed in Table 5.3 show discrepancies between the hydrogen bonding behaviour of OPLS and CHARMM22 NMA which can be attributed both to differences between the two forcefields and to the different hydrogen bonding criteria used. Without further investigation it is not possible to quantify the two contributions. Although qualitatively the radial distribution functions for both forcefields are similar the coordination numbers of the first peaks of $g(r_{OH})$ and $g(r_{NO})$ are slightly larger for the OPLS model (1.01 and 1.15 compared to 0.98 and 1.00 for CHARMM22 at 373K) indicating that the hydrogen bonding behaviour of the two forcefields is subtly different.

	OPLS	CHARMM22	
		no angular cutoff	$\theta_{N-H\cdots O} > 140°$
HB per molecule	1.77	1.94	1.47
$n = 0$	2.6	1.7	9.3
$n = 1$	27.0	22.6	40.3
$n = 2$	61.4	56.8	44.2
$n = 3$	8.8	17.6	6.1
$n = 4$	0.2	1.2	0.0

Table 5.3: Comparison of the average number of hydrogen bonds formed per molecule, and the percentages of molecules forming between 0 and 4 hydrogen bonds for the CHARMM22 all atom and OPLS combined atom forcefields at 373K. OPLS data from reference [68].

5.4.4 Criteria for strong hydrogen bonding

A clear definition of strong hydrogen bonding between the peptide groups of NMA molecules must now been established. Each interaction should satisfy the following statements:

1. the hydrogen bond length (separation between donated hydrogen and accepting oxygen) must be less than the position of the first minimum in $g(r_{OH})$.

2. the hydrogen bond angle $N - H \cdots O$ must be between 140° and 180°.

5.4.5 General hydrogen bonding behaviour

In the 308K simulation just over 50% of the NMA molecules form two hydrogen bonds according to the above definition which implies that the molecules form linear chains in the liquid. As the temperature increases the proportion of molecules forming two or more hydrogen bonds decreases in favour of molecules which form none or only one. Interestingly, although the number of molecules forming three hydrogen bonds decreases with increasing temperature, it does not change as dramatically as one might expect from the changes in proportions of molecules forming 0, 1 or 2 hydrogen bonds.

5.4.6 Acceptor behaviour

The NMA oxygen is capable of accepting two hydrogen bonds and Figure 5.9(a) shows the fraction of molecules which accept either 0, 1 or 2 hydrogen bonds. From this it can be seen that the majority of NMA oxygens do not fulfil their potential hydrogen bonding capacity at any of the three temperatures. The small minority of NMA oxygens which do accept two hydrogen bonds seem to be less temperature dependent. This could be because conformations where the oxygen fulfils its full hydrogen bonding potential are more stable and thus less likely to be broken up by increasing the kinetic energy

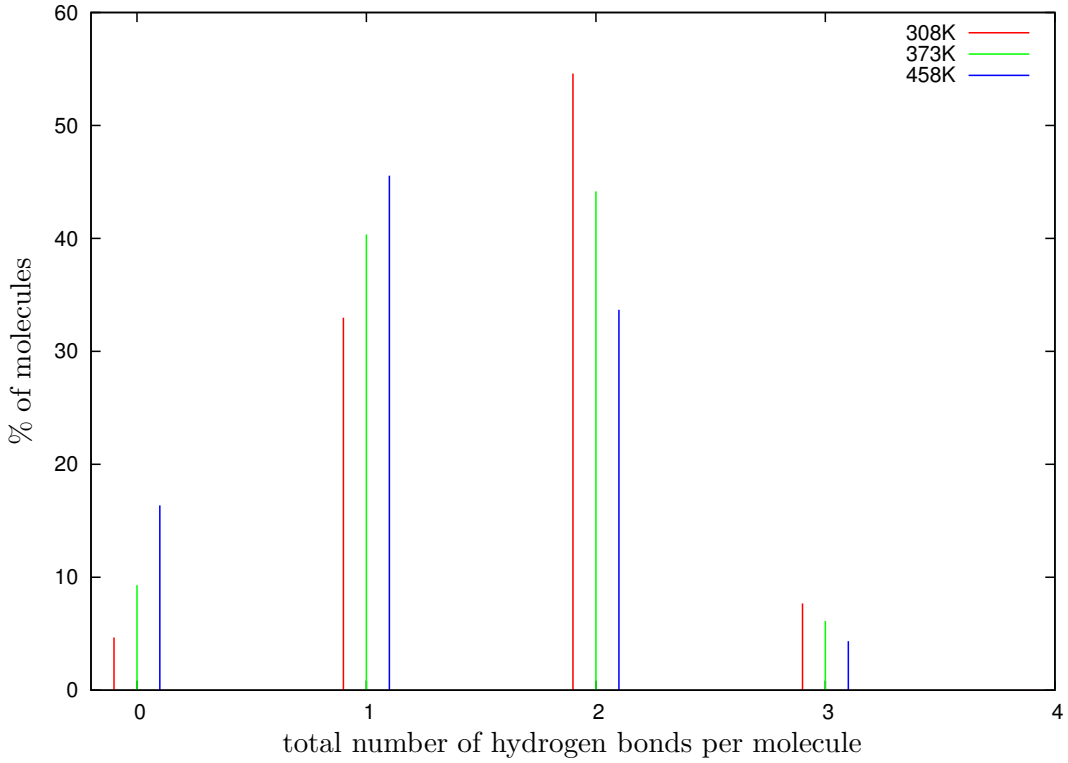

Figure 5.8: The percentage of NMA molecules forming 0, 1, 2, 3 or 4 hydrogen bonds with $N-H\cdots O$ angles greater than 140°.

of the system. This seems unlikely as it has been shown that the acceptance of a second hydrogen bond weakens the hydrogen bond between the carbonyl oxygen and its original partner[67]. However, this discovery was made using *ab initio* simulation where charge transfer is modelled explicitly. Classical simulations cannot model cooperativity other than implicitly through the tuning of the empirical forcefield parameters so this explanation cannot be ruled out. Alternatively, this feature could arise from the dynamics - the doubly hydrogen bonded configuration could be a midpoint in the process where the oxygen switches hydrogen bonded partners and therefore appear as a common feature at all temperatures in these time-averaged results.

5.4.7 Donor behaviour

The donor behaviour of NMA molecules changes as expected with increasing temperature and is shown in Figure 5.9(b). At 308.15K almost all of the amide hydrogens are donated to form hydrogen bonds in agreement with experimental studies of the vibrational spectra. Huang *et al* investigated the IR spectra of liquid NMA for the temperature range 303 to 343K and estimated that 14% of the $N-H$ groups do not form hydrogen bonds[62]. This is quite similar to the 17% calculated for hydrogen bonds

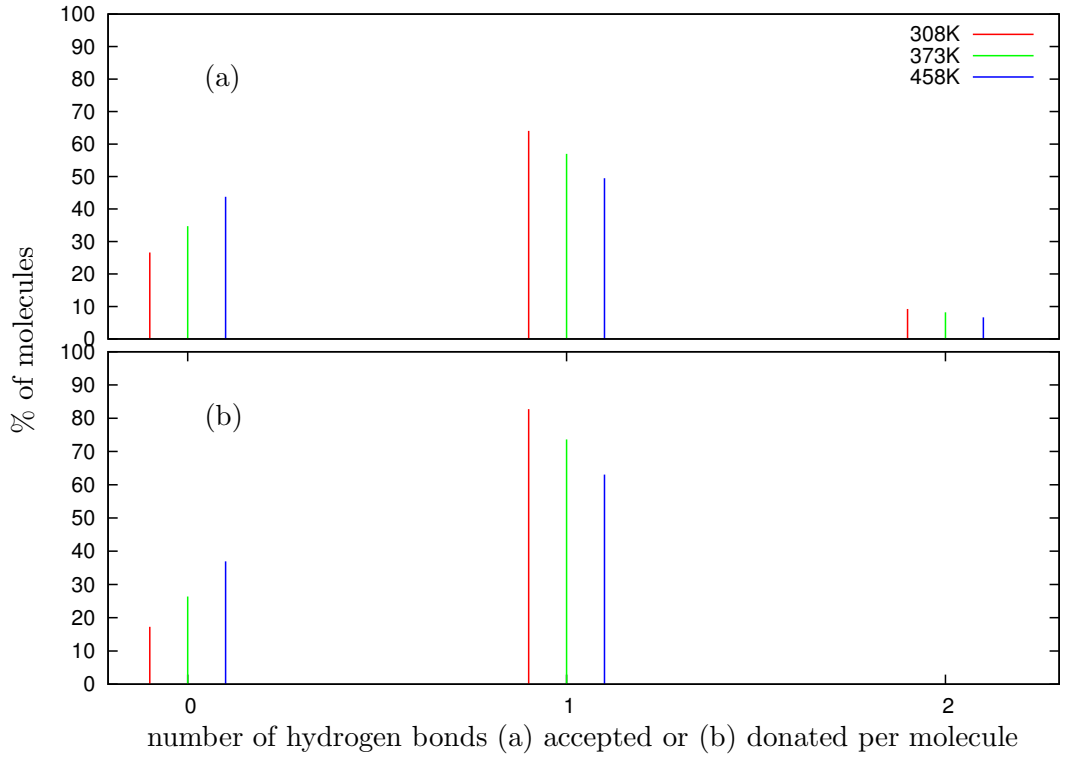

Figure 5.9: The percentage of NMA molecules which (a) accept 0, 1 or 2 hydrogen bonds with $N-H\cdots O$ angles greater than 140° and (b) donate 0 or 1 hydrogen bonds.

in the 308K simulation. The proportion of free NH groups increases with increasing temperature in both these simulations and the experimental studies[61, 62]. It can also be seen from Figure 5.9(b) that the hydrogen bond definition used eliminates almost all bifurcated contacts (where a single hydrogen is associated with two acceptor atoms.)

5.4.8 Summary of hydrogen bonding in NMA

From these results it is possible to infer the topology of the intermolecular hydrogen bonding structure. Figure 5.9(b) shows that even at the highest temperature studied the majority of molecules donate their amide hydrogen to form a hydrogen bond with another NMA molecule. At lower temperatures the proportion is much higher, about 80% at 308.15K. The NMA molecule has a total of three hydrogen bonding sites: the first is the amide hydrogen which can be donated; the second and third are accepting sites on the carbonyl oxygen. Thus there is an imbalance in the ratio of acceptor to donor sites and the question arises as to how the molecules arrange themselves in response to this imbalance. One can imagine the two extrema of this situation. There could be a 'fair' distribution of hydrogen bonds with each molecule donating exactly one and accepting exactly one. Alternatively, the opposite extreme would be the case where

every carbonyl oxygen accepts its maximum of two hydrogen bonds until there are no more donor atoms left, thus half the oxygen atoms in the system would be unable to accept any hydrogen bonds at all. Naturally the true situation falls somewhere between these two extremes. Indeed, even at the lowest temperature the full donor capacity of the system is not utilised and nearly 20% of the amide hydrogens do not participate in hydrogen bonding interactions. The most common situation is for the carbonyl oxygen to accept a single hydrogen bond although since a significant percentage do accept two, approximately a third of the oxygens in the system do not accept any hydrogen bonds at all.

If each molecule accepted and donated exactly one hydrogen bond the system would comprise an endless chain of NMA molecules. In reality, some of the molecules form no hydrogen bonds at all and others accept two, thus can be considered as 'branchpoints' in otherwise linear chains. Therefore it can be concluded that the medium-range static liquid structure of CHARMM22 NMA comprises a series of predominantly linear but also branched chains as well as isolated molecules. The CHARMM22 NMA molecule has a higher tendency to form branched chains than the OPLS model although this could be due to differences between the hydrogen bond criteria applied here and in reference [68]. The proportion of isolated molecules increases almost four-fold (from 4.6% to 16.9%) when the temperature is raised from 308K to 458K. Recent experimental evidence suggests that the break up of hydrogen bonded chains with temperature is not entirely linear but has a sudden transition at 393K[65].

5.4.9 Hydrogen bonded clusters

Using the hydrogen bond definition, the simulation trajectories can now be re-analysed by grouping molecules into clusters associated via hydrogen bonding. In this manner the number and size of clusters which form and disperse during the simulation run can be counted. Figure 5.10 illustrates the results of such an analysis and plots the cluster mole fraction $X_n = \frac{N_n}{N_{tot}}$ as a function of n, the number of molecules in the cluster, where N_n is the number of clusters of a particular size n and N_{tot} is the total number of clusters. Huang *et al* assume NMA molecules only form linear chains and therefore that the cluster population distribution can be easily described by the statistics of linear polycondensation[62]. In the framework of this theory the probability of forming a hydrogen bond is p and conversely the probability of not forming a hydrogen bond is $(1-p)$. The probability of selecting at random a hydrogen bonded chain of length n molecules is equal to the mole fraction X_n of n-mers in the distribution. This can

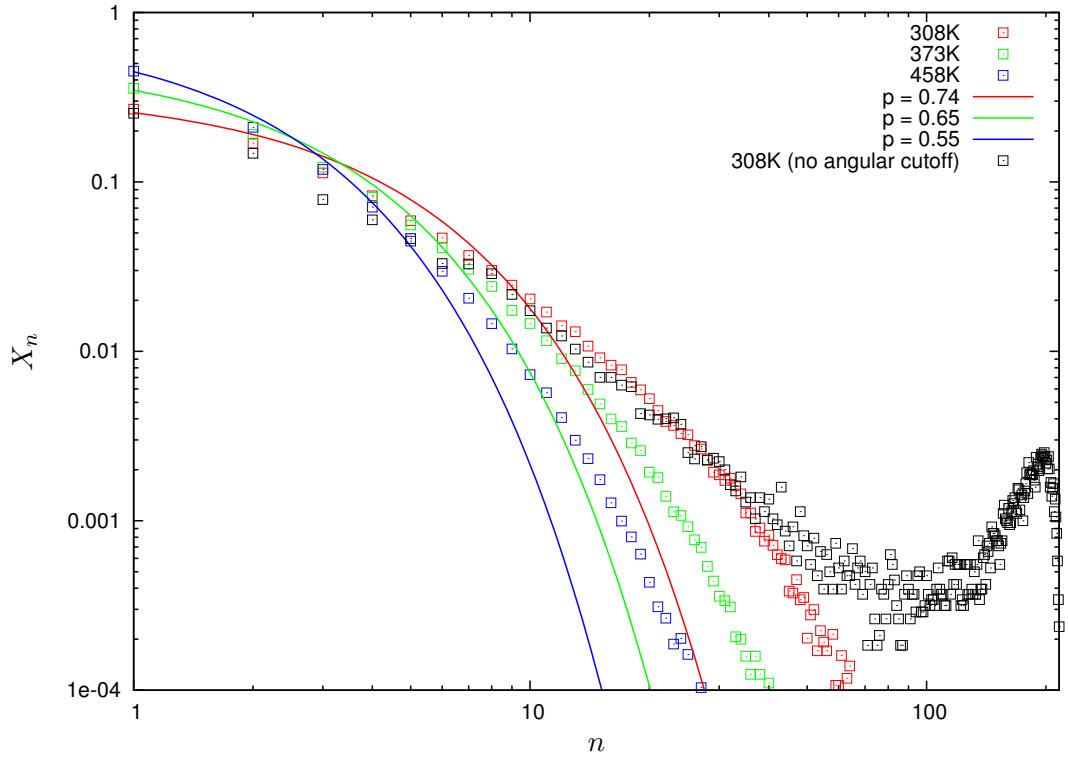

Figure 5.10: The population distribution of hydrogen bonded NMA clusters at 308, 373 and 458K. X_n is the cluster mole fraction defined in Equation 5.1 and n is the cluster size. The solid lines are best fits to data points of the same colour using Equation 5.1. The two sets of data at 308K illustrate the difference in cluster populations when no angular cutoff is applied (black) and when $\theta_{N-H\cdots O} > 140°$ (red). When no angular criterion is applied system size (216 molecules) hydrogen bonded networks form.

easily be written down as

$$X_n \equiv \frac{N_n}{N_{tot}} \;=\; P_n \equiv (1-p)p^{(n-1)} \tag{5.1}$$

since there are $(n-1)$ molecules in a linear n-mer which have formed a hydrogen bond and only 1 which has not. The best fits of this equation to the population distributions calculated from the simulations are plotted in Figure 5.10 from which it is clearly seen that the statistics of linear polycondensation do not describe NMA cluster distributions. In fact, the computed cluster distributions are biased towards larger cluster sizes for all three temperatures simulated. This can be attributed to the fact that CHARMM22 NMA forms branched as well as linear chains as found in the previous section. There may also be an additional effect in that it has been shown that hydrogen bonding in NMA is positively cooperative favouring the formation of longer chains[57] and this propensity may be implicitly included in the empirical forcefield parameters.

Populations of NMA clusters have been published by Akiyama who derived the distributions by calculating the enthalpies and entropies for oligomeric species on the basis of *ab initio* calculations and the ideal gas model[66]. Cluster populations have also been investigated by Ludwig *et al* who calculated partition functions for linear clusters of size one to nine molecules also using *ab initio* methods[70, 48]. These two studies give conflicting results, neither of which agree with the population distributions found in the classical simulations presented in this thesis. Figure 5.11(a) compares

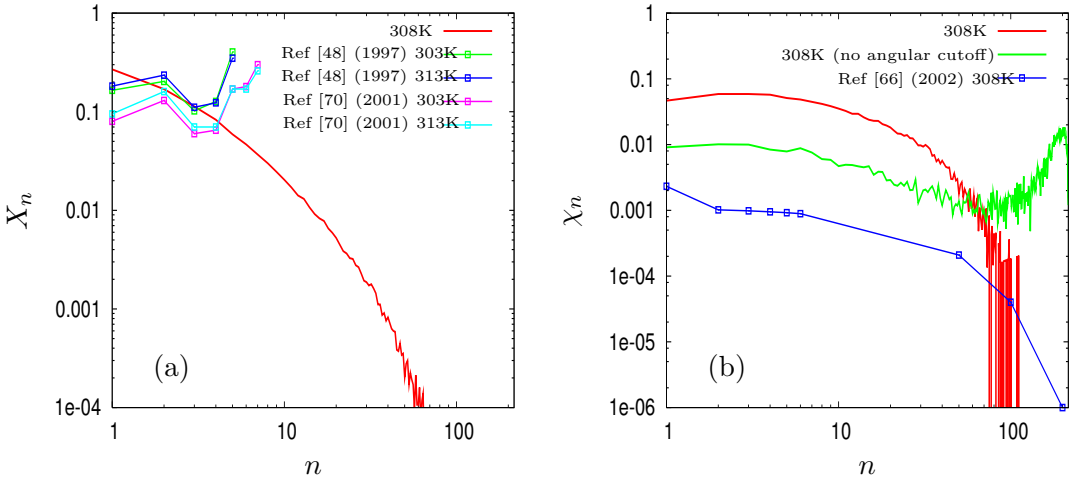

Figure 5.11: (a) compares the cluster mole fraction $X_n = \frac{N_n}{N_{tot}}$ at 308K with the population distributions calculated by Ludwig *et al*[70, 48]. Akiyama calculates the molecular mole fraction $\chi_n = \frac{M_n}{M_{tot}}$ where M_n is the number of molecules contained in clusters of size n and M_{tot} is the total number of molecules in the system. His results are plotted in (b) along with the corresponding distribution from the simulations both with and without an angular cutoff.

the cluster population distribution X_n of the 308K simulation with those calculated by Ludwig and co-workers at 303 and 313K. In their first paper only five linear species up to pentamers were considered[48] while in a later publication nonamers were considered, but did not survive thermodynamic calculations leading to a maximum chain length of seven[70]. From Figure 5.11(a) it is obvious that their population distributions depend much more strongly on the maximum chain length considered than upon temperature. Considering the classical simulation results it is very surprising that Ludwig found octamers and nonamers to be unimportant, especially when their exclusion leads to a peculiar distribution in which heptamers are the most prevalent species but oligomers one molecule longer do not exist. χ_n, the probability of choosing a molecule at random and finding that it belongs to a cluster of size n is plotted in Figure 5.11(b) for the classical simulation at 308K and the values published in reference [66] at 303K. 303K is only just above the melting point of NMA (301K) and this may partially explain why Akiyama's results show a bias towards very large cluster sizes. Unfortunately Akiyama's paper contains inconsistencies: Equation (7) appears to be incorrect and

even when correctly derived from the information given in the Appendix I have been unable to reproduce the exact values listed in the paper for the cluster distribution of NMA. Despite these differences in the detail, the general principle that NMA molecules form a wide range of cluster sizes, certainly not curtailed at seven molecules, is one that is strongly supported by the classical simulations.

Figure 5.10 also shows that the population distribution is dependent on the specific definition used for hydrogen bonding. When no angular cutoff is used "hydrogen bonds" form a network throughout the system and a peak is observed in the distribution just smaller than the system size (216 molecules). This peak near the system size is also observed when a larger simulation box of 60Å is used. However, when the definition of hydrogen bonding given in section 5.4.4 is applied the population distributions for both the 30Å and 60Å boxes become identical confirming that a box size of 30Å is sufficient to overcome finite size effects in the cluster distributions. Applying an angular cutoff reduces the maximum cluster size to less than 100 molecules even in the lowest temperature simulation. At 308K half of the molecules are contained within clusters of 10 or fewer molecules and three-quarters within clusters smaller than 23 molecules. Nonetheless, large clusters do form over the course of the simulation with the result that 10% of the molecules, averaged over all timesteps, belong to clusters with more than 36 molecules (16% of the system size) at 308K. Both the average and maximum size of cluster decrease as the temperature is raised. At 458K, on average, only 10% of the molecules are members of clusters larger than 10 molecules and the maximum cluster size is about 50 molecules. At 308 and 373K a randomly chosen molecule is most likely to belong to a dimer or trimer. At 458K the behaviour is different and molecules are most likely to be isolated.

The clustering algorithm identifies hydrogen bonded molecular clusters within which individual molecules can be interrogated to discover the number of hydrogen bonds which each forms. If a molecule only forms one hydrogen bond it must be an endpoint of the cluster. Figure 5.12 shows the distribution of the number of endpoints for different sizes of cluster and their variation with temperature. From this it can be seen that the vast majority of five molecule clusters have two endpoints and therefore are linear chains. Less than 20% of five molecule clusters are branched chains with three endpoints. However, as the temperature increases the likelihood of a five molecule cluster possessing a branch increases. The results for clusters containing 8, 10 and 15 molecules also plotted in Figure 5.12 show that this trend is also true for larger cluster sizes. A plausible explanation for this observation is as follows: firstly, at higher temperatures there will be more movement; secondly, chain-like structures may be more easily broken than denser, heavily branched structures. Large clusters may therefore

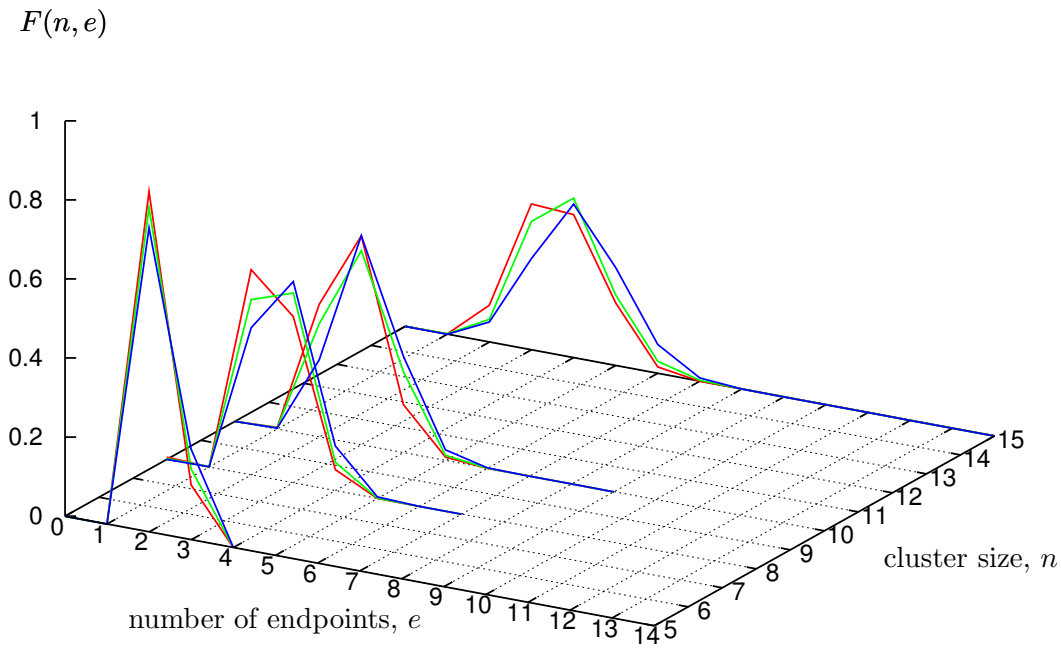

Figure 5.12: F(n,e) = the number of clusters with e endpoints and n molecules in the cluster normalised by the total number of clusters of size n. red = 308K, green = 373K, blue = 458K.

tend to break at their linear chain-like sections, leaving smaller, branched clusters thus explaining the trend towards more branched structures shown in Figure 5.12.

5.5 $g(r_{OO})$, $g_2(r_{O\cdots O}, \theta_{O\cdots O=C})$ and inter-chain contact

The oxygen - oxygen radial distribution function in liquid NMA shown in Figure 5.13 displays an unexpected wing at radial distances shorter than the first peak at 4.9Å . This first peak corresponds to the distance between oxygen molecules in the direction of the hydrogen bonded chain and as discussed previously in the context of amide hydrogen correlations in Section 5.3 this separation characterises the length of a single chain "link". This assignment only becomes more apparent when the radial distribution is split into two dimensions by considering correlations between the $O\cdots O$ vector and the angle it makes with the covalent bond of the carbonyl group, $O=C$. Figure 5.14(a) shows this radial-angular distribution with contour intervals starting at $g_2(r_{O\cdots O}, \theta_{O\cdots O=C}) = 1$. (Therefore this plot only shows regions which are over-structured relative to the long-range density.) From this it can be seen that the first peak in the radial distribution arises from two different angular peaks, one at $\theta_{O\cdots O=C} = 180°$ labelled A and the other at $\theta_{O\cdots O=C} = 0°$, labelled B. These peaks are both due to nearest neighbour contacts but in opposite directions along the hydrogen bonded chain. The second peak in $g(r_{OO})$ is at 9.5Å, just less than twice the distance

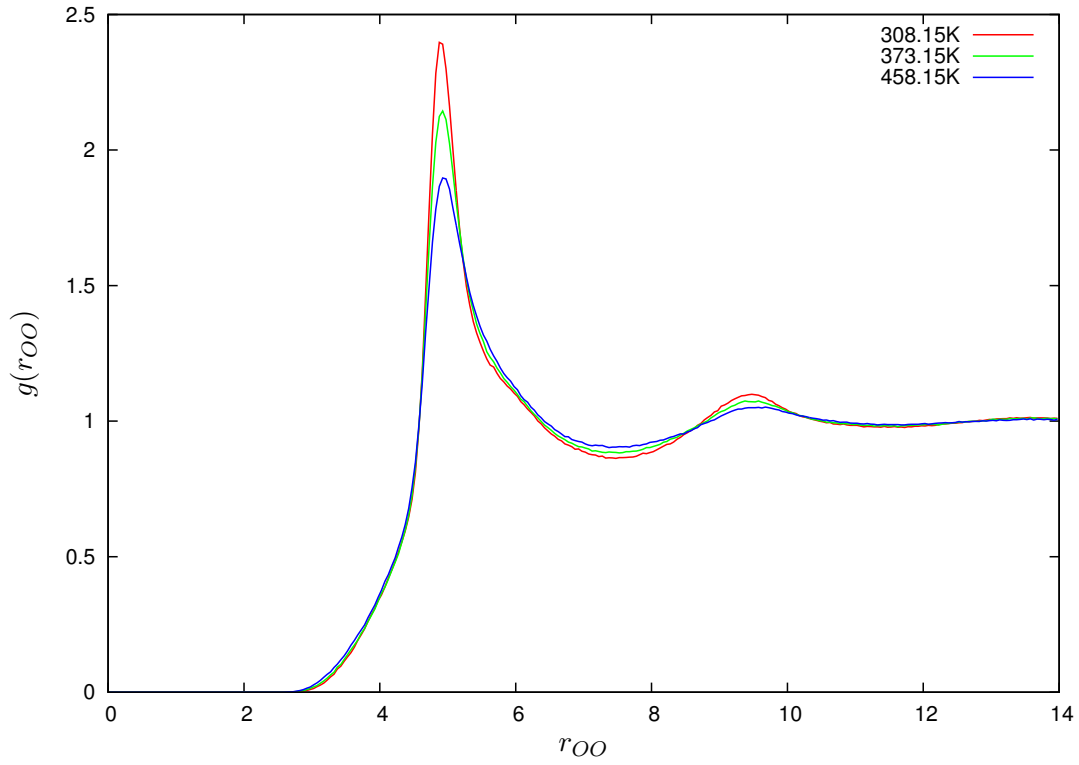

Figure 5.13: $g(r_{OO})$ at 308K, 373K and 458K.

of the first peak, and therefore corresponds to second nearest neighbours along the chain. The angular plot confirms this since the peaks at 9.5Å are centred on 0° and 180° (labelled C in Figure 5.14) and therefore correspond to the two possible chain directions as before.

5.5.1 Low r contributions

In order to understand the contribution to $g(r_{OO})$ at low r it is necessary to plot the joint radial-angular distribution with contour intervals starting below 1 as shown in Figure 5.14(b). From this it is apparent that the contribution at low r arises from configurations where $\theta_{O\cdots O=C}$ lies between 60° and 120° and therefore this feature is perpendicular to the chain direction. The coordination numbers listed in Table 5.4 support this statement as they show that integrating $g(r_{OO})$ up to the position of the first peak already gives $n(r) = 2$. If the only contributions were from neighbouring molecules along the chain direction the value of $n(r)$ at this point could be expected to be equal to 1 with each chain neighbour contributing $\frac{1}{2}$. As the actual value is greater than this there must be a contribution from molecules outwith the hydrogen bonded linear chain. Thus the low r wing of $g(r_{OO})$ can be attributed to interactions between molecules in neighbouring chains rather than between molecules in the same chain.

T (K)	1^{st} peak position (Å)	n(r)	1^{st} min position (Å)	n(r)
308.15	4.875	2.08	7.475	12.96
373.15	4.925	2.11	7.525	12.44
458.15	4.925	1.91	7.675	12.48

Table 5.4: Extrema positions and cumulative number densities for the 1^{st} peak in $g_{OO}(r)$.

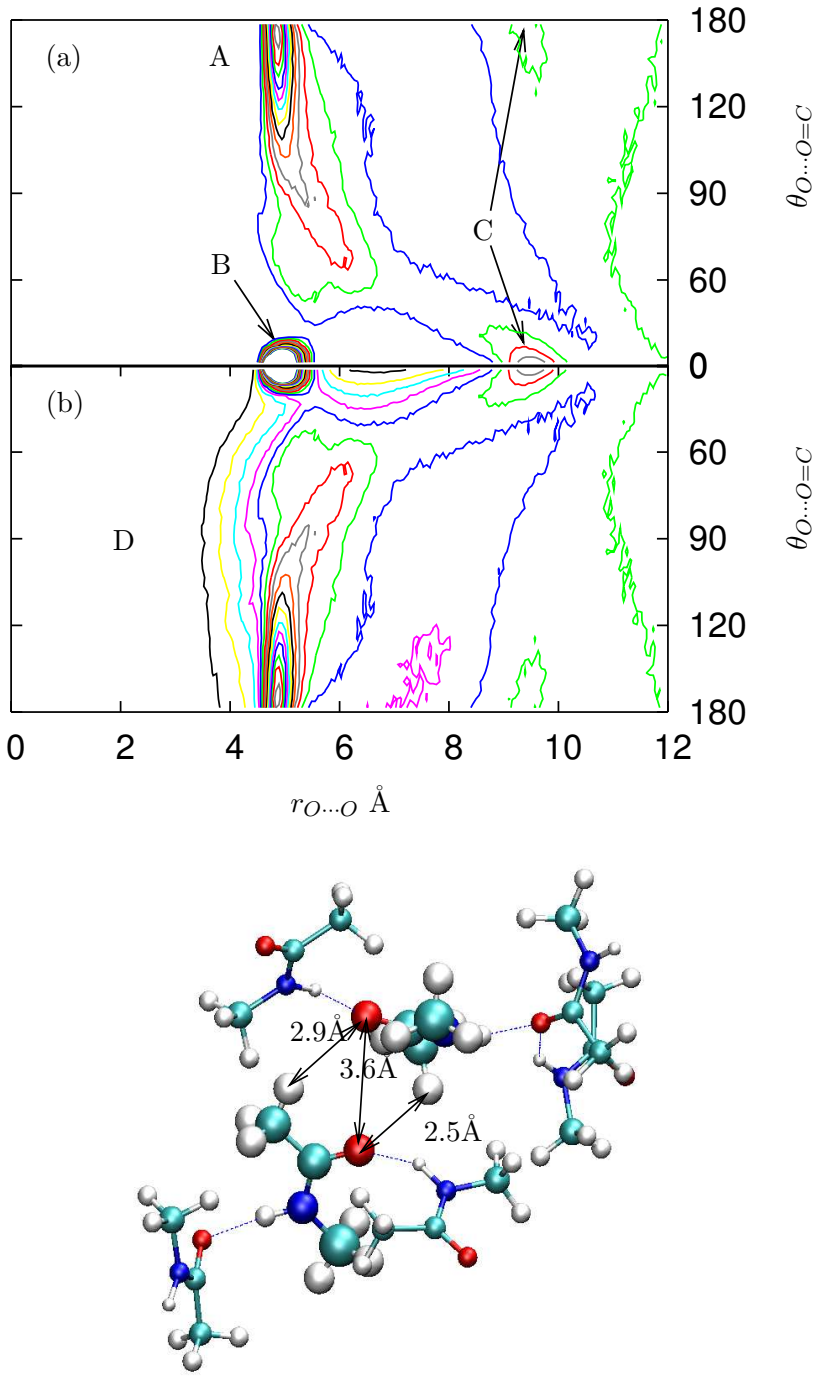

Figure 5.14: (a) and (b) both plot contours of $g_2(r_{O\cdots O}, \theta_{O\cdots O=C})$ which is the radial angular distribution obtained by correlating oxygen - oxygen separation, $r_{O\cdots O}$ with the angle $\theta_{O\cdots O=C}$. In both Figures the $g_2(r_{O\cdots O}, \theta_{O\cdots O=C}) = 1$ is blue. This is the first contour level in the upper plot while in the lower Figure (b) the first contour is $g_2(r_{O\cdots O}, \theta_{O\cdots O=C}) = 0.001$ (orange). In both cases contour levels increase in 0.2 increments up to a value of 5. The simulation snapshot shows two neighbouring NMA chains, the upper comprising four molecules and the lower three. The central molecules approach each other closely with an oxygen - oxygen separation of only 3.6Å and contribute to the low r wing of $g_2(r_{O\cdots O}, \theta_{O\cdots O=C})$ shown in (b). Labels A, B, C and D refer to interactions described in the text.

In fact, a similar feature was observed in Section 5.2 in the interactions between the amide and carbonyl groups of NMA molecules and from these results it can be concluded that the most likely interchain distance is about 5Å although a wide range of separations appear in the simulations with a minimum interchain distance of just over 3Å. It is quite surprising that oxygen atoms in parallel chains approach each other this closely as within the CHARMM22 model each is given a negative charge and repels similarly charged atoms in accordance with Coulomb's Law. The energy penalty associated with such close approaches may be alleviated somewhat by associations between oxygen atoms and methyl hydrogens in what could be termed "weak" hydrogen bonds. The simulation snapshot in Figure 5.14 shows a typical interchain oxygen - oxygen close contact with which are associated two oxygen - methyl hydrogen separations of less than 3Å . The electric potential at a point due to a system of point charges is equal to the sum of the potentials due to each individual point charge:

$$\phi(q_i, r_i) = \frac{1}{4\pi\epsilon_\circ} \sum_i \frac{q_i}{r_i} \tag{5.2}$$

From this it is possible to calculate that the methyl associations shown in Figure 5.14 reduce the magnitude of the electric potential due to the close proximity of another oxygen by 20% (for $r_{OM} = 2.9$Å) and 25% (for $r_{OM} = 2.5$Å).

5.5.2 Weak or improper methyl hydrogen bonding

In this section the oxygen - methyl hydrogen associations introduced in the previous section are explored and it will be shown that they arise not only from interactions with neighbouring chains but also from specific intra-chain structural motifs. A closer look at Figure 5.14 in Section 5.5 reveals that both the distances labelled are between oxygen atoms and hydrogens belonging to the methyl group covalently bonded to the carbonyl carbon known as H_C (see Figure 2.5 for atomic nomenclature). Turning to the radial distribution functions shown in Figure 5.15 it is apparent that this observation must be generally true since $g(r_{OH_C})$ has a peak at 2.9Å whereas $g(r_{OH_N})$ has none. From the radial-angular distribution $g_2(r_{O\cdots H_C}, \theta_{O\cdots H_C - C_C})$ shown in Figure 5.16(a), it is clear that the first peak in $g(r_{OH_C})$ has a reasonably well defined angular range centred on about 135°, labelled A in this Figure. The angular localisation suggests hydrogen bonding, even though with true hydrogen bonding the angular range is expected to be centred on 180°. The hydrogens belonging to the amide methyl group, known as H_N are generally to be found farther away from carbonyl oxygens and $g_2(r_{O\cdots H_N}, \theta_{O\cdots H_N - C_N})$ shown in Figure 5.16(b), reveals a particular lack of configurations with $120° < \theta_{O\cdots H_N - C_N} < 180°$, the range expected for hydrogen bonding.

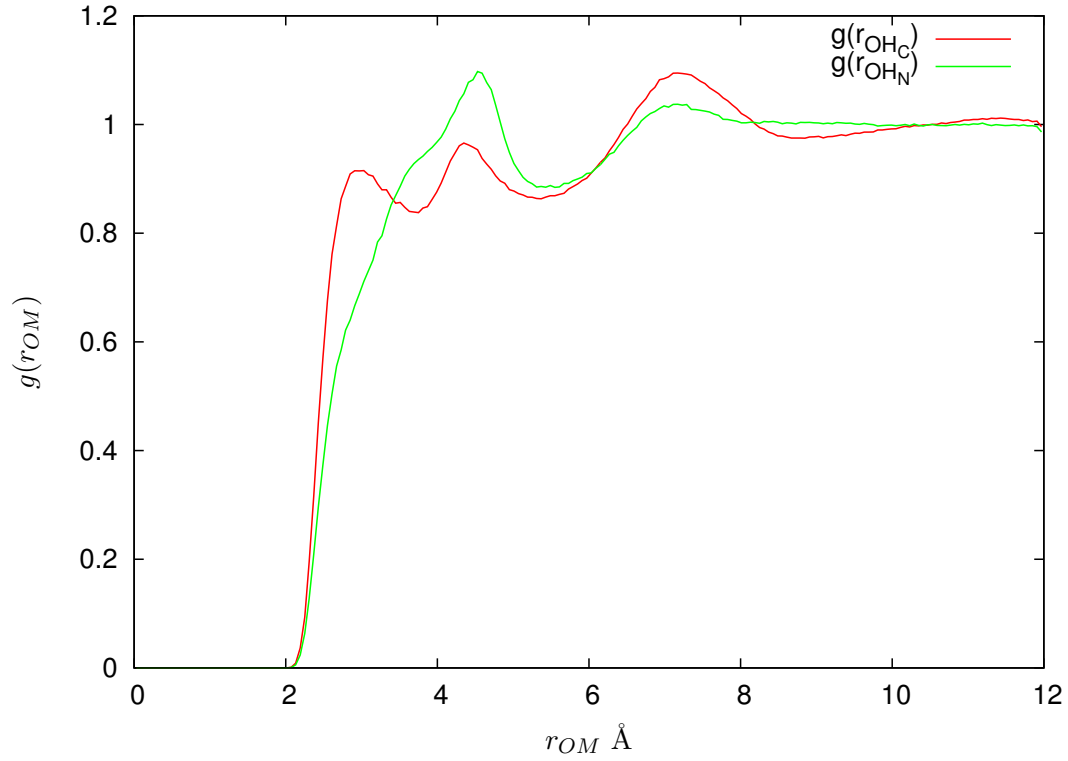

Figure 5.15: Radial distribution functions $g(r_{OM})$ for interactions between the NMA oxygen O and methyl hydrogen M, where M belongs either to the carbonyl methyl group H_C or the amide methyl group H_N, at 308K.

At first glance the NMA molecule appears to be roughly symmetric but this symmetry is broken by the positive and negative charges assigned to the amide hydrogen and carbonyl oxygen respectively. As a result the two methyl groups occupy quite distinct chemical environments and this is the origin of the differences between their radial and angular distribution functions.

Why is it in general the carbonyl methyl hydrogens which approach oxygen atoms more closely than those belonging to the amide methyl group? The simplest explanation is that the carbonyl methyl group is closer to the NMA oxygen, separated from it by only two covalent bonds, whereas the amide methyl group is separated by three. When two oxygen atoms approach each other closely, by necessity the H_C atoms belonging to one molecule are brought into close contact with oxygen of the other. In the CHARMM model of NMA all the methyl hydrogens are assigned the same value of positive charge ($0.09e$) and so there is no reason to suppose any more complex explanation than this. However, inter-chain associations do not give the full story behind Figures 5.15 and 5.16. *Intra*-chain motifs make a contribution which can also be explained with reference to

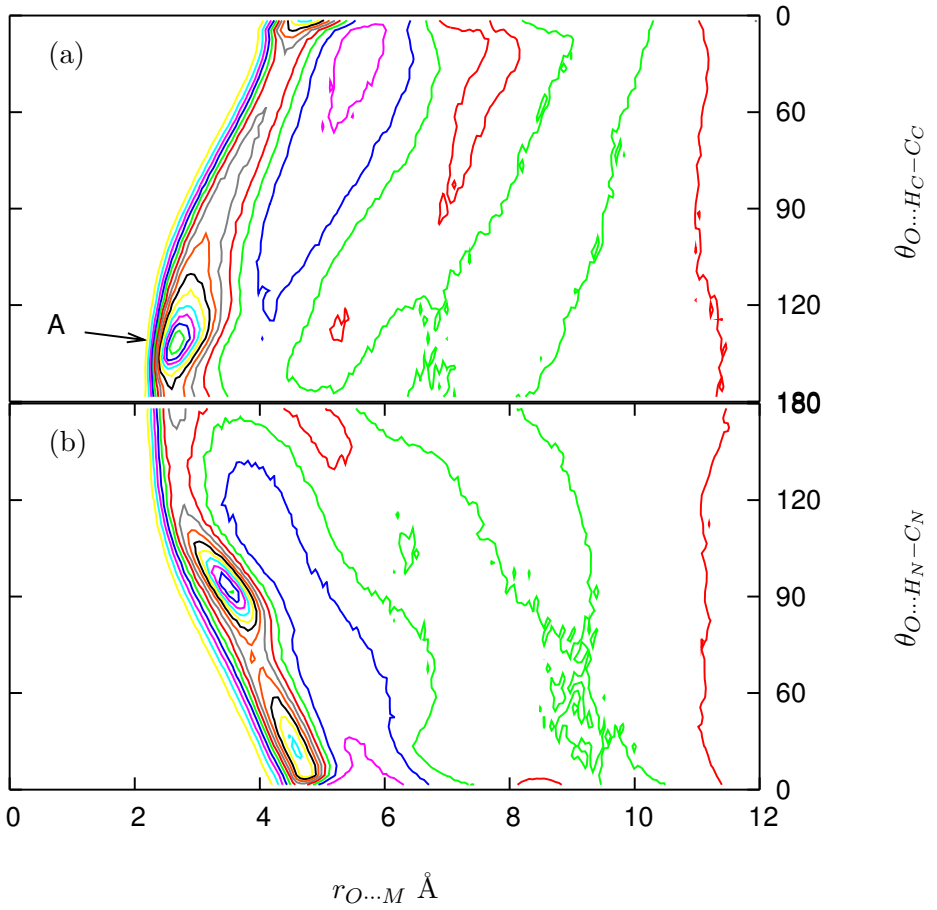

Figure 5.16: Radial angular distributions for (a) $g_2(r_{O\cdots H_C}, \theta_{O\cdots H_C-C_C})$ and (b) $g_2(r_{O\cdots H_N}, \theta_{O\cdots H_N-C_N})$. In the x-axis label M refers to either H_N or H_C. Contour levels begin at 0.2 (turquoise), increasing in increments of 0.2 to 5. The $g_2(r,\theta) = 1$ contour is green, the blue and red contours are $g_2(r,\theta) = 0.8$ and 1.2 respectively.

the geometric shape of the molecule.

The C-N-C-C backbone of the NMA molecule resembles a planar zig-zag. This intrinsic shape of the NMA molecule means that when a hydrogen bond is formed by the amide group the carbonyl methyl group usually points towards the accepting molecule. This can be clearly seen in the interaction between the central and right-hand molecules illustrated in Figure 5.17. The radial - angular distribution corresponding to this methyl group is shown in Figure 5.16(a) where r_{OH_C} is plotted against the putative hydrogen bond angle $\theta_{OH_CC_C}$. When decomposed into the angular domain, $g_2(r_{OH_C}, \theta_{OH_CC_C})$ forms a diagonal with shorter separations corresponding to $O\cdots H_C - C_C$ angles closer to 180° and longer distances to angles approaching 0°. The first diagonal between 2 and 6Å corresponds to the carbonyl methyl group of a molecule hydrogen bonded to the central oxygen atom from which the distribution is calculated but also includes

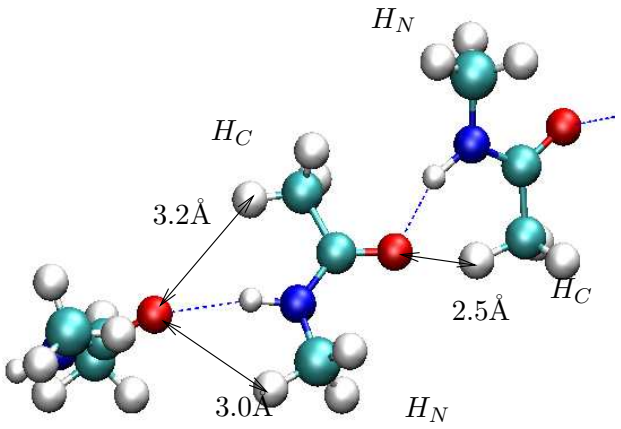

Figure 5.17: Simulation snapshot of three molecules from an NMA pentamer showing the intrachain associations contributing to the peaks in Figure 5.16. Hydrogens belonging to the methyl group bonded to the carbonyl carbon are labelled H_C while those belonging to the group bonded to the nitrogen are labelled H_N. The angle made by the 2.5Å $H_C \cdots O$ separation with the $H_C - C_C$ covalent bond is 133° and therefore this particular structure contributes to the 1^{st} peak in $g_2(r_{H_C\cdots O}, \theta_{O\cdots H_C - C_C})$ illustrated in Figure 5.16(a). In general the H_C methyl group is closer to the accepting oxygen but this is by no means always the case as is shown by the 3.0Å(H_N) and 3.2Å(H_C) associations. This snapshot epitomises the message of Figure 5.16 which is that although there are some favoured structural configurations the flexibility of NMA hydrogen bonded chains is such that a wide range of OM distances and OMC_M angles are permitted.

contributions from neighbouring chains. The second diagonal (broken red contour) partly corresponds to interactions with a second neighbour in the same chain direction but again also includes other contributions from other secondary or tertiary neighbours. The continuity in both the radial and angular directions indicates that a continuum of $O \cdots H_C$ separations and $O \cdots H_C - C_C$ angles appear in the simulations. This is mainly due to averaging over the three indistinguishable methyl hydrogens. Within the diagonal of Figure 5.16(a) there is a localised peak labelled A at 2.9Å and 135° corresponding to the first peak of the radial distribution function $g(r_{OH_C})$. The methyl - oxygen interaction labelled "2.5Å" in Figure 5.17 falls precisely within this peak and motivates an structural explanation for this feature of $g_2(r_{OH_C}, \theta_{OH_CC_C})$. When a hydrogen bond forms between the amide and carbonyl groups of neighbouring NMA molecules the carbonyl methyl group tends to orient itself such that one of its methyl hydrogens points towards the accepting oxygen. In a sense this methyl hydrogen - oxygen interaction "pins" the entire methyl group. It is questionable as to whether these intra-chain "pinning" interactions can truly be termed hydrogen bonds as they are essentially a secondary effect and their geometry is mainly dependent on that of the strong amide-carbonyl hydrogen bond. That said, they do satisfy the angular and distance constraints often used to define weak or improper hydrogen bonds[12]. This structural motif has also been identified by Vargas *et al* in electronic structure

calculations of *cis* and *trans*-NMA dimers[73]. It was found to be the only stable form of *trans*-NMA dimer at the MP2 level of theory with the augmented correlation consistent basis set aug-cc-pVTZ. According to their calculations the $O \cdots H_C - C_C$ interaction stabilises the dimer arrangement formed by the amide-carbonyl hydrogen bond by 1.1kcal/mol.

The contour lines of $g_2(r_{OH_C}, \theta_{OH_C C_C})$ diverge slightly as $\theta_{OH_C C_C} = 180°$ is approached which suggests that at least some carbonyl methyl hydrogens may form linear hydrogen bonds though whether these are primarily intra- or inter-chain configurations cannot be said without further investigation.

A comparison between (a) and (b) of Figure 5.16 shows that while the distance and angular ranges occupied are similar the features within these ranges are quite distinct. As in Figure 5.16(a), Figure 5.16(b) which shows $g_2(r_{OH_N}, \theta_{OH_N C_N})$ also displays a slight divergence of contour lines as $\theta_{OH_N C_N} = 180°$ is approached which suggests that some amide methyl hydrogens form weak hydrogen bonds with NMA oxygen atoms. In general however there is a distinct lack of structural configurations with $\theta_{OH_N C_N}$ in the range [120°, 180°]. The amide methyl hydrogens are generally found further away from the oxygen than the carbonyl methyls and are more evenly distributed through the angular range. As before this can be attributed to the inherent shape of the trans-NMA molecule which often ensures that when a hydrogen bond is formed the amide methyl group is pointed away from the accepting molecule. However, the orientational flexibility of the hydrogen bond interaction, not to mention that of the CHARMM model of NMA, is such that separation distances between the accepting oxygen and both sets of methyl hydrogens range between 2.5Å and 5Å over a wide angular range as illustrated by the hydrogen bonded interaction between the central and left-hand molecules of Figure 5.17.

5.6 Dynamics

Liquids are not static but in perpetual motion therefore no complete characterisation of the liquid state is possible without consideration of the system dynamics. In this section NMA self-diffusion and hydrogen bond dynamics are investigated as a function of temperature. Investigation of the system at 308K also provides baseline data against which to compare the dynamical behaviour of NMA molecules in aqueous solution.

5.6.1 Self-diffusion

Self diffusion coefficients are readily calculated from the mean square displacement plots shown in Figure 5.18 by means of the Einstein relation as described in Chapter 3, Equation 3.23. This method yields self diffusion coefficients of 0.06, 0.17 and 0.41

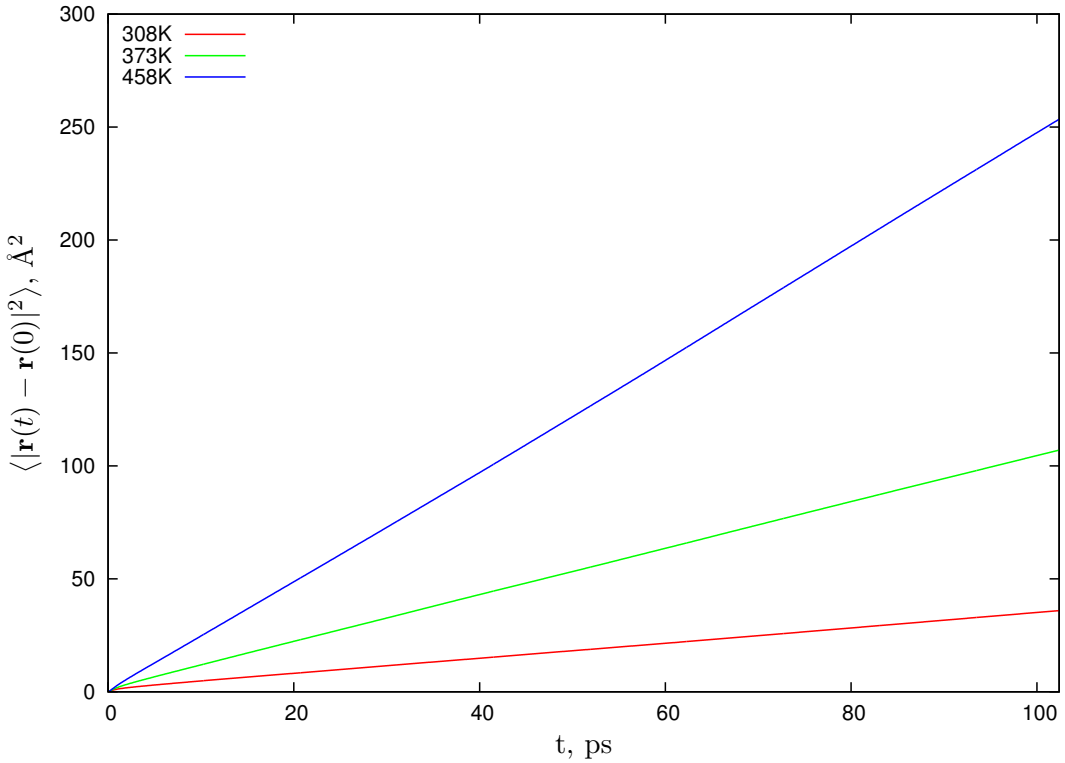

Figure 5.18: Centre of mass mean square displacement $\langle|\mathbf{r}(t)-\mathbf{r}(0)|^2\rangle$ for the CHARMM22 NMA model at 308K, 373K and 458K

Å^{2}ps^{-1} for NMA at simulation temperatures of 308K, 373K and 458K respectively. The experimental value for NMA self-diffusion at 308K is 0.0411Å^{2}ps^{-1}[130] which is 0.02Å^{2}ps^{-1} lower than the simulation result. A NMR study published in 2000 using the pulsed field gradient spin echo method (PFG-NMR) investigated diffusion in NMA over a range of temperatures and pressures[131]. The results at ambient pressure are shown in Figure 5.19 where they are compared with the simulation results across the same temperature range. It is common for forcefields to predict faster self-diffusion than is found experimentally and the same was true for the water models investigated in Chapter 4. The value obtained for CHARMM22 NMA gives equally good agreement with experiment as the closest of the water models, SPC/E. Self-diffusion for the 308K simulation was calculated by splitting a 1ns trajectory into four parts and calculating diffusion coefficients for each separately giving a mean value of 0.057 and standard deviation 0.003Å^{2}ps^{-1}. This mean agrees with the value obtained by analysing the

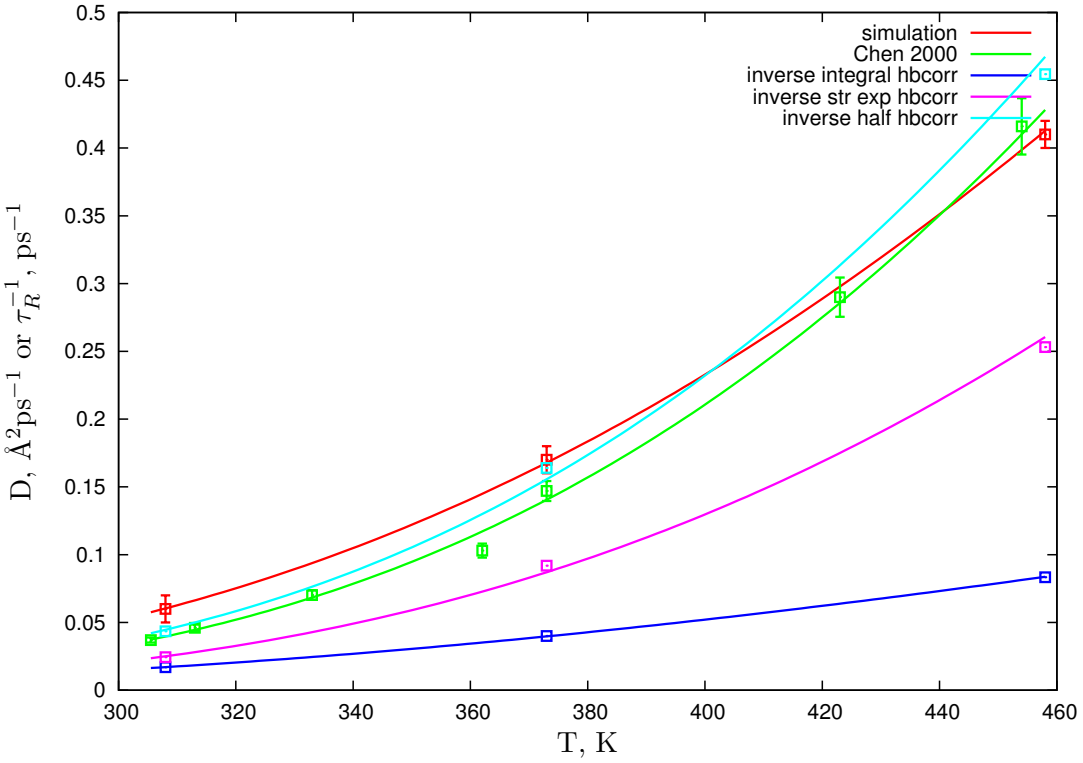

Figure 5.19: Self diffusion coefficients calculated from the CHARMM22 NMA simulations at 308K, 373K and 458K compared with the experimental results of Chen *et al*[131]. Inverse relaxation times calculated from hydrogen bond autocorrelations as detailed in the following section are also plotted. The lines are fits to the datapoints assuming an Arrhenius dependence upon temperature.

trajectory as a whole. Standard deviations calculated in this way for the entire suite of simulations give values ranging between 0.001 and 0.01Å^{2}ps^{-1} and since it is best to err on the side of caution all calculated diffusion coefficients are quoted to two significant figures with an error of ± 0.01Å^{2}ps^{-1}. A more confident estimate of the error could be made using the block averaging method of Flyvberg and Petersen[132] but given time constraints is beyond the scope of this work. Both the experimental data and the simulation results are well fitted by assuming an Arrhenius dependence upon temperature. These fits give activation energies for the self diffusion process of 190 ± 5meV (154 ± 4cm^{-1} $= 18.4 \pm 0.5$kJmol^{-1}) for the experimental results and 156 ± 2meV (126 ± 2cm^{-1} $= 15.0 \pm 0.2$kJmol^{-1}) for the simulations. Self-diffusion in liquid NMA will be further discussed in Chapter 7 where the results of quasielastic neutron scattering experiments are presented.

5.6.2 Hydrogen bond dynamics

The history independent hydrogen bond autocorrelation function, $c(t)$, defined by Equation 3.25 in Chapter 3 is plotted in Figure 5.20 for NMA at three different temperatures. Integrating this function to obtain a relaxation time gives values of

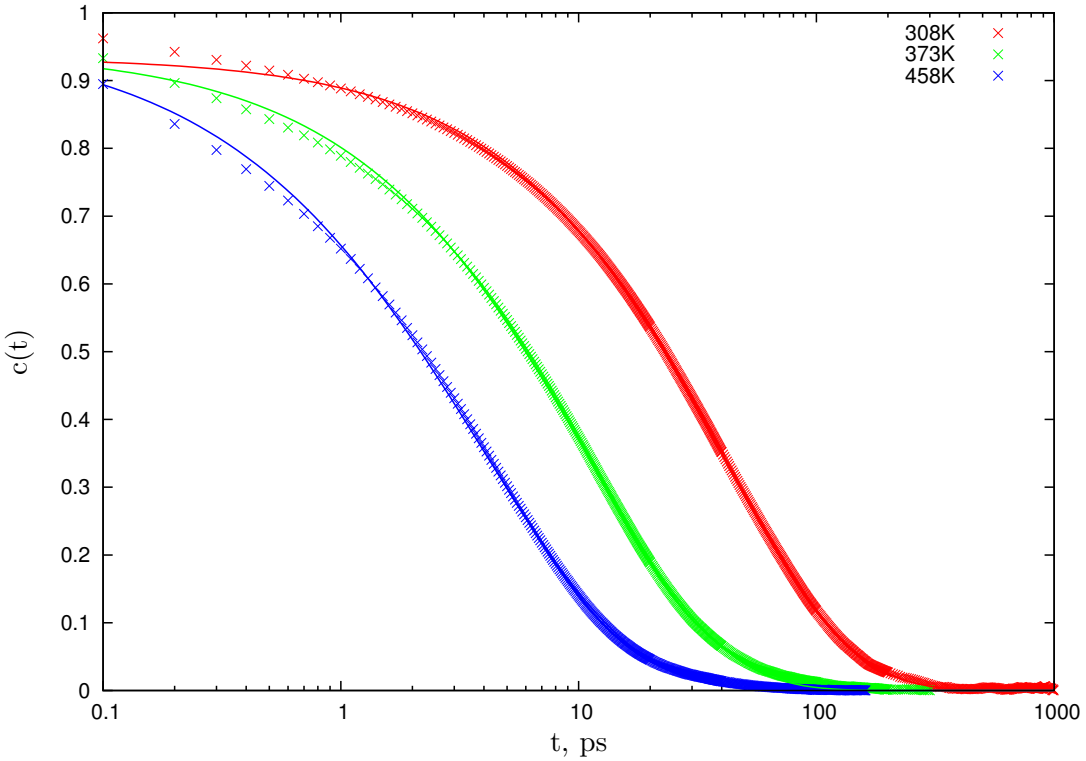

Figure 5.20: History independent hydrogen bond autocorrelations for CHARMM NMA at 308K, 373K and 458K (fits using stretched exponentials)

59ps, 25ps and 12ps respectively for 308K, 373K and 458K. As for water this correlation function can be fitted using a stretched exponential and fits for the three temperatures are also plotted in Figure 5.20. The values of the fitting parameters are listed in Table 5.5. According to this definition of the relaxation time the time over which hydrogen bonds are autocorrelated decreases by a factor of 10 over the temperature range 308K to 458K. It is clear that the formation of hydrogen bonds will impede the diffusion of

T (K)	$A_\circ$	β	τ_R
308	0.934±0.001	0.809±0.001	41.01±0.07
373	0.944±0.002	0.757±0.003	10.88±0.04
458	0.967±0.003	0.691±0.003	3.95±0.02

Table 5.5: Fitted parameters and errors in the fit of the simulation data to equation 3.26.

NMA molecules and Figure 5.19 shows that the temperature dependence of the self-

diffusion coefficient is very similar to that of the relaxation time particularly when calculated by fitting the autocorrelation function to a stretched exponential or taking the half-life. The half-life of $c(t)$, the time for which the autocorrelation function drops to half its initial value, is 23.0, 6.1 and 2.2ps for the 308K, 373K and 458K simulations respectively. The temperature dependence of each of the three "definitions" is well fitted by an Arrhenius relation as illustrated in Figure 5.19. However, the integral method of calculating the activation energy yields a lower result that that of the other two methods as can be seen from Table 5.6. This is quite probably due to errors in the calculation introduced by poor statistics in the long time limit which is a known problem with this method. The similarity of the activation energies calculated for the hydrogen bond autocorrelations with those calculated for the simulated and experimental diffusion coefficients indicates, as one would expect, that the two properties are interdependent. The relaxation times derived from history independent autocorrelation functions do

definition E_a	meV	cm^{-1}	$kJmol^{-1}$
integral	129 ± 1	104.1 ± 0.8	12.5 ± 0.1
stretched exponential	190 ± 7	153 ± 6	18.3 ± 1.3
half life	190 ± 7	153 ± 6	18.3 ± 0.7
D (simulation)	156 ± 2	126 ± 2	15.0 ± 0.2
D (expt[131])	190 ± 5	154 ± 4	18.4 ± 0.5

Table 5.6: Activation energies calculated from the Arrhenius fits in Figure 5.19 to the data sets corresponding to the different definitions of the times calculated from the hydrogen bond autocorrelations and for simulated and experimental diffusion coefficients.

not define hydrogen bond lifetimes as no count is made of how often the hydrogen bond breaks and reforms between the initial and final times of the autocorrelation. Instead, these values quantify timescales over which two molecules remain sufficiently close to one another so as to be able to participate in hydrogen bonding. At 308K, seven degrees above the melting point, correlation times are long, easily lasting for over 50ps. As the temperature increases correlation times decrease until at 458K, where molecules are most likely to be isolated monomers, those molecules which do manage to form hydrogen bonds rarely maintain the association for longer than 10ps.

Correlation times calculated by the history independent method are expected to be several times larger than those calculated by the history dependent method. Therefore the results presented above at least do not contradict the following simulation and experimental results. A hydrogen bond lifetime of 12ps for NMA molecules hydrogen bonded to methanol has been calculated using the history dependent method[60]. In this calculation hydrogen bonds lasting less than 2ps were discounted and bonds which broke and reformed in less than 2ps were considered to be unbroken. Using this

hydrogen bond definition the authors achieve good agreement with a lifetime of 10 to 15ps derived from their room temperature time-resolved 2D-IR experiment. The collective reorientation time of liquid NMA just above room temperature has been found to be approximately 19ps by OHD-OKE spectroscopy[65]. This value is similar to that calculated for the NMA-methanol hydrogen bond lifetime and as pointed out by the authors of this paper it is reasonable to suppose that the rotational-diffusive motion of NMA molecules is governed by the breakage of hydrogen bonds. To the best of my knowledge there are no other experimental or theoretical predictions of hydrogen bond lifetimes in liquid NMA with which to compare the results of the computer simulations.

5.7 Discussion

Crystalline NMA is composed of vertically stacked planes of parallel linear chains of NMA molecules hydrogen bonded through the amide and carbonyl groups. When the solid melts at just above room temperature these linear hydrogen bonded chains remain the dominant medium range structure but are joined by a variety of other features not present in the crystal. While in solid NMA each molecule forms exactly two hydrogen bonds, one accepted and one donated, in liquid NMA some carbonyl oxygens accept two hydrogen bonds leading to the existence of branched chains. Additionally, the disordered liquid structure allows for closer contact between oxygen atoms than occurs in the crystal. This is manifested in an unusual wing in the oxygen-oxygen radial distribution at low r which cannot be due to interactions between oxygen atoms within the same hydrogen bonded chain. Therefore these are inter-chain contacts and provide evidence for close packing of parallel hydrogen bonded chains[54]. Corrobation of this statement comes from amide - carbonyl group correlations.

Hydrogens belonging to the methyl group bonded to the carbonyl carbon atom show directionality in their interactions with the oxygen atom suggestive of hydrogen bonding. This interaction which could be termed weak hydrogen bonding appears to be associated with both inter- and intra-chain structural configurations. The intra-chain contribution is specifically associated with the hydrogen bond formed between the oxygen and amide hydrogen in agreement with *ab initio* results[73].

The cluster population distributions show that there is no average size of hydrogen bonded NMA cluster. Instead, cluster sizes are distributed across a wide range of up to 100 molecules at 308K and are shown to be biased towards larger clusters when compared to theoretical predictions of the population distribution of linear non-cooperative chains. In this respect the use of the classical molecular dynamics simulation technique is advantageous as it permits the simulation of larger systems

than *ab initio* methods which are restricted to small clusters. Previously, it has been argued that NMA oligomers containing more than seven molecules do not exist based upon calculations of the thermodynamic properties of oligomers less than nine molecules in length[70]. The classical simulations presented here contradict those results, which yielded a surprising cluster population distribution in which seven molecule oligomers were the maximum and also the most frequent cluster size. The population distributions calculated from the classical simulations predict a more rational dependence upon cluster size in which the frequency tails off to zero showing that the maximum cluster size even at 308K is approximately 100 molecules less than the system size. Consequently the population distribution obtained is independent of the number of molecules in the simulation for this and larger system sizes. However, one drawback of the analysis technique used to count clusters is that the precise quantitative population distribution depends upon the criteria used for the hydrogen bonding definition. DFT calculations predict that the frequency shift induced by cooperative hydrogen bonding will converge to a constant before the number of molecules in an NMA oligomer reaches ten[47]. This result indicates that spectroscopy will be unable to confirm or deny the existence of the larger clusters predicted by the classical simulations.

As the temperature of the simulations is increased the hydrogen bonded clusters break up: at 308K and 373K randomly selected molecules are most likely to belong to dimers or trimers whereas at 458K a randomly chosen molecule is most likely to be a monomer. A comparison of clusters of the same size across the three temperatures investigated reveals that clusters become more branched as the temperature increases. This suggests that as the system gains more kinetic energy large clusters break up at linear sections rather than at branchpoints. Over the temperature range investigated the self diffusion coefficient increases from 0.06 to $0.41 \text{Å}^2 \text{ps}^{-1}$ in agreement with PFG-NMR measurements and exhibits a similar Arrhenius temperature dependence to relaxation times derived from the history independent hydrogen bond autocorrelation function c(t). The activation energies calculated for these processes are quite reasonable, falling within the range expected for intermolecular vibrational modes such as hydrogen bond stretching.

Chapter 6

Aqueous *N*-methylacetamide

In this chapter aqueous NMA is studied across its full concentration range. This allows access to two important limiting cases: firstly, that of dilute NMA in water which offers a model system relevant to the hydration of the peptide group and gives insight into the persistence and structure of hydrogen-bonded oligomers in aqueous solution. At the opposite extreme, concentrated NMA represents a simple system where water exists in an environment rich in peptide groups. Though largely unexplored this concentration regime may be relevant to the issue of interior or buried water molecules trapped in the internal cavities of many proteins and hydrogen-bonded to the main chain backbone[133]. Firstly, the disruption of conventional and weak methyl-donating hydrogen bonding between NMA peptide groups by dilution in water is determined via quantitative analysis of hydrogen-bonded cluster populations, as well as consideration of radial and generalised 2D distribution functions. Secondly, inter-species interactions are explored as a function of concentration and thirdly, the structure and dynamics of water in these mixtures and the progressive changes which result from increasing the peptide group concentration are investigated.

6.1 Comparison of TIP3P, SPC/E and F3C results

First of all it is appropriate to make a brief digression to examine how the quantities in question depend upon the water model used for the aqueous simulations. Figure 6.1 shows a representative sample of intra- and inter-species radial distribution functions for a aqueous solution of concentration $X_{NMA} = 0.2$ (0.2 mole fraction NMA). In this Figure the plotted data are coloured according to the water potential, F3C, SPC/E or TIP3P, used for the simulation of the aqueous NMA mixture. At this concentration water molecules form percolating networks as they do in the bulk and Figure 6.1(a) is clearly very similar to $g(r_{O_W O_W})$ for pure water shown in Figure 4.2. As discussed in Chapter 4 it is apparent that there are differences between TIP3P and the other two

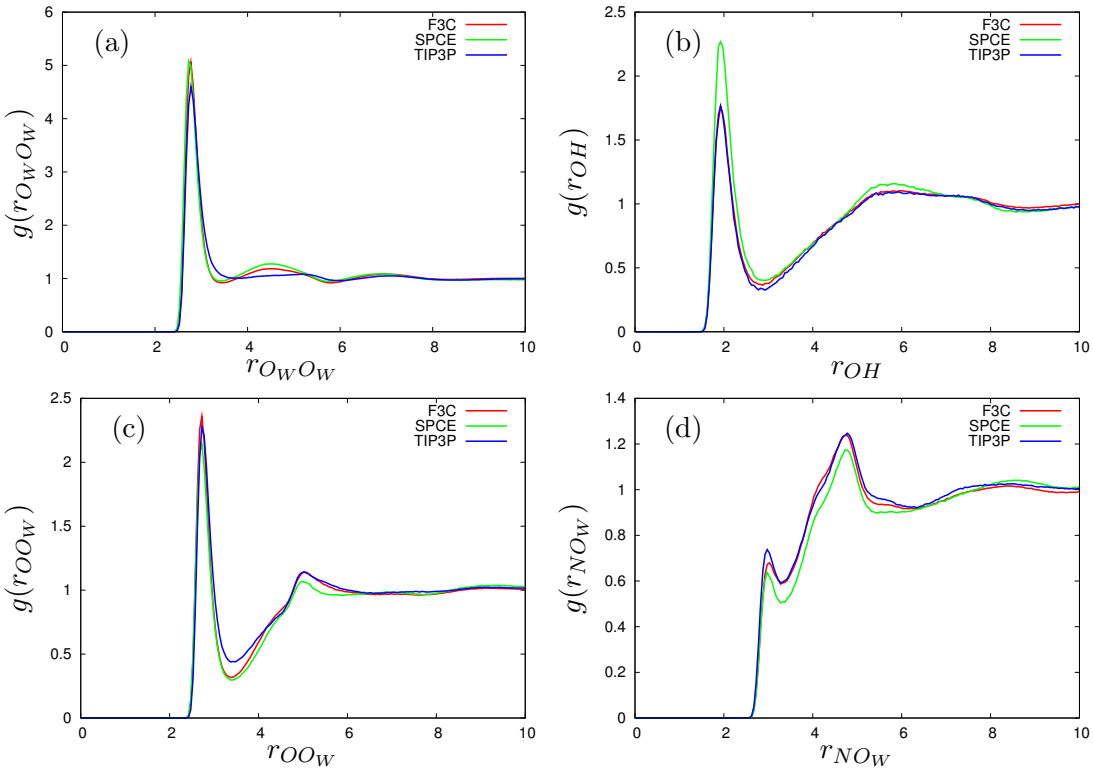

Figure 6.1: Radial distribution functions for $X_{NMA} = 0.2$ aqueous solutions using the CHARMM forcefield for NMA combined with the F3C (red), SPC/E (green) or TIP3P (blue) water models (a) water - water interaction, (b) NMA - NMA interaction, (c) and (d) show water - NMA interactions.

models which more closely match experimental data. $g(r_{OH})$ for interactions between the NMA oxygen and hydrogen is shown in Figure 6.1(b). The radial distribution functions for simulations using the TIP3P and F3C models are very similar while both peaks are overstructured when the SPC/E model is used. The CHARMM forcefield is parametrised to the TIP3P model and therefore this combination should, in a sense, give the "correct" result. The developers of the CHARMM model warn against the use of the CHARMM forcefield with other models, especially SPC/E water as mentioned in Section 3.2.2. The inter-species radial distributions in Figures 6.1(c) and (d) again show close agreement between the results for F3C and TIP3P water models.

The CHARMM forcefield has been parametrised to the TIP3P model and this is, therefore, the logical choice for aqueous simulations. However, the TIP3P model is notoriously bad at reproducing water's tetrahedral extended structure. SPC/E water gives much closer agreement with experimental data on liquid water but for aqueous simulations with the CHARMM forcefield there is likely to be an imbalance in the relative strengths of the water-water and water-solute interactions[76]. The F3C model

offers a compromise as it not only gives a much better representation of liquid water structure than TIP3P but also appears to have the same effect on CHARMM NMA interactions as the correctly parametrised TIP3P. Therefore the remainder of the results presented in this chapter derive from simulations of CHARMM NMA and F3C water.

6.2 Hydration of NMA chains

In Chapter 5 it was shown that NMA molecules hydrogen bond through the amide and carbonyl groups to form extended linear and branched chain structures. At 308.15K molecules are most likely to be found in dimers or trimers but a wide range of cluster sizes appear in the simulated liquid, up to a maximal size of about 100 molecules. Half of the molecules in the system are contained in clusters of 10 or fewer molecules. The aim of this section is to discover how the addition of water affects this extended intermolecular structure of liquid NMA.

6.2.1 Hydrogen bonding between NMA molecules

The radial distribution function for the hydrogen bonding interaction between the NMA oxygen and the NMA amide hydrogen, $g(r_{OH})$, is shown in Figure 6.2. The oxygen - hydrogen separation corresponding to the first peak increases with dilution implying that the inter-NMA hydrogen bond length increases slightly as water is added to the system. This is corroborated by a similar increase in the oxygen - nitrogen (acceptor - donor) first peak distance. Peak positions and coordination numbers are listed in Tables 6.1 and 6.2. Over the concentration range investigated the coordination number of

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
0.1	OH	1.925	2.725	0.15
0.2	OH	1.925	2.875	0.26
0.3	OH	1.925	2.875	0.33
0.4	OH	1.925	2.875	0.43
0.5	OH	1.875	2.925	0.52
0.6	OH	1.875	2.875	0.58
0.7	OH	1.875	2.925	0.67
0.8	OH	1.875	3.025	0.79
0.9	OH	1.875	3.075	0.89
1.0	OH	1.875	3.075	0.99

Table 6.1: Extrema positions and coordination numbers for the 1^{st} peak in $g(r_{OH})$: hydrogen bonding between NMA molecules.

the hydrogen bonded peak falls from 1.0 in pure NMA to 0.15 for $X_{NMA} = 0.1$ showing that there is marginally more hydrogen bonded association between NMA molecules

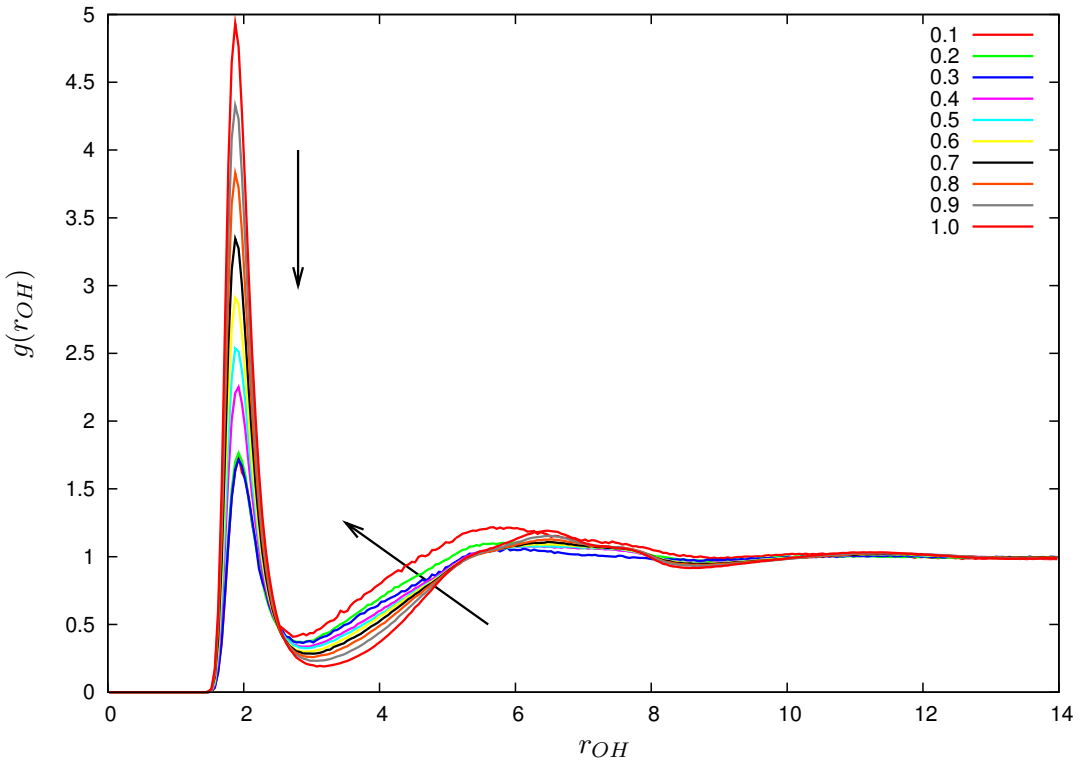

Figure 6.2: $g(r_{OH})$ radial distribution for interactions between the amide hydrogen and carbonyl oxygen of NMA molecules. Arrows point in the direction of increasing X_{H_2O}. The legend indicates the concentration of the mixture over the range 0.1 mole fraction NMA to pure NMA (1.0).

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
0.1	NO	2.925	3.375	0.16
0.2	NO	2.925	3.375	0.25
0.3	NO	2.925	3.475	0.34
0.4	NO	2.875	3.525	0.47
0.5	NO	2.875	3.625	0.59
0.6	NO	2.875	3.525	0.61
0.7	NO	2.875	3.575	0.71
0.8	NO	2.875	3.625	0.82
0.9	NO	2.875	3.625	0.92
1.0	NO	2.875	3.675	1.04

Table 6.2: Extrema positions and coordination numbers for the 1^{st} peak in $g(r_{NO})$.

in the dilute solutions than expected if hydrogen bonding between NMA molecules decreased linearly in proportion to the concentration. The black arrows in Figure 6.2 indicate the direction of movement of the peak positions of the radial distribution function $g(r_{OH})$ as the mole fraction of water is increased. The height of the first peak

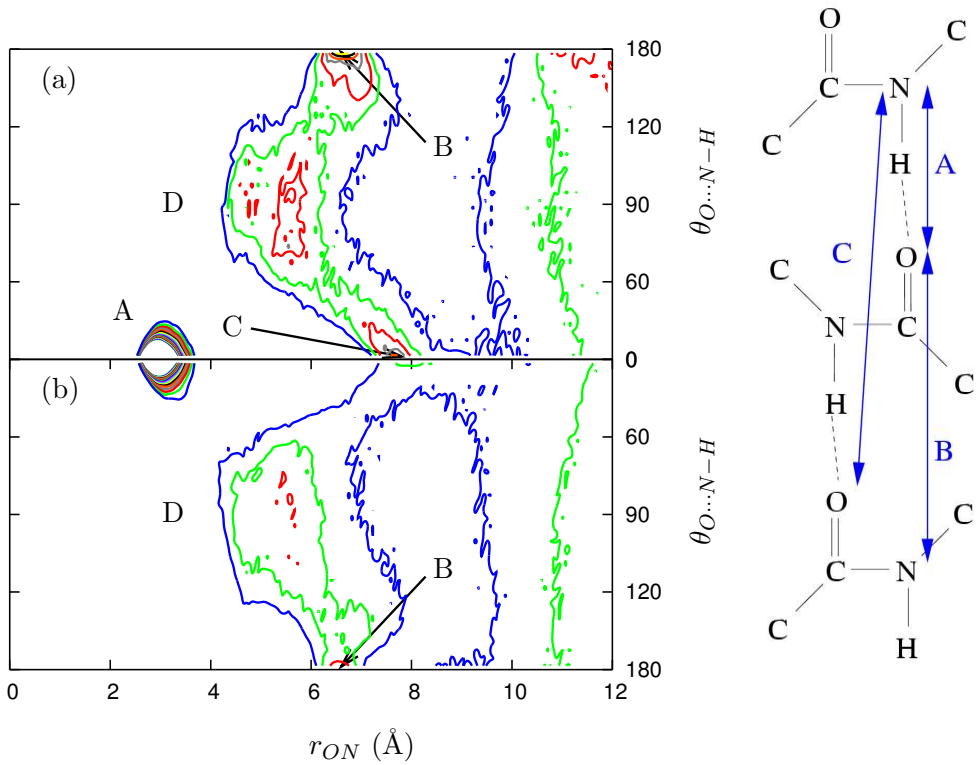

Figure 6.3: $g_2(r_{O\cdots N}, \theta_{O\cdots N-H})$ for (a) $X_{NMA} = 0.4$ and (b) $X_{NMA} = 0.9$ aqueous solutions. Contour levels start at $g_2(r, \theta) = 1$ (blue), increasing in increments of 0.2 up to 5. The labelled features are referred to in the text of Section 6.2.1.

decreases dramatically as the water mole fraction increases. The height represents the number density at this distance relative to that at large r therefore the decrease simply indicates that the number density at the peak location is closer to that in the bulk as the proportion of NMA in the mixture falls. The second peak of $g(r_{OH})$ shifts to lower r as the water mole fraction increases reflecting the smaller size of the water molecule relative to NMA as this peak includes second neighbour molecules both in the intra and inter-chain directions: NMA molecules separated by a water molecule will be closer together than NMA molecules separated by another NMA.

The two dimensional radial-angular distribution function $g_2(r_{ON}, \theta_{ONH})$ is obtained by considering the donor - acceptor distance, $r_{O\cdots N}$ and the (inverse) hydrogen bond angle $\theta_{O\cdots N-H}$. It is illustrated in Figure 6.3 for representative weak (a) $X_{NMA} = 0.9$ and concentrated (b) $X_{NMA} = 0.4$ aqueous NMA solutions. The schematic (methyl hydrogens not shown) illustrates the structural motifs giving rise to the features of $g_2(r_{ON}, \theta_{ONH})$. It is straightforward to assign the prominent peak labelled A at $r_{ON} = 2.9$Å, $\theta_{ONH} = 0°$ to the hydrogen bonded contact between adjacent chain members. The feature at $r_{ON} = 6.5$Å with the opposite, 180°, angle labelled B,

is the corresponding contribution from the amide and carbonyl groups which are not participating in the hydrogen bond. These two associations can be termed the "near" (A) and "far" (B) dimer contacts. The feature labelled C has the same angular dependence as A but is further out with $r_{ON} = 7.6$Å. In fact, A and C are separated by a distance corresponding to one "unit" of the NMA hydrogen bonded chain as previously identified in Section 5.3 and therefore C can be identified as the "near" trimer contact. This feature is also apparent in Figure 6.3(b) which shows $g_2(r_{ON}, \theta_{ONH})$ for a more dilute concentration $X_{NMA} = 0.4$ from which it can be concluded that NMA trimers continue to be present even in relatively dilute solutions.

The broad feature labelled D between 4 and 6Å centred on $\theta_{ONH} = 90°$ is perpendicular to the chain direction and thus arises from inter-chain contributions. A comparison of Figures 6.3 (a) and (b) reveals that the general size and shape of this feature does not change significantly on moving from a concentrated to a dilute NMA solution. The minimum separation between non-hydrogen bonded NMA molecules does not appear to increase and thus one plausible model of NMA hydration in which water molecules infiltrate the spaces between parallel NMA chains can be rejected as according to this hypothesis the inter-chain separation ought to increase.

The broad band between 10 and 12Å arises from a number of inter- and intra-chain configurations including the far trimeric contact, tetramer contacts and neighbouring chain contributions.

6.2.2 Hydrogen bonded clusters

The same geometric criteria are used to define hydrogen bonds between NMA molecules in solution as were defined for liquid NMA in Chapter 5 Section 5.4.4. The cutoff

X_{NMA}	$n = 1$	$n = 2$	$n = 3$	$n = 4$
0.1	74.7	18.6	4.7	1.7
0.2	65.7	24.3	6.8	2.2
0.3	59.0	25.0	9.5	3.8
0.4	45.0	27.3	13.8	7.5
0.5	39.2	25.0	14.8	8.9
0.6	29.6	21.0	15.4	10.9
0.7	23.7	22.0	15.0	11.7
0.8	15.4	15.5	13.8	10.6
0.9	9.3	10.7	10.0	8.9
1.0	4.7	5.9	5.9	5.7

Table 6.3: The percentage of NMA molecules contained in clusters of size $n = 1, 2, 3$ or 4

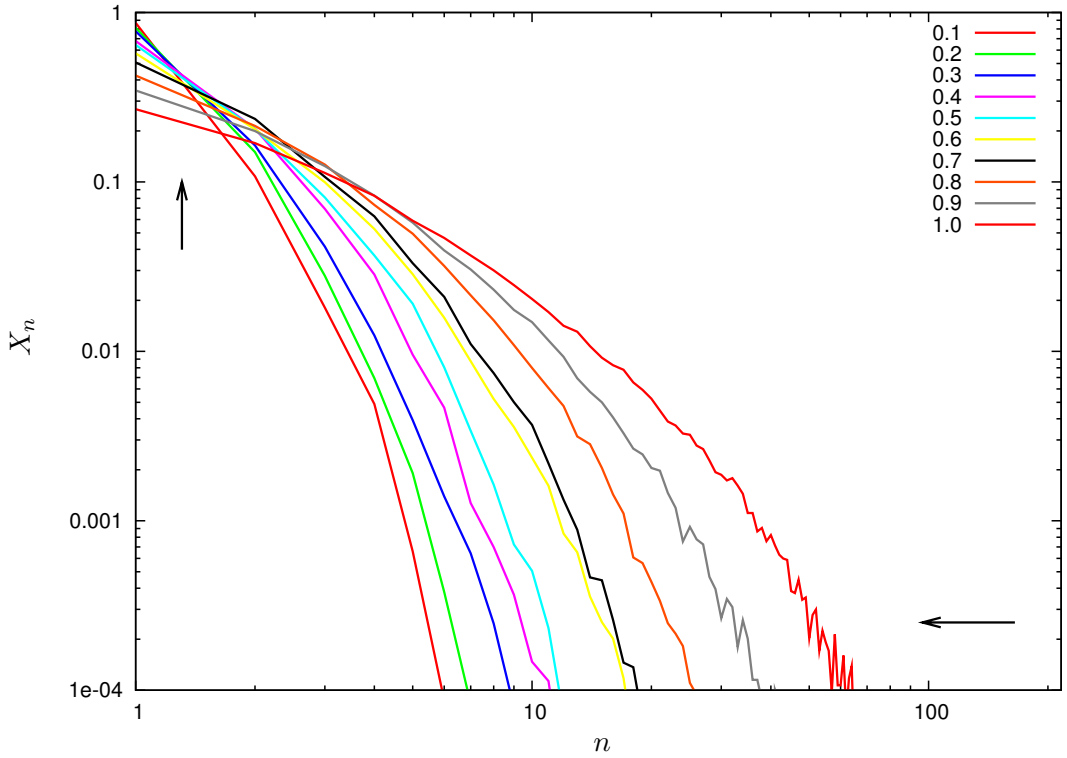

Figure 6.4: Hydrogen bonded NMA cluster population distributions for pure NMA ($X_{NMA} = 1.0$)and a range of aqueous NMA solutions between $X_{NMA} = 0.1$ and $X_{NMA} = 0.9$. The cluster mole fraction $X_n = \frac{N_n}{N_{tot}}$ is plotted as a function of the cluster size n where N_n is the number of clusters of size n molecules and N_{tot} is the total number of clusters identified in the trajectory. Arrows indicate the direction of changes associated with increasing hydration.

distance is the location of the first minimum in the relevant radial distribution function and only $\theta_{O\cdots H-N}$ angles between 140° and 180° are permitted. By this means clusters of NMA molecules connected by continuous networks of NMA-NMA hydrogen bonds can be identified in the aqueous solutions. The distribution of cluster sizes of NMA molecules in the pure liquid and across the full aqueous concentration range is illustrated in Figure 6.4. As the mole fraction of water is increased large NMA clusters are depleted in favour of smaller clusters and monomers. In the NMA-rich solutions a higher percentage of NMA molecules belong to dimers and trimers than exist as monomers as can be seen from the values listed in Table 6.3.

By $X_{NMA} = 0.7$ monomers have become the dominant species. In the water-rich region NMA monomers are overwhelmingly dominant in good agreement with the results of others [46, 53] and the maximum cluster size is less than ten. The arrows in Figure 6.4 illustrate the trends seen with increasing hydration: both the increase in NMA monomers and the decrease in the maximum NMA cluster size. Although NMA molecules show very strong aggregration in inert solvents with self-association

occuring even at concentrations of 0.001M[61], water has been found to impede the self-association of NMA. Schweitzer-Stenner *et al* found that concentrations greater than 6M were required for "effective" aggregation of NMA in aqueous solution[134]. In the classical simulations the $X_{NMA} = 0.1$ solution has a molarity of 4.2M while the $X_{NMA} = 0.2$ solution is 6.8M. The molarities of the other concentrations are listed in Appendix C, Table C.3. At $X_{NMA} = 0.1$ three quarters of the NMA molecules are isolated and borrowing Schweitzer-Stenner's terminology this concentration can be said to be effectively un-aggregated. As the NMA mole fraction is increased the cluster population distribution changes smoothly with concentration and there is no single concentration at which the behaviour of NMA molecules undergoes a sudden change from non-associative to self-associative.

6.2.3 $g(r_{OO})$ and interchain separation in aqueous NMA

Further insight into the breakup of NMA chains on hydration is found by examining the correlation functions for NMA oxygen atoms, $g(r_{OO})$ shown in Figure 6.5 and $g_2(r_{OO}, \theta_{OOC})$ shown in Figure 6.6. In the pure liquid, the first peak of $g(r_{OO})$ at 4.9Å, is attributed to the oxygen - oxygen separation of NMA molecules connected by a strong hydrogen bond, with the low-r wing providing evidence for a range of NMA-NMA inter- and intra-chain close contact geometries as described in Section 5.5 and by Whitfield *et al*[54]. A plausible hypothesis of the hydration process of NMA chains might involve water molecules firstly hydrating the chain ends, leaving the pure NMA intermolecular structure reasonably intact at high concentrations. Subsequently, water molecules might be expected to be dispersed between NMA chains, thus pushing the chains apart. Consequently, this would lead to a decrease in the height of the low-r wing of $g(r_{OO})$. However, evidence in Section 6.2.1 did not support this hypothesis, indicating instead that the interchain separation does not increase on dilution. If the interchain separation does not change upon hydration then the low-r wing of $g(r_{OO})$ might be expected to show little difference as water is added. Figure 6.5 shows $g(r_{OO})$ for pure NMA together with a range of solution concentrations, and illustrates, quite unexpectedly, that the low-r wing is enhanced on dilution, even at the lowest concentrations explored. Curiously, $g(r_{OO})$ for $X_{NMA} = 0.3$ does not quite follow the same trend as the other concentrations. As will be shown in Section 6.2.7 an unusually high proportion of NMA molecules converted to the *cis*-isomer form during the simulation. Since the carbonyl oxygen and the amide hydrogen are both on the same side of the molecule in the *cis*-form when this isomer hydrogen bonds the oxygen - oxygen separation is less than 4.9Å and this explains the shift of number density from the 4.9Å to lower r for this concentration.

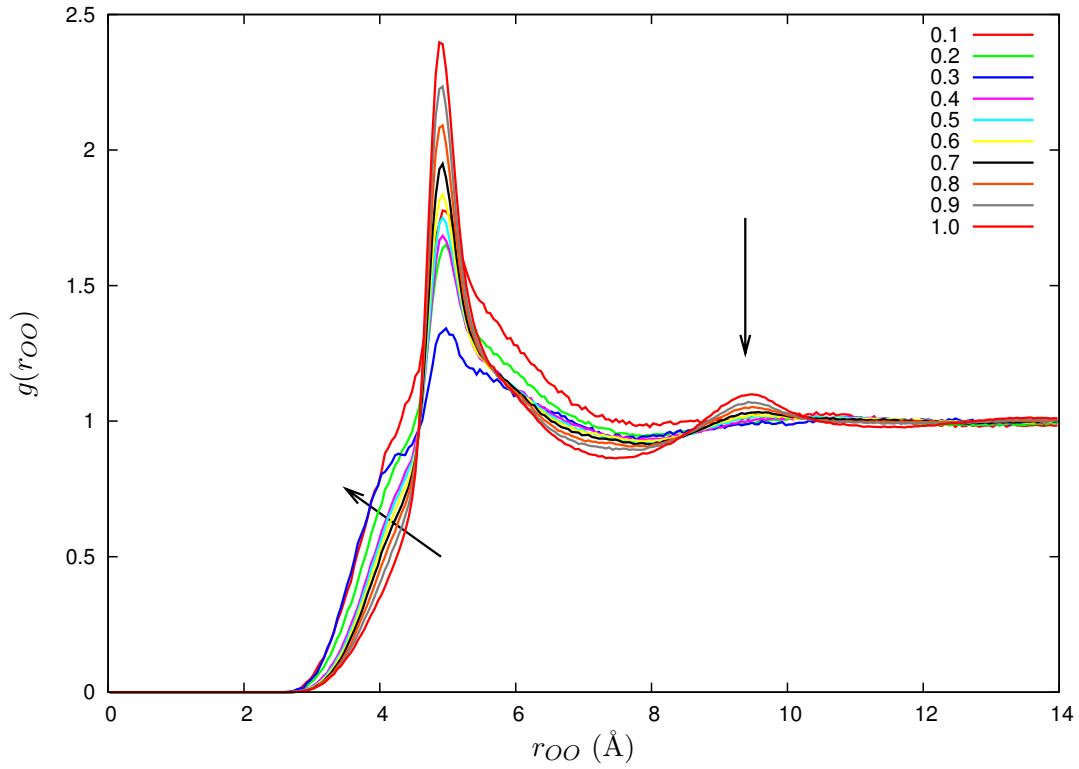

Figure 6.5: $g(r_{OO})$ for pure NMA and its aqueous solutions of concentrations between X_{NMA} =0.1 and 0.9. Bridges formed by hydrogen bonded water molecules force close contact between the oxygen atoms of NMA molecules and cause the low r tail of $g(r_{OO})$ to increase with water mole fraction as indicated by the arrow. Such bridges can be formed by more than one water molecule. A single molecule "water bridge" is illustrated in Figure 6.7. A second arrow indicates the decrease in the second neighbour peak with increasing hydration.

The enhancement of the low-r wing of $g(r_{OO})$ is clear in the joint radial-angular distributions: Figure 6.6 (a) shows that $g_2(r_{OO}, \theta_{OOC})$ for the concentrated $X_{NMA} = 0.9$ mixture has lower relative number density at low-r than its dilute counterpart shown in Figure 6.6 (b). Table 6.4 lists values of n(r) for the position of the first peak in $g(r_{OO})$, revealing, for example, that in a $X_{NMA} = 0.1$ aqueous solution there are on average 0.93 other NMA oxygen atoms within 4.9Å of a randomly chosen NMA oxygen. One might suppose that this number density would decrease linearly with NMA mole fraction thus following the pink line plotted in Figure 6.7. However, as this plot illustrates the actual values are in excess of this prediction and confirm that the apparent enhancement of the low-r wing is not merely an artifact of the normalisation of $g(r_{OO})$. In contrast, the normalised coordination numbers for the hydrogen bonded interactions $g(r_{OH})$ and $g(r_{NO})$ show ideal dependence upon concentration down to $X_{NMA} = 0.6$. At lower concentrations there appears to be slightly enhanced NMA hydrogen bonded self-association but clearly insufficient to explain the significant increase in oxygen - oxygen association. The enhancement

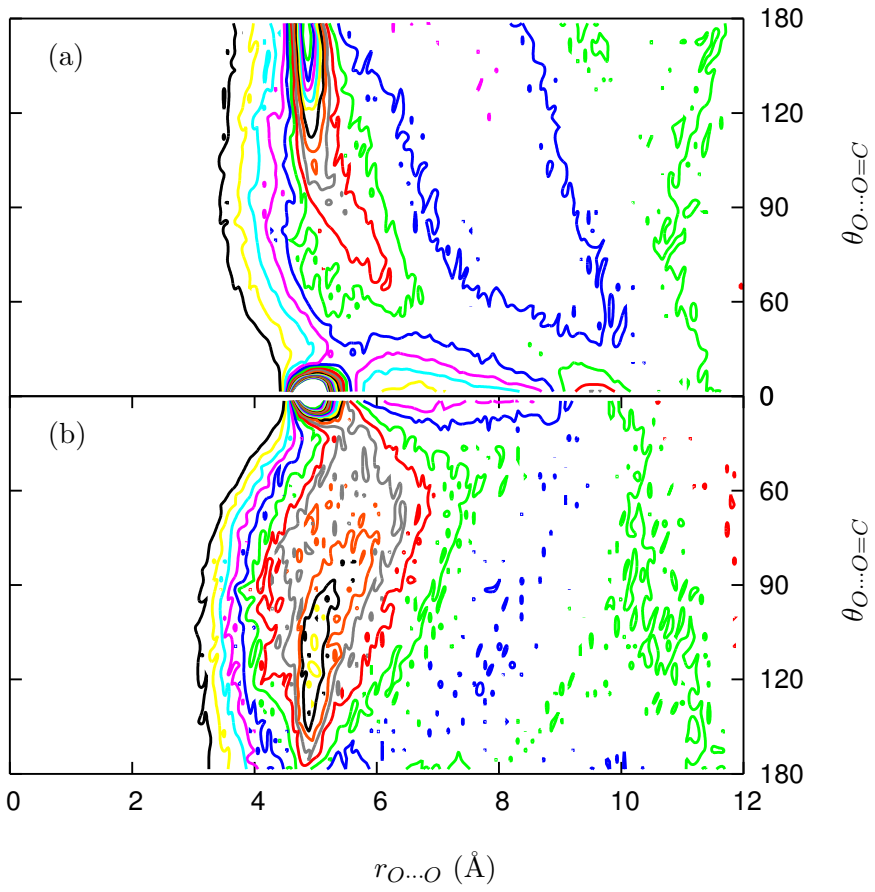

Figure 6.6: $g_2(r_{O\cdots O}, \theta_{O\cdots O=C})$ for (a) $X_{NMA} = 0.9$ and (b) $X_{NMA} = 0.1$ aqueous solutions. Contour levels start at $g_2(r, \theta) = 0.2$ (black), increasing in increments of 0.2 up to 5. The $g_2(r, \theta) = 1$ contour is blue.

of the low-r "inter-chain" correlation at concentrations where most NMA molecules have no hydrogen-bonded neighbours of their own species indicates that water - NMA interactions must be responsible for promoting close NMA - NMA contacts.

6.2.4 Water molecules bridge the gap between NMA molecules

NMA molecules are not only able to form strong hydrogen bonds with their own species but also with water molecules. This can result in configurations where one or more water molecules form a hydrogen-bonded "bridge" between two NMA molecules as illustrated in Figure 6.7. These bridge structures promote close contact between the connected NMA molecules leading to oxygen - oxygen separations as low as 3Å, which contribute to the low-r wing of $g(r_{OO})$. This observation illustrates that the addition of water has an unexpected and non-trivial effect on oxygen - oxygen separation in this model peptide and similar motifs may well also be visible in aqueous proteins. A recent

X_{NMA}	g(r)	1^{st} peak position (Å)	n(r_{peak})	n(r=7.775Å)
0.1	OO	4.925	0.93	5.39
0.2	OO	4.975	1.42	8.07
0.3	OO	4.975	1.72	9.64
0.4	OO	4.925	1.74	11.11
0.5	OO	4.925	1.90	12.09
0.6	OO	4.925	2.02	12.86
0.7	OO	4.925	2.12	13.44
0.8	OO	4.925	2.23	13.88
0.9	OO	4.925	2.30	14.21
1.0	OO	4.925	2.36	14.44

Table 6.4: 1^{st} peak position in $g(r_{OO})$ and cumulative number densities for a range of concentrations. As $g(r_{OO})$ does not have a well defined first minimum the cumulative number density at a constant distance of 7.775Å is listed as this gives an indication of the dependence of oxygen - oxygen coordination number upon concentration even though it cannot strictly be called a coordination number itself.

investigation of the energy transfer and reorientational dynamics of the equivalent motif in aqueous *N,N*-dimethylacetamide (DMA) suggests that water bridges such as these act as an "extension cord" between amide molecules, connecting groups that would otherwise be sterically hindered[135]. The restricted geometries of both the water molecule and the water - amide hydrogen bond would seem to indicate that the amide - amide associations ought to have a "specific structural rigidity" as suggested by Timmer and Bakker. This has implications for the ability of water molecules to act as effective molecular chaperones in protein folding and in molecular recognition. If such structural rigidity were present in the aqueous NMA simulations then it should be apparent in $g_2(r_{OO}, \theta_{OOC})$ shown in Figure 6.6 as a directional peak in the low-r region. No such feature appears in the radial angular distribution indicating that although the formation of a water bridge brings the two carbonyl groups into close proximity their relative orientation remains fairly unrestricted. This property may be useful to the activity of water as a molecular chaperone as it implies that peptides brought into close contact by bridging waters are not locked into a peptide - water - peptide structure but remain able to perform an orientational search for the optimum peptide - peptide configuration. The bridging water motif may be a transition state in biomacromolecules rather than or in addition to an optimised structure.

The second peak of $g(r_{OO})$ at 9.5Å diminishes as water is added and seems to have vanished almost completely at the lowest concentrations. This is easily understood by considering the clustering results illustrated in Figure 6.4. At $X_{NMA} = 0.2$ only 10% of the NMA molecules form clusters of three or more molecules and so are able to

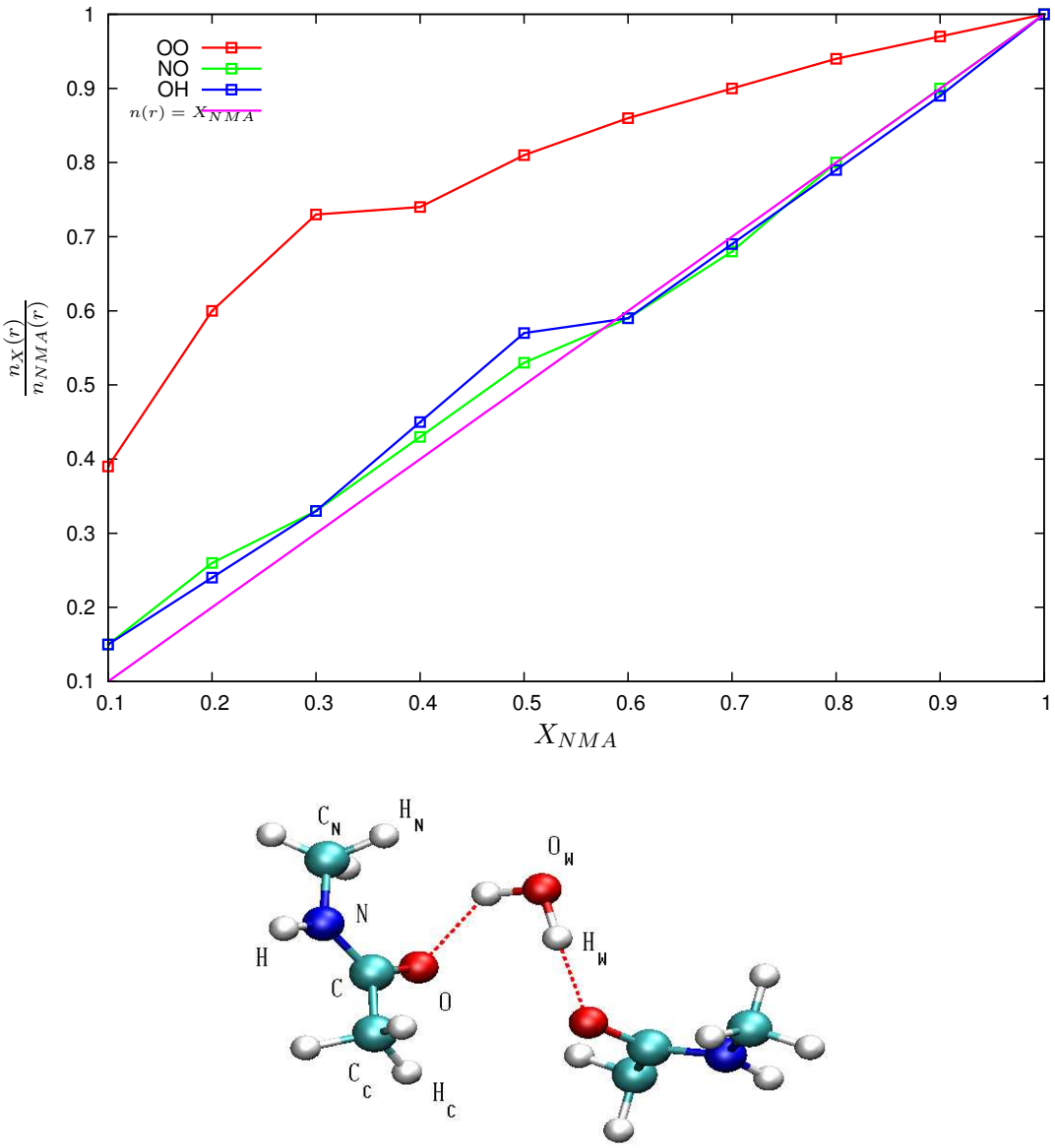

Figure 6.7: This figure plots the coordination numbers of the first peaks in $g(r_{OH})$ (blue) and $g(r_{NO})$ (green) listed in Tables 6.1 and 6.2 normalised to the value at $X_{NMA} = 1.0$. The number densities at the first peak of $g(r_{OO})$ (red) listed in Table 6.4 normalised to the value at $X_{NMA} = 1.0$ are also plotted and compared to the ideal situation where number density is a linear function of the NMA mole fraction (pink). The increased oxygen - oxygen coordination is due to structural motifs such as the "water bridge" illustrated.

contribute to this peak.

6.2.5 NMA cluster topology as a function of concentration

NMA oxygen atoms are capable of accepting two hydrogen bonds and do so in just less than 10% of cases in the pure liquid. Consequently, these doubly accepting molecules form branchpoints in hydrogen-bonded chains. It is possible to obtain a rough

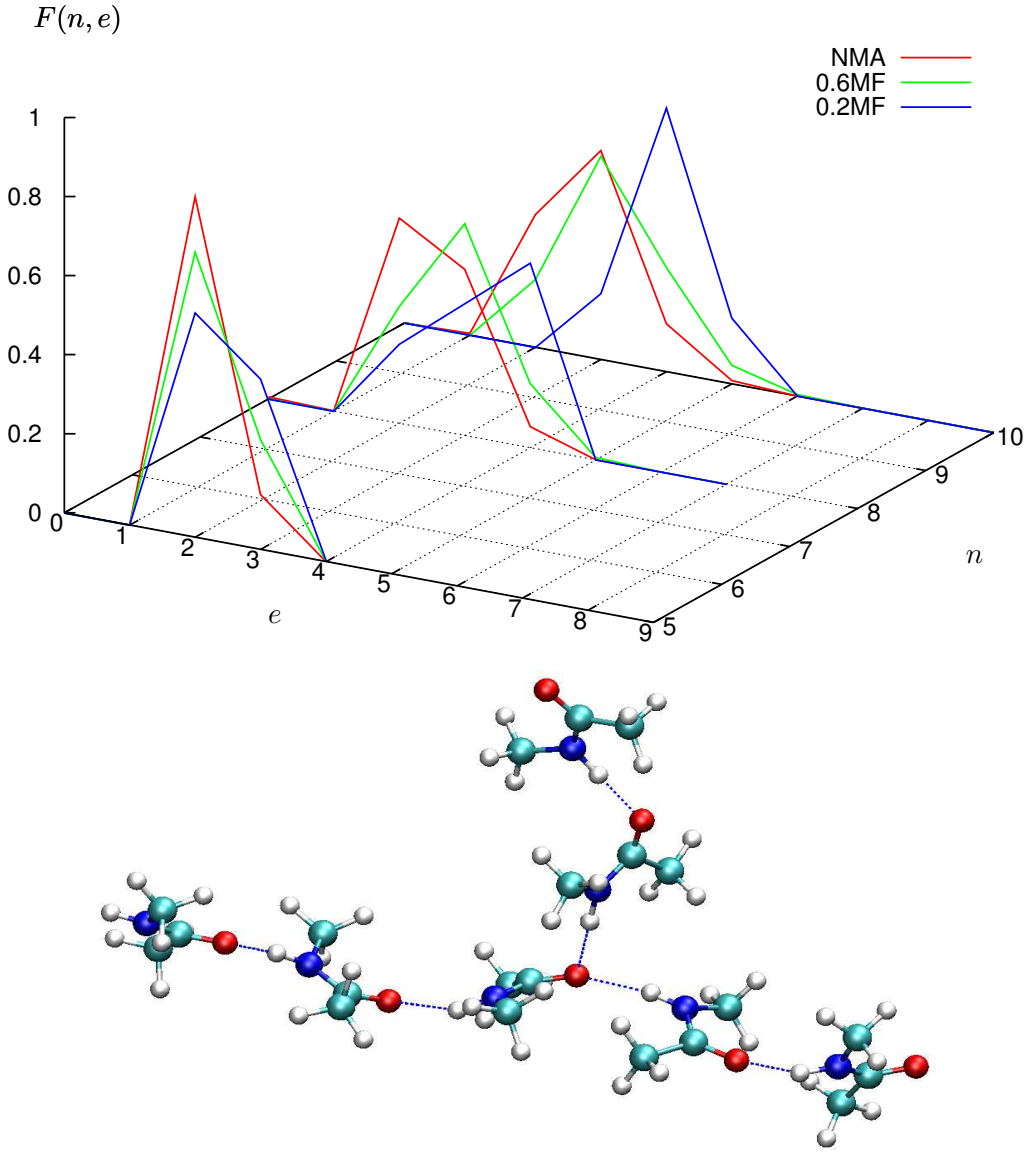

Figure 6.8: $F(n,e)$, the proportion of the total number of clusters containing n molecules which have e endpoints is plotted as a function of n and e for a range of concentrations. The simulation snapshot shows a seven molecule cluster with 3 endpoints (1 branch).

estimate of the degree of branching of the NMA hydrogen bonded networks by counting the number of molecules in a given cluster which only have one hydrogen-bonded neighbour. Such molecules are "endpoints" of the cluster. Preliminary investigations

confirmed that a population of *trans*-NMA molecules are unlikely to hydrogen bond to form rings and therefore the number of branches in a given cluster is equal to the number of endpoints - 2. Figure 6.8 presents the variation in number of endpoints as a function of concentration for three different cluster sizes and an illustrative seven molecule branched cluster. Comparing NMA clusters of like size reveals a trend whereby the average number of branches increases with decreasing NMA concentration. This trend suggests the following theory of hydration of NMA chains: as water is added to the system it is more likely to be inserted into the NMA chains at linear sections. Probably this is simply be due to steric effects. It will be easier for water molecules to approach linear segments of NMA chain as opposed to branchpoints. However, a similar trend was observed when the simulation temperature of liquid NMA was increased suggesting that there may also be an energetic preference for breakage to occur at linear chain sections. The trend towards increasingly branched clusters explains the slightly enhanced hydrogen bonded NMA self-association observed for $X_{NMA} < 0.6$ in Figure 6.7.

6.2.6 Methyl group interactions

In Chapter 5 the interactions of the two NMA methyl groups with the carbonyl oxygen atom were discussed and it was concluded that there are associations present in liquid NMA between the oxygen and the methyl hydrogens on the carbonyl side of the molecule which can be described as weak or improper hydrogen bonds. These arise from both inter-chain contacts and from the specific hydrogen bond geometry within $N - H \cdots O = C$ hydrogen bonded chains. In this section the effect of hydration upon these weak associations between NMA molecules is explored. It has already been observed that the NMA inter-chain separation does not increase on solvation so it might be anticipated that there will be little change in methyl - oxygen interactions. On the other hand, a new mechanism for close NMA oxygen - oxygen contacts has been identified in the aqueous solutions and these water bridges may turn out to have a non-trivial effect on methyl - oxygen interactions. Figure 6.9 shows the radial angular distributions $g_2(r_{OH_C}, \theta_{OH_CC_C})$ and $g_2(r_{OH_N}, \theta_{OH_NC_N})$ for the interactions of the carbonyl methyl hydrogens and the amide methyl group hydrogens respectively with the NMA oxygen in $X_{NMA} = 0.1$ and $X_{NMA} = 0.4$ aqueous solutions. On comparing these distributions with those of the pure liquid in Figure 5.16 it can be seen that the general features remain the same and there do not appear to be any new features which can be definitely associated with the formation of water bridges between NMA molecules. On hydration the coordination numbers and spatial distributions of the methyl hydrogens change in a manner consistent with the break-up of NMA chains. Weak interactions between the methyl hydrogens and the carbonyl oxygen continue to

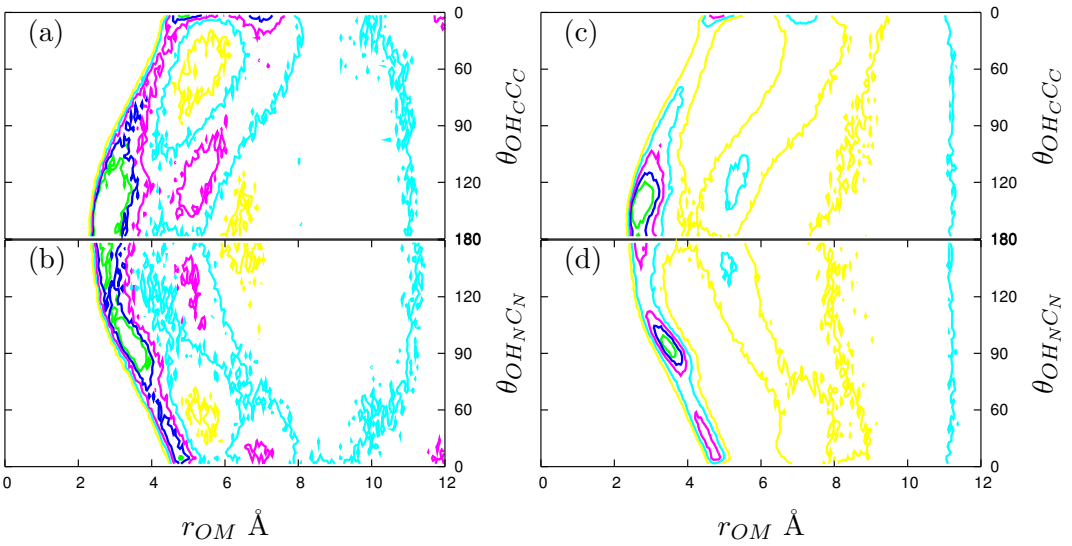

Figure 6.9: Contour plots of the radial angular distributions for $g_2(r_{O\cdots H_C}, \theta_{O\cdots H_C - C_C})$ and $g_2(r_{O\cdots H_N}, \theta_{O\cdots H_N - C_N})$ in $X_{NMA} = 0.1$ (a) and (b) and $X_{NMA} = 0.4$ (c) and (d) aqueous solutions. In the x-axis label M refers to either H_N or H_C. Contour levels begin at 1 (yellow) and increase in increments of 0.2 to 2.8. These plots can be compared to Figure 5.16 which shows the same correlation function for liquid NMA.

be apparent wherever there are strongly hydrogen bonded NMA chains or dimers.

6.2.7 O=C-N-H dihedral angle

The NMA molecule has two naturally occurring isomers, the most common being the trans-isomer with dihedral angle $\phi_{OCNH} = 180°$ as this form is more stable by 2.07 kcal mol^{-1}[76]. The *cis*-NMA population has been measured as 1.5% by ultraviolet Raman spectroscopy at room temperature (298K)[136] and as 2.7% at 293K using FT-IR[57]. The potential energy barrier associated with transformations between the *cis* and *trans* isomers is implemented in the CHARMM model as described in Chapter 3, Section 3.2.4. The initial configuration for all the simulations comprised a set of *trans*-NMA molecules, however from Figure 6.10 it is apparent that during four of the 1ns simulations some NMA molecules made a transition to their *cis*-isomeric form. The majority of molecules however remain in their initial *trans*-isomeric state. In the pure NMA simulation the *cis*-population is 0.93% on average. The other simulations in which *cis*-isomers appear are $X_{NMA} = 0.7$, 0.5 and 0.3. In the $X_{NMA} = 0.7$ and 0.5 simulations the $cis-$populations are 0.52% and 1.14% respectively. During the $X_{NMA} = 0.3$ simulation a particularly high proportion of the initally *trans*-NMA molecules isomerised to the *cis* form to give an average *cis* population of 30.7%. The impact of this high percentage of *cis*-isomers can be seen, for example, in the radial distribution function $g(r_{OO})$ shown in Figure 6.5. The high degree of isomerisation at

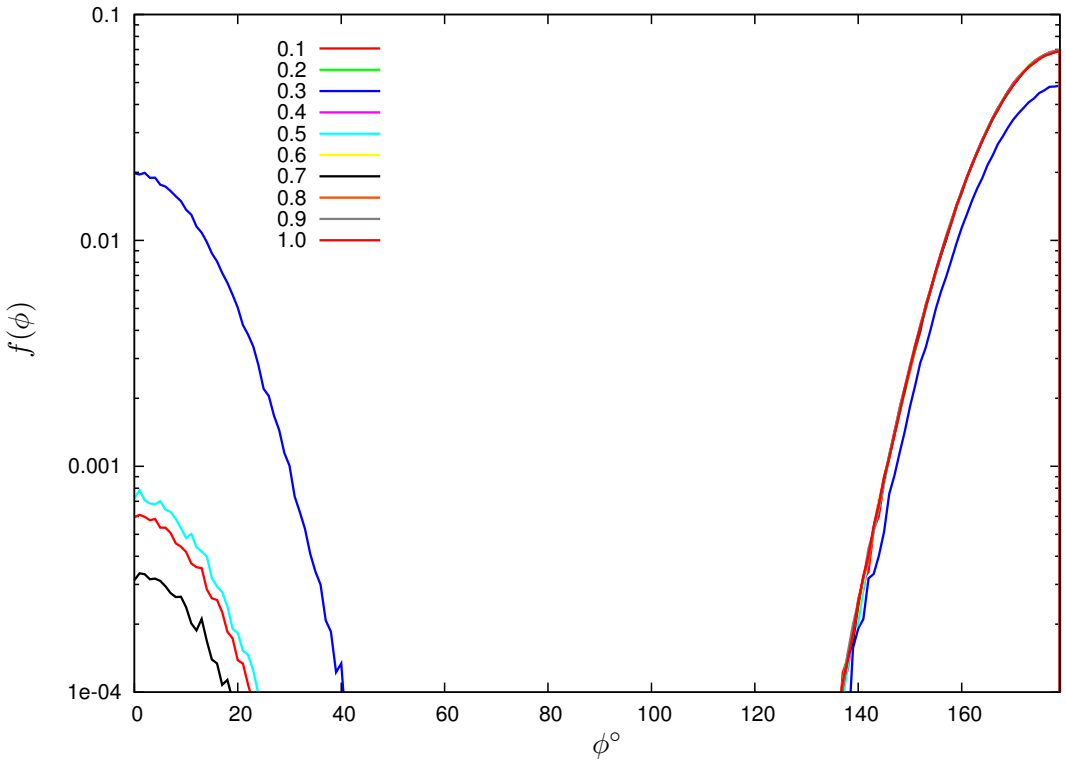

Figure 6.10: Frequency distribution of the NMA O=C-N-H dihedral angle, ϕ_{OCNH}, as a function of concentration, X_{NMA}

this concentration does not appear to be due to the combination of the CHARMM22 parameters with the F3C water forcefield as an independent simulation using the CHARMM22 TIP3P forcefield resulted in an average *cis*-population of 37.9%. The CHARMM model is flexible and this leads to a range of possible ϕ_{OCNH} angles. There is no change in the frequency distribution of this angle upon hydration.

6.2.8 Discussion of NMA hydration

In its liquid phase NMA forms hydrogen bonded linear and branched chains. Upon hydration these chains break up at linear sections into shorter, more highly branched, chain segments with an increasing dominance of trimers, dimers and isolated molecules. Judging by the trend in coordination numbers there is slightly more hydrogen bonded association between NMA molecules at low concentrations than one might expect which can be attributed at least in part to the proportional increase in doubly accepting oxygens. A quarter of NMA molecules are hydrogen bonded to other NMA molecules at $X_{NMA} = 0.1$ even though there are nine water molecules for each NMA at this concentration. As well as breaking up hydrogen bonded chains the addition of water appears to cause a slight lengthening of the hydrogen bond formed between the amide

and carbonyl functional groups. Despite the disintegration of the NMA hydrogen bonded extended structure NMA molecules show significant non-hydrogen bonded self-association even at the lowest concentrations explored due to the formation of water "bridges" between NMA molecules. These structural motifs cause the enhancement of the low-r wing of $g(r_{OO})$ which in pure NMA is due to inter-chain associations. Methyl group interactions with the NMA oxygen which in the pure liquid were found to appear in both inter- and intra-chain structures appear to decrease in a manner consistent with the break up of NMA chains on hydration.

6.3 Interspecies hydrogen bonding

As discussed in the previous section the addition of water to liquid NMA has an unexpected and non-trivial effect on NMA - NMA interactions due to the formation of hydrogen bonded water bridge structures. In this section the cross-species interactions are considered in more detail. An obvious feature of the radial distribution functions presented in this section is that the heights of the $g(r)$ peaks change markedly as a function of concentration. In general this is simply because the distribution is normalised to the bulk density of either water or NMA depending on the correlation under consideration. Therefore, as the proportion of water/NMA in the system is changed the difference between local and bulk densities becomes more or less pronounced[137].

6.3.1 Correlations between the NMA and water oxygen atoms

The radial distribution function for contacts between the NMA and water oxygen atoms is shown in Figure 6.11 and provides insight into inter-species hydrogen-bonding within the liquid. The first peak at 2.7Å in $g(r_{OO_W})$ corresponds to hydrogen bonded configurations where the NMA molecule acts as an acceptor of one or more water hydrogens. This is illustrated in the simulation snapshot pictured in Figure 6.12, with this type of $O \cdots O_W$ interaction labelled as A. The second distinct peak at 4.9Å arises from configurations in which the NMA molecule acts as a hydrogen bond donor (the $O \cdots O_W$ interaction labelled B in Figure 6.12). This peak broadens to lower r and becomes less distinct as the the NMA concentration decreases in the aqueous solutions. It is apparent that a variety of different NMA - water structural motifs contribute to the 4.9Å peak of $g(r_{OO_W})$:

- water molecules separated by two or three hydrogen bonds from the NMA oxygen (C in Figure 6.12)
- water molecules hydrogen-bonded to the NMA donor group (B in Figure 6.12)

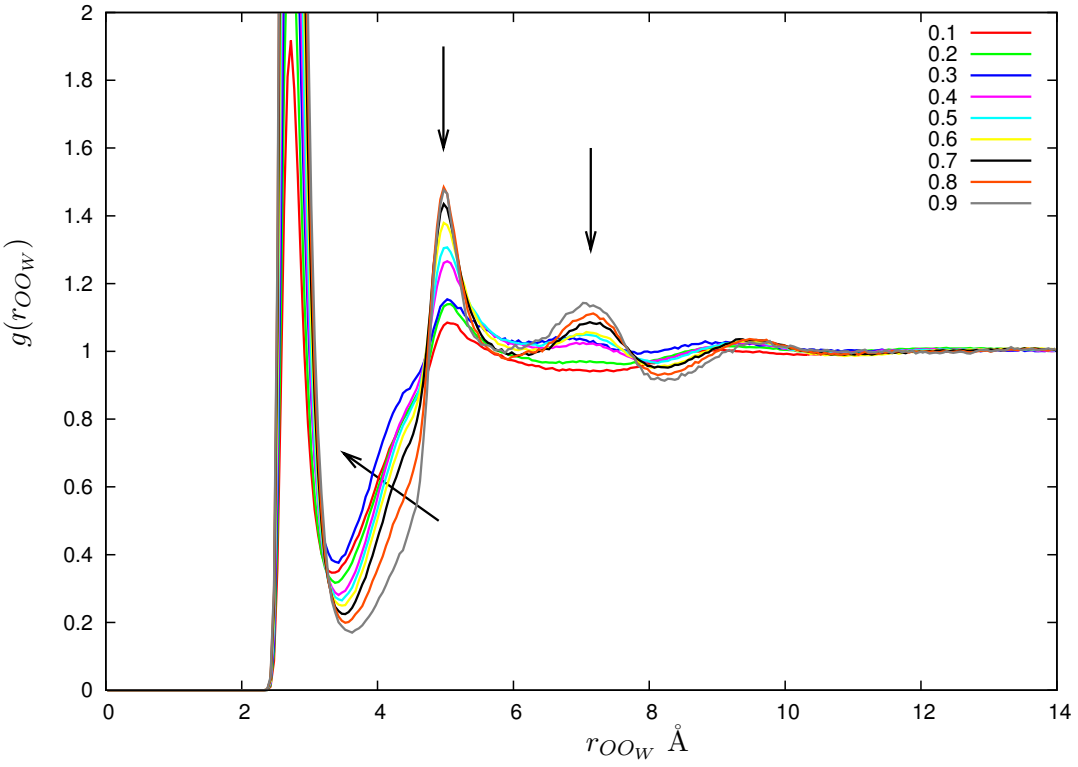

Figure 6.11: $g(r_{OO_W})$ for aqueous NMA in the concentration range $X_{NMA} = 0.1$ to $X_{NMA} = 0.9$. Arrows indicate trends as the mole fraction of water is increased.

- water molecules close to but not necessarily hydrogen-bonded to the NMA donor group (not shown)

Of these, the principal set of molecules contributing to the region between the two peaks are water molecules connected to the NMA molecule by two hydrogen bonds, labelled C in Figure 6.12. This structural motif is increasingly prevalent as water is added to the system and consequently $g(r_{OO_W})$ in this region increases as indicated by the arrow in inverse proportion to the NMA mole fraction.

$g(r_{OO_W})$ for the simulation at $X_{NMA} = 0.3$ has a slightly different r dependence to those of the other concentrations. Its r dependence is similar to that of $g(r_{OO})$ for this concentration shown in Figure 6.5. As before the shift away from the 4.9Å peak towards lower r can be attributed to the unusually high proportion of *cis*-isomers present at this concentration as identified in Section 6.2.7.

In the NMA-rich region a new peak appears at about 7Å. This feature is separated from the first peak by a distance corresponding to a single unit in the NMA chain and therefore can be attributed to the correlation between a water molecule hydrogen bonded to the end of an NMA chain and the second molecule in that chain. The NMA-

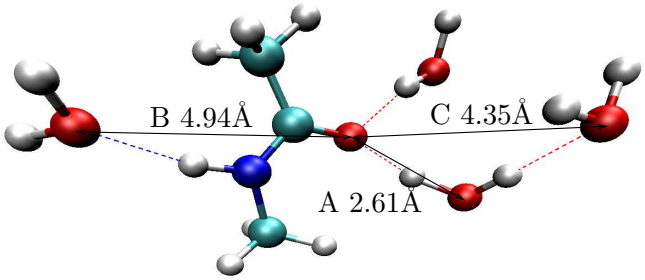

Figure 6.12: A snapshot taken from a simulation of $X_{NMA} = 0.1$ showing a single NMA molecule with three hydrogen-bonded water neighbours and a second nearest neighbour water connected to the NMA by two hydrogen bonds. Hydrogen bonds donated by the NMA molecule are indicated with a blue dotted line. Those donated by waters with a red dotted line. The black arrows and letters refer to the interactions described in the text.

water structural motifs described above explain the form of the other inter-species distributions not explicitly shown such as $g(r_{NO_W})$. In the following sections the two different types of peptide - water hydrogen bonds will be examined in more detail, firstly considering the interaction where the NMA molecule acts as a donor.

6.3.2 Donation of the NMA amide hydrogen to the water oxygen

Figure 6.13 shows the radial distribution function for interactions between the amide hydrogen and water oxygens. There is a prominent first peak at just under 2 Å which is characteristic of hydrogen bonding. Directionality is the distinguishing feature of hydrogen bonding as opposed to van der Waals interactions and by decomposing the radial distribution into the angular domain as in Figure 6.14 it can be seen that the peak at 2Å labelled A in this figure, is centred on $\theta_{O_W \cdots H-N} = 180°$. Table 6.5 lists the coordination numbers for this peak for the full range of aqueous NMA concentrations.

As the proportion of NMA in the mixture increases the coordination number of this

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
0.1	HO_W	2.025	2.625	0.75
0.2	HO_W	1.975	2.725	0.73
0.3	HO_W	2.025	2.625	0.57
0.4	HO_W	1.975	2.675	0.53
0.5	HO_W	1.975	2.675	0.45
0.6	HO_W	1.975	2.775	0.41
0.7	HO_W	1.925	2.725	0.31
0.8	HO_W	1.925	2.825	0.22
0.9	HO_W	1.925	2.825	0.11

Table 6.5: Extrema positions and coordination number for the 1^{st} peak in $g(r_{HO_W})$: hydrogen bonds accepted by the water oxygen from NMA amide hydrogens.

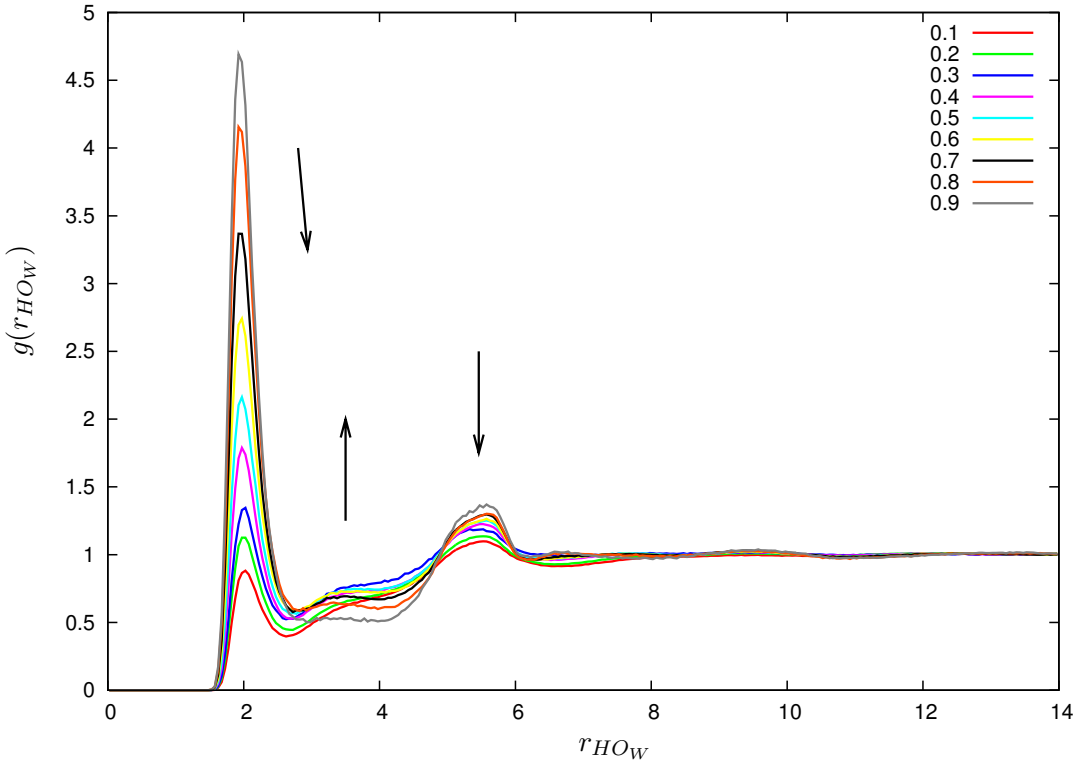

Figure 6.13: $g(r_{HO_W})$ radial distribution for interactions between the amide hydrogen of the NMA molecule and water oxygens. Arrows indicate hydration changes.

peak decreases from 0.75 ($X_{NMA} = 0.1$) to 0.11 ($X_{NMA} = 0.9$). The position of the first peak changes by 0.1Å indicating a decrease in the length of the HO_W hydrogen bond with increasing NMA content. Note that the bin width of the histogram used to collect the data for g(r) is 0.05 and this limits the accuracy with which peak positions can be pinpointed. A change of 0.1Å corresponds to 2 bin widths. The peak and minimum positions and coordination numbers of the donor - acceptor association are listed in Table 6.6 and show the same variation as the hydrogen - acceptor interaction.

There is also a prominent second peak between 5 and 6Å in $g(r_{HO_W})$ corresponding to feature B in the radial angular distribution. It may be recalled that the position of the first peak in $g(r_{OO})$ is at 4.9Å and that this distance therefore characterises the length of a single "link" in the NMA hydrogen bonded chain. Figure 6.14 shows that the second peak is formed from contributions with $\theta_{O_W HN} = 0°$. By connecting these two pieces of information it can be deduced that the molecular configuration contributing to this peak is one in which the water molecule is donating to the NMA oxygen on the opposite side of the NMA molecule to the N-H amide group.

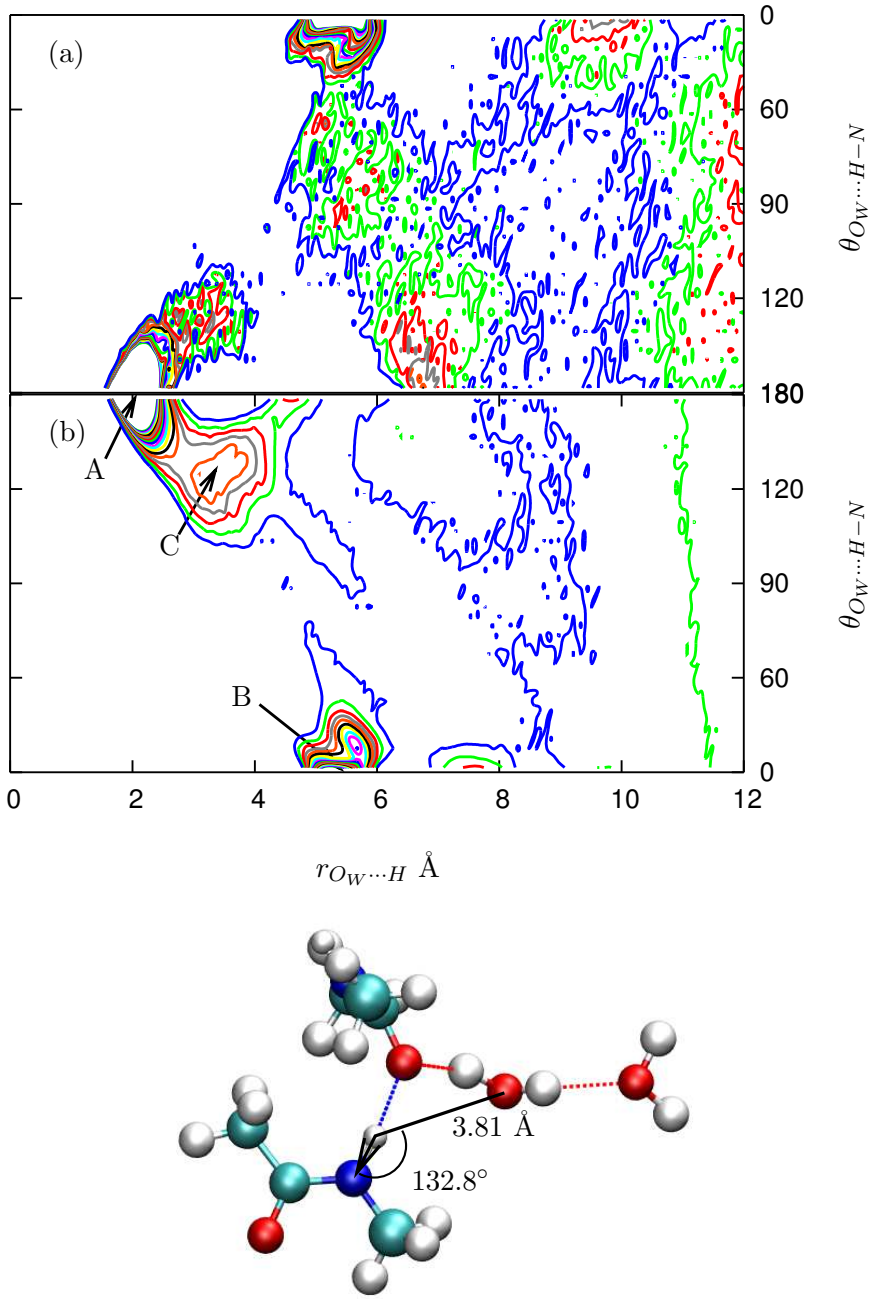

Figure 6.14: Contour plot of the $g_2(r_{O_W\cdots H}, \theta_{O_W\cdots H-N})$ radial-angular distribution for interactions between the amide donor group of the NMA molecule and water oxygens in (a) $X_{NMA} = 0.9$ and (b) $X_{NMA} = 0.4$ solutions. Contour levels start at $g_2(r, \theta) = 1$ (blue) increasing in 0.2 increments. A snapshot from a $X_{NMA} = 0.4$ simulation showing the water oxygen - amide group interaction contributing to feature C

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
0.1	NO_W	3.025	3.225	0.63
0.2	NO_W	3.025	3.275	0.62
0.3	NO_W	2.975	3.275	0.53
0.4	NO_W	2.975	3.425	0.59
0.5	NO_W	2.975	3.425	0.49
0.6	NO_W	2.975	3.475	0.43
0.7	NO_W	2.925	3.475	0.33
0.8	NO_W	2.925	3.525	0.24
0.9	NO_W	2.925	3.525	0.12

Table 6.6: Extrema positions and coordination number for the 1^{st} peak in $g(r_{NO_W})$: hydrogen bonds accepted by the water oxygen from NMA amide hydrogens. The NMA nitrogen is the donor atom.

As the proportion of NMA in the mixture increases there is an interesting effect on the region of $g(r_{HO_W})$ between the two peaks. At $X_{NMA} = 0.1$ this region presents itself as a sloping shoulder leading gradually up to the second peak. As the NMA concentration increases this region becomes more step-like and reduces in height, eventually appearing flat at $X_{NMA} = 0.9$. From the contour plot of the radial angular distribution $g_2(r_{O_WH}, \theta_{O_WHN})$ in Figure 6.14 it can be seen that at mid range concentrations ($X_{NMA} = 0.4$ is shown in Figure 6.14(b)) this region of the radial distribution decomposes in the angular direction into a broad peak labelled C. This peak lies between 3 and 4Å with an angular range centred on 135°. It can be attributed to interactions of the amide group with second neighbour waters as illustrated in Figure 6.14.

Comparing $g_2(r_{O_WH}, \theta_{O_WHN})$ for two different concentrations $X_{NMA} = 0.9$ in Figure 6.14(a) and $X_{NMA} = 0.4$ in Figure 6.14(b) highlights some of the other changes which occur upon hydration. At the higher concentrations the interaction between water molecules and the far neighbour in an NMA dimer can be seen: peaks at $g_2(r_{O_WH} = 6.5\mathring{A}, \theta_{O_WHN} = 180°)$ and $g_2(r_{O_WH} = 9.6\mathring{A}, \theta_{O_WHN} = 0°)$. These peaks are barely apparent in the $X_{NMA} = 0.4$ solution. Conversely, a new peak at $g_2(r_{O_WH} = 7.6\mathring{A}, \theta_{O_WHN} = 0°)$ has appeared at the lower concentration. Interactions between water molecules connected by two hydrogen bonds to the carbonyl group contribute to this peak (the $O_W \cdots H - N$ interaction between the far right water molecule in Figure 6.14 with the amide group of the upper central NMA molecule). This peak has quite a narrow angular range, even allowing for the use of a linear $\cos\theta$ scale and suggests that the hydration shell around the carbonyl group of NMA molecules is well ordered.

6.3.3 Donation of water hydrogens to the NMA oxygen

As in the amide hydrogen - water oxygen radial distribution function, the radial distribution for interactions of water hydrogens with the NMA oxygen shown in Figure 6.15 also displays the characteristic peak of hydrogen bonding at just under 2Å. By decomposing the radial distribution into the angular dimension using the hydrogen bond angle $\theta_{OH_WO_W}$ it becomes apparent that this interaction labelled A in Figure 6.16 is firmly centred on 180° and consequently deserving of the hydrogen bond moniker. The smaller, second peak labelled B at about 3.1Å is simply due to the second hydrogen belonging to the water molecule which forms the hydrogen bond contributing to the first peak. The region between the second and third peaks of the radial distribution is revealed in Figure 6.16 to be a concentration dependent peak at 4Å labelled C, with an angular range of $\theta_{OH_WO_W} > 120°$ centred on 180°. This peak is due to second neighbour

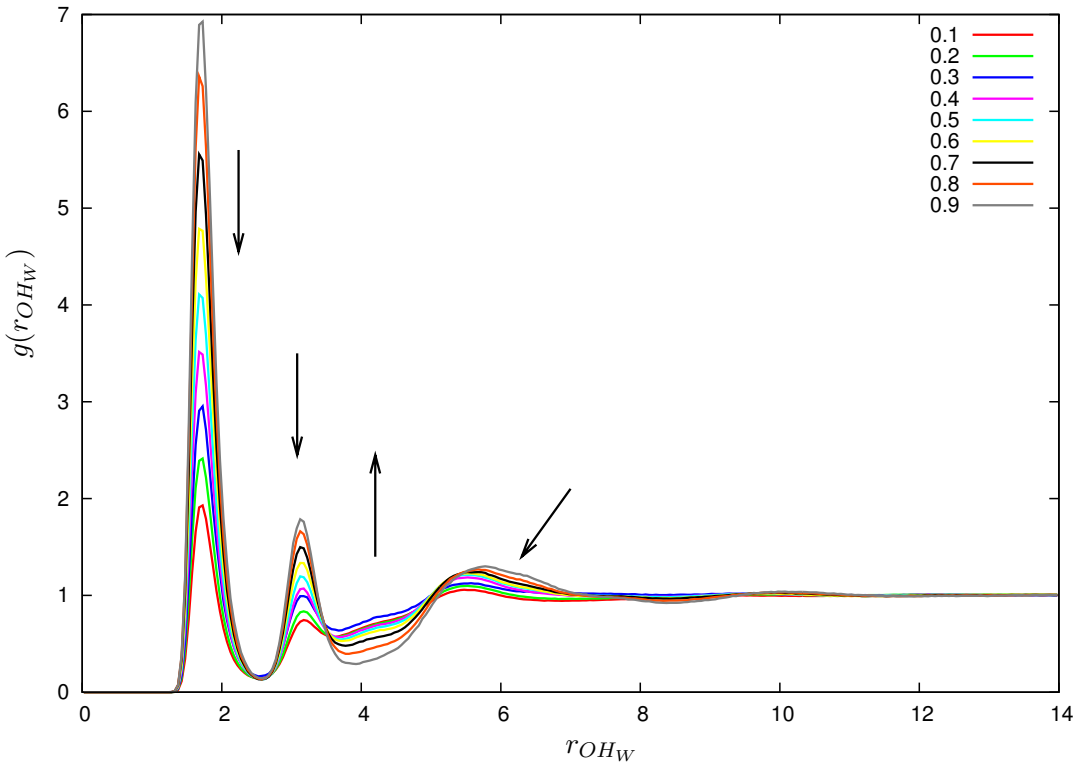

Figure 6.15: $g(r_{OH_W})$ radial distribution for interactions between the carbonyl oxygen of the NMA molecule and water hydrogens.

waters and therefore increases as the NMA mole fraction decreases. The broad third peak in the radial distribution between 5 and 7Å arises from configurations at all angles resulting in a band labelled D in Figure 6.16. This band is due to a variety of both hydrogen bonded and non hydrogen bonded configurations. The peak within this band which is more prominent in the concentrated solution shown in Figure 6.16(a) is due to the correlation between the carbonyl oxygen and water molecules accepting a hydrogen bond from the amide group on the opposite side of the NMA molecule. Table 6.7 lists the peak positions and coordination numbers for the different concentrations simulated and shows that unlike the other inter-species hydrogen bond (HO_W), the length of the OH_W hydrogen bond does not seem to change with concentration. Admittedly, the position of the first peak does decrease by 0.05Å for concentrations between $X_{NMA} = 0.2$ and 0.8 but returns to its $X_{NMA} = 0.1$ value at $X_{NMA} = 0.9$. However, the separation of the donor and acceptor atoms as indicated by the first peak in $g(r_{OO_W})$ does not change with concentration and as a change of 0.05Å corresponds to a single bin width in the histogram used to collect data, the conclusion that the length of the OH_W hydrogen bond does not change with concentration appears to be reasonable.

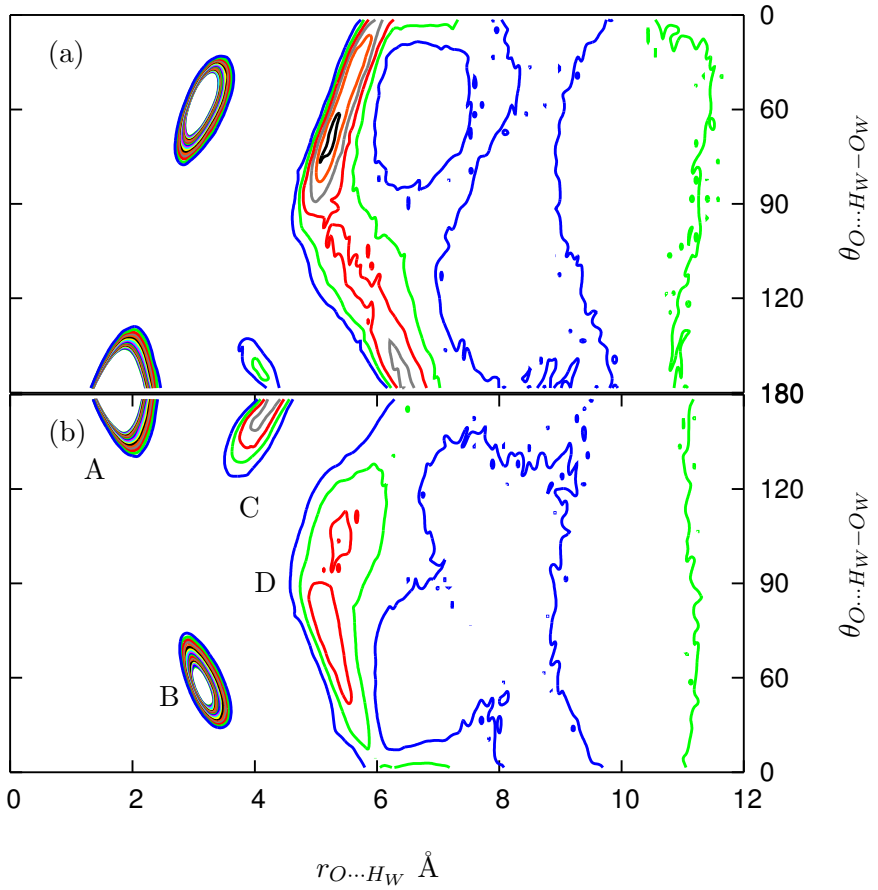

Figure 6.16: Contour plot of the $g_2(r_{O\cdots H_W}, \theta_{O\cdots H_W-O_W})$ radial-angular distribution for interactions between water hydrogens and the acceptor carbonyl group of the NMA molecule in (a) $X_{NMA} = 0.8$ and (b) $X_{NMA} = 0.4$ solutions. In this interaction the acceptor is the NMA oxygen and the donor atom is the water oxygen. Contour levels start at $g_2(r, \theta) = 1$ increasing in 0.2 increments to 5.

6.3.4 Discussion of interspecies hydrogen bonding geometries

The results presented above show that both NMA and water molecules form interspecies hydrogen bonds. The hydrogen bond formed when the NMA amide hydrogen is donated to a water oxygen is slightly longer than the hydrogen bond formed when water hydrogens are donated to the NMA carbonyl oxygen. Therefore one expects that the $O \cdots H_W$ hydrogen bond is stronger than $O_W \cdots H$ and indeed it has been reported by Köddermann and Ludwig that the hydrogen bond in which water donates a proton to the NMA carbonyl group is 50% stronger than the one in which water accepts a proton from NMA[57]. The $O_W \cdots H$ hydrogen bond length decreases by 0.1Å on increasing the NMA concentration from $X_{NMA} = 0.1$ to $X_{NMA} = 0.9$ implying that this hydrogen bond weakens as the proportion of water in the system is increased. In

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
0.1	OH_W	1.725	2.575	1.75
0.2	OH_W	1.725	2.525	1.49
0.3	OH_W	1.725	2.525	1.33
0.4	OH_W	1.675	2.575	1.11
0.5	OH_W	1.675	2.575	0.92
0.6	OH_W	1.675	2.575	0.75
0.7	OH_W	1.675	2.575	0.57
0.8	OH_W	1.675	2.575	0.37
0.9	OH_W	1.725	2.625	0.20

Table 6.7: Extrema positions and coordination number for the 1^{st} peak in $g(r_{OH_W})$: hydrogen bonds accepted by the NMA oxygen from water hydrogens.

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
0.1	OO_W	2.725	3.325	1.85
0.2	OO_W	2.725	3.375	1.60
0.3	OO_W	2.725	3.425	1.45
0.4	OO_W	2.725	3.425	1.16
0.5	OO_W	2.725	3.475	0.96
0.6	OO_W	2.725	3.475	0.77
0.7	OO_W	2.725	3.525	0.59
0.8	OO_W	2.725	3.525	0.38
0.9	OO_W	2.725	3.625	0.20

Table 6.8: Extrema positions and coordination number for the 1^{st} peak in $g(r_{OO_W})$: hydrogen bonds accepted by the NMA oxygen from water hydrogens, in this case the water oxygen is acting as a donor atom.

contrast the $O \cdots H_W$ hydrogen bond length is invariant with respect to concentration.

6.4 Quantitative hydrogen bonding analysis

The previous sections have shown that NMA and water molecules interact in a way which can be qualitatively described as hydrogen bonding. In this section the problems associated with quantitative definitions of hydrogen bonds are discussed and a specific definition is presented and justified. This definition is then used to characterise the intermolecular hydrogen bonded structure as a function of concentration.

6.4.1 Definition of intra- and inter-species hydrogen bonds

In Chapter 5 two NMA molecules were defined as hydrogen bonded if

- Their hydrogen to acceptor (oxygen) separation was less than the position of the first minimum in $g(r_{O\cdots H})$

- $\theta_{O\cdots H-N}$ was between 140° and 180°

The first criterion is easily justified by referral to $g(r_{OH})$ as it implies that all molecular pairs contributing to the first peak of the radial distribution are considered to be hydrogen bonded. The precise choice of angle in the second criterion is essentially arbitrary and was arrived at by an assessment of how a variation in cutoff angle affected the hydrogen bonding characteristics of pure NMA. This was described in Section 5.4.

The key question to consider is whether or not the definition of a hydrogen bond should be fixed or allowed to vary as a function of the donor - acceptor pairing, concentration or external variables such as temperature. An example of a fixed definition would be taking the cutoff criteria used for liquid NMA ($r_{ON} < 3.675$Å, $r_{OH} < 3.075$Å, $\theta_{OHN} < 140°$) and applying these numerical values to all other NMA - NMA interactions in the full range of aqueous solutions. By extension of this logic, exactly the same "hydrogen bond" criteria should be applied to the NMA donating to water, water donating to NMA and water - water interactions. However, a cursory glance at the radial distributions shows that these four are actually different types of interaction, each with a different mean separation and angular variance. The alternative is to choose a definition of hydrogen bonding which allows for this variation between different interactions. This can be achieved by identifying the distance cutoff as the location of the first minimum in the relevant radial distribution function. Therefore the numerical value of the cutoff distance varies with atom type, concentration and temperature. The choice of angular criteria is more problematic. Since directionality is a crucial aspect of hydrogen bonding it is necessary to impose some kind of constraint. In the literature a cutoff angle of 150° is commonly used for water-water hydrogen bonding whilst 140° was identified as a suitable cutoff for liquid NMA based on the analysis in Section 5.4. This begs the question as to the correct choice of cutoff angle for defining interspecies hydrogen bonds. For consistency it was decided to use a cutoff angle of 140° to define all four types of hydrogen bonds in the system: NMA-NMA, NMA-water, water-NMA and water-water. Therefore the actual values given in the following results will not necessarily be comparable to other investigations which use different hydrogen bonding criteria. As there are many different choices of such criteria the main aim of this and subsequent sections is to describe trends resulting from changes in concentration.

The working definition of hydrogen bonding used for analysis of all the aqueous solutions and pure liquids unless otherwise stated is therefore:

- Hydrogen - acceptor separation less than the position of the first minimum in $g(r_{HA})$

- Hydrogen bond angle $\theta_{A\cdots H-D}$ between 140° and 180°

6.4.2 Average number of hydrogen bonds per molecule

Using the above definition of hydrogen bonding the simulations can be analysed to investigate the relative propensities of NMA and water to form intra- and inter-species hydrogen bonds. The average number of hydrogen bonds per NMA molecule is plotted

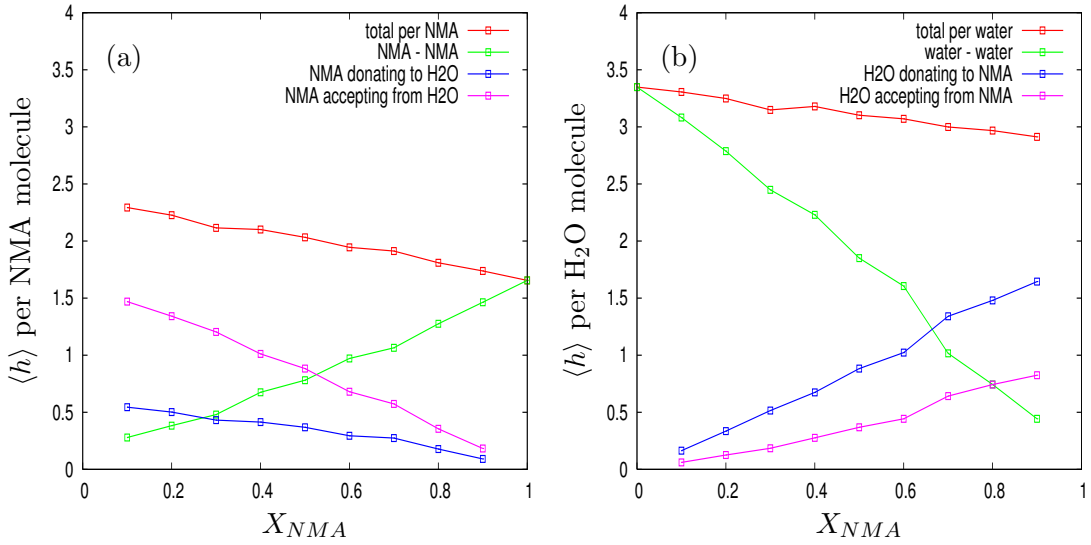

Figure 6.17: These figures show the variation with concentration of the average number of hydrogen bonds per molecule of species A (red); the average number of hydrogen bonds per species A molecule formed with other molecules of the same species A (green); the average number of hydrogen bonds donated by a species A molecule to species B molecules (blue); the average number of hydrogen bonds per species A molecule accepted from species B molecules (pink). In (a) species A = NMA and B = water whereas in (b) A = water and B = NMA.

in Figure 6.17(a). In general the NMA molecule is considered to possess two hydrogen bond acceptor sites on the carbonyl oxygen but only one donor group. Therefore an individual NMA molecule can form a maximum of 3 hydrogen bonds but because there are twice as many acceptor sites as donor atoms this maximal number is rarely achieved in the pure liquid. In fact, the average number of hydrogen bonds per molecule is 1.66. Introducing water into the system does not change the number of excess acceptor sites as each water molecule has two acceptor sites which are balanced exactly by its two donor groups. If the inter-species interaction is "ideal" meaning that it is the arithmetic average of the pure species interactions then the average number of hydrogen bonds

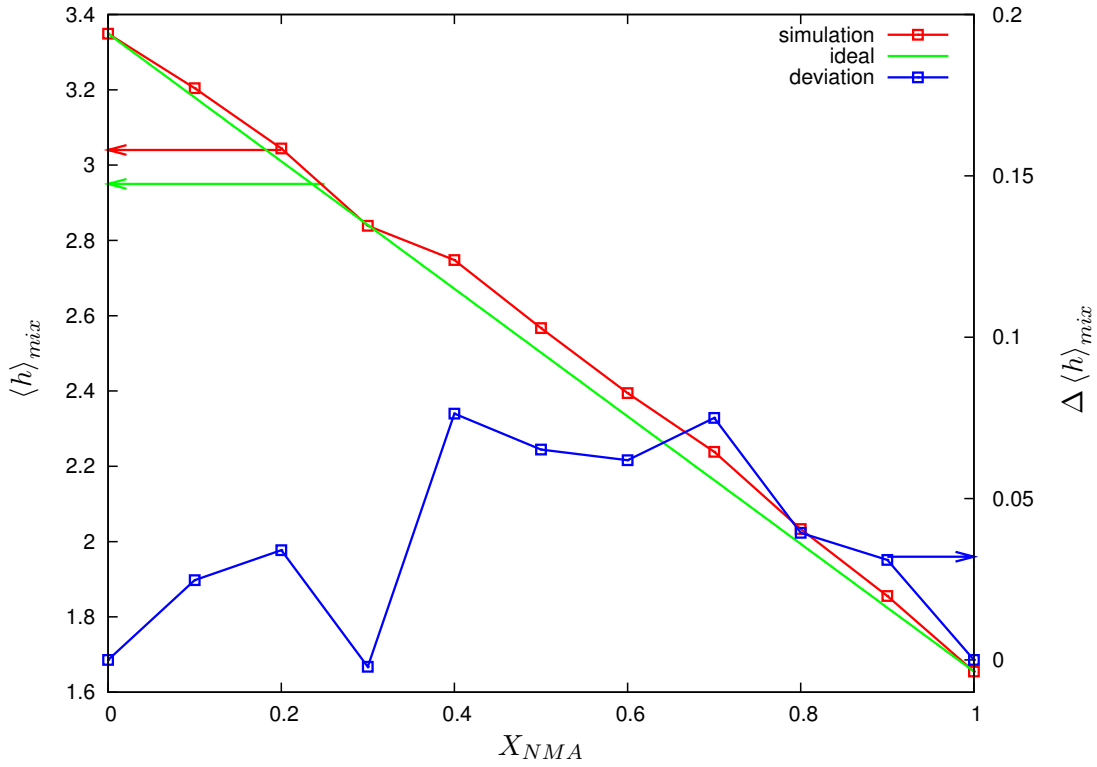

Figure 6.18: The average number of hydrogen bonds per molecule in the simulations as a function of NMA mole fraction (red). In this case the average is taken over both NMA and water molecules. This is compared with the values which would be obtained if the two liquids mixed ideally in accordance with Equation 6.1 (green). The deviation $\Delta < h >_{mix}$ of the simulation result from the ideal value is plotted in blue against the right hand y axis. The value obtained for $X_{NMA} = 0.3$ is probably anomalous because of the high proportion of trans-NMA which converted to cis-NMA during this simulation.

per molecule in the mixture can be described by

$$< h >_{ideal\ mixture} = X_{NMA} < h >_{NMA} + (1 - X_{NMA}) < h >_{H_2O} \qquad (6.1)$$

This function is plotted in Figure 6.18 along with the actual values found in the simulations. From this it can be seen that the interspecies interactions are not ideal although the average number of hydrogen bonds per molecule in the mixture is only slightly greater than that predicted by ideal mixing. Calculations using the Kirkwood-Buff theory of solutions for a binary mixture of NMA and water also found only slight deviations from ideality[138]. This was interpreted to mean that in spite of interactions between the amide and water, the distribution of molecules in the solution is essentially random. The small deviation of the simulation result from the ideal value $\Delta < h >_{mix} = < h >_{simulation} - < h >_{ideal}$ is also plotted as a function of NMA mole fraction. This would ideally show a smooth dependence on concentration as do

the excess functions discussed in Chapter 2. The angularity probably indicates that the simulations have not fully sampled the phase space of these aqueous solutions. The excess volume and viscosity deviations measured experimentally for a range of aqueous NMA solutions by Victor and Hazra show extrema at $X_{NMA} = 0.4$ indicative of association through hydrogen bonding[91]. While the results of these simulations obviously support the assertion of hydrogen bonded associations between water and NMA it is not possible to identify a specific mole fraction at which the maximum deviation from ideality occurs. The result for $X_{NMA} = 0.3$ is significantly different to those of the other simulations which can probably be attributed to the high proportion of *cis*-isomers which formed at this concentration. Without further investigation it is not possible to say whether the almost perfectly ideal behaviour of the $X_{NMA} = 0.3$ simulation can be accounted for by statistical variation or if it is a true reflection of the different hydrogen bonding behaviour of aqueous *cis*-NMA.

6.4.3 Donor and acceptor behaviour

In addition to calculating the average number of hydrogen bonds per molecule, the cluster analysis also identifies the specific number and type of hydrogen bonds donated or accepted by each molecule and these results are plotted in Figure 6.19. First of all consider the donor behaviour of the NMA molecule characterised in Figure 6.19(a). As the NMA mole fraction decreases in the aqueous solutions the percentage of NMA molecules which donate to other NMAs decreases while the percentage which donate to water increases. The crossover from the majority donating to NMA to the majority donating to water occurs at $X_{NMA} = 0.5$ suggesting that the propensity of the amide group to donate to water is the same as to other NMA molecules. The situation is not quite this clear cut however, as the proportion of free amide groups which do not form hydrogen bonds increases with hydration as illustrated in Figure 6.20. In liquid NMA only 17% of the amide hydrogens are not donated to form hydrogen bonds with other NMA molecules. This result is in reasonable agreement with experimental estimations of the number of free amide groups in the liquid[62]. On hydration, the percentage of free amide groups increases to 32% in $X_{NMA} = 0.1$ solution. This implies that NMA is not a very effective donor of hydrogen bonds to water molecules in accordance with other studies[57, 71]. The high percentage of free amide groups at $X_{NMA} = 0.3$ may indicate that *cis*-NMA isomers are less able to donate hydrogen bonds than *trans*-isomers. This result explains the lower than average number of hydrogen bonds formed in the $X_{NMA} = 0.3$ solution as found in the previous section.

In Section 6.4.2 it was found that the average number of hydrogen bonds per NMA molecule increased with hydration. As the number of amide groups participating

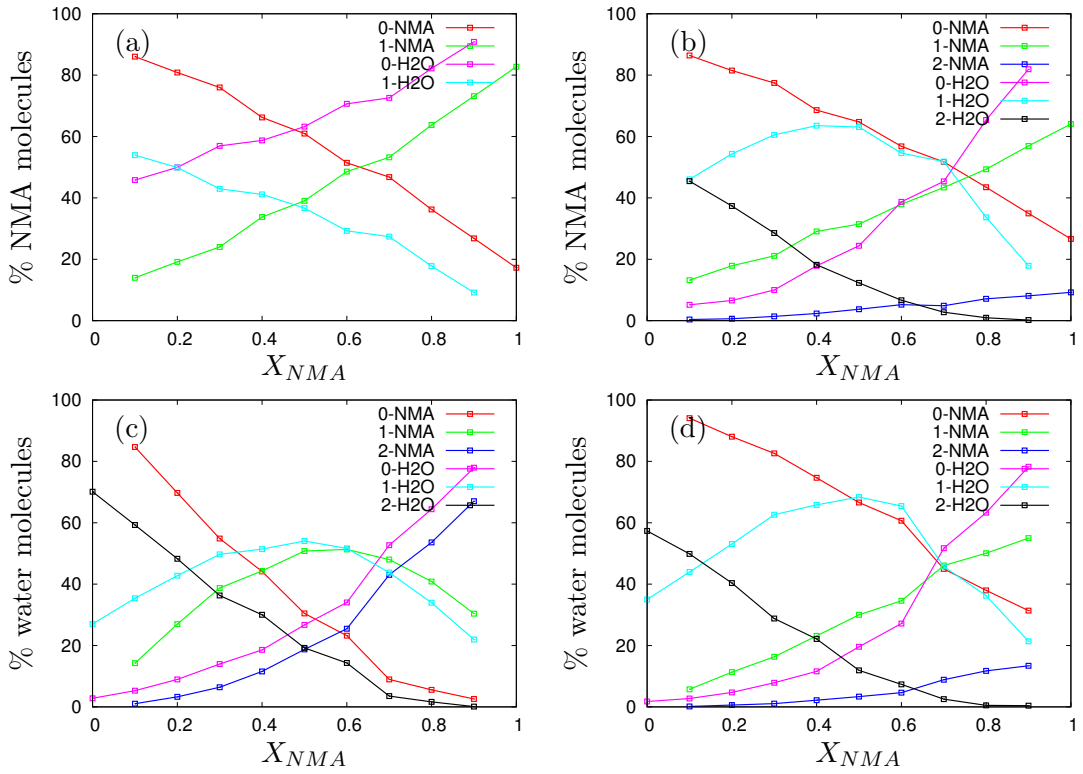

Figure 6.19: The percentage of NMA or water molecules donating or accepting 0 (red), 1 (green) or 2 (blue) hydrogen bonds to/from NMA molecules and donating or accepting 0 (pink), 1 (turquoise) or 2 (black) hydrogen bonds to/from water molecules averaged over all timesteps. (a) NMA donor behaviour, (b) NMA acceptor behaviour, (c) water donor behaviour, (d) water acceptor behaviour.

in hydrogen bonds decreases with increasing water mole fraction the deficit must be compensated for by a large increase in the number of hydrogen bonds accepted by the NMA carbonyl groups. Figure 6.19(b) shows the acceptor behaviour of NMA molecules in the simulations. The percentages of NMA molecules accepting 1 or 2 hydrogen bonds from other NMAs decrease proportionately to the decrease in NMA mole fraction. Conversely, the number of hydrogen bonds accepted from water molecules increases much more rapidly in the NMA-rich region than expected from the change in concentration.

A relatively high proportion of NMA carbonyl groups do not form hydrogen bonds even in liquid NMA (26.6%). Consequently, in the NMA-rich region water molecules typically donate their hydrogen atoms to these free carbonyl groups at the ends of NMA hydrogen bonded chains as seen in Figure 6.21 for the interaction of the central water with the rightmost NMA molecule. It is also possible for water molecules to be inserted into existing NMA chains at the free lone pair site of carbonyl oxygens which already

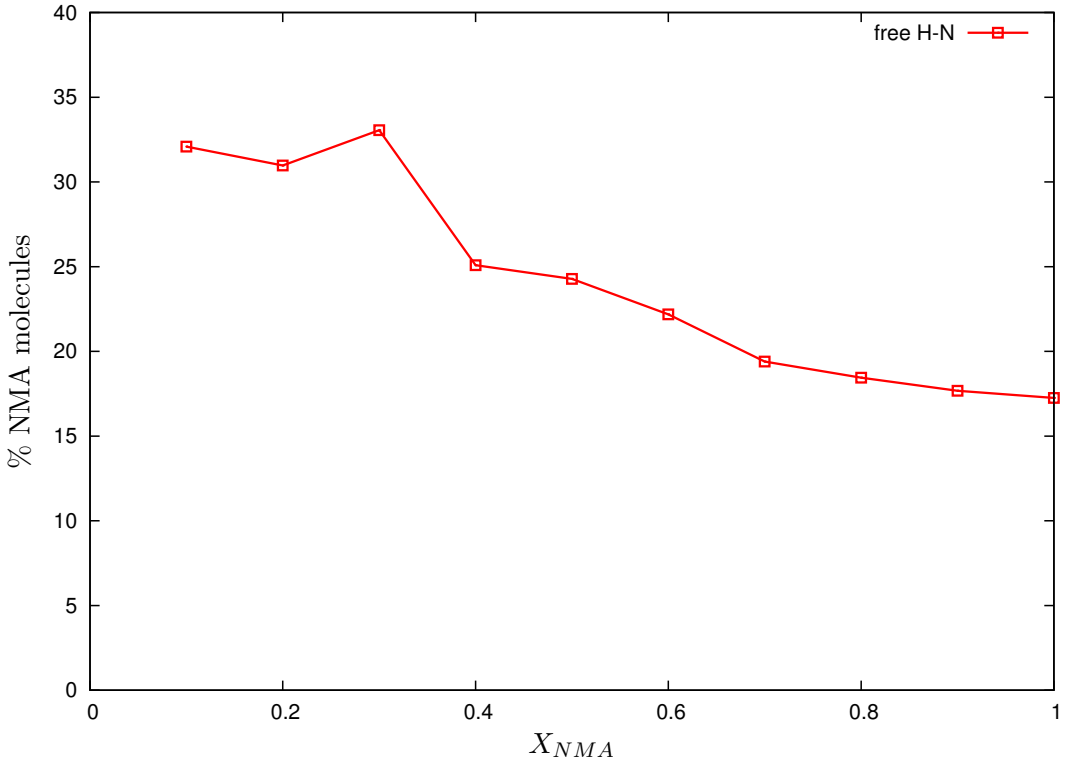

Figure 6.20: % of NMA molecules which are not hydrogen bonded through the $N-H$ group. The value obtained for $X_{NMA} = 0.3$ is probably anomalous because of the high proportion of trans-NMA which converted to cis-NMA during this simulation.

have one hydrogen bond to another NMA molecule. This is illustrated in Figure 6.21 where the NMA molecule at the bottom left of the simulation snapshot is part of an NMA chain but has also formed a hydrogen bond to the central water. 67% of the water molecules at $X_{NMA} = 0.9$ donate both their hydrogen atoms to NMA carbonyl groups forming the bridge-like structure seen in Figure 6.21 which is responsible for the increase in the low-r wing of $g(r_{OO})$ with hydration.

Even in the NMA-rich region it is rare for waters to accept two hydrogen bonds from NMA molecules as can be seen from the plot of water acceptor behaviour shown in Figure 6.19(d). At $X_{NMA} = 0.9$, 55% of the water molecules accept a single hydrogen bond from an NMA and therefore the typical structural motif for high NMA concentrations is that seen in Figure 6.21 where water molecules donate both hydrogens to NMA carbonyl groups but only accept one hydrogen bond from an NMA amide group.

The characteristic hydrogen bonded water structure found in the NMA-rich simulations is in agreement with the FT-IR spectroscopy results of Czarnecki and Haufa[61]. They

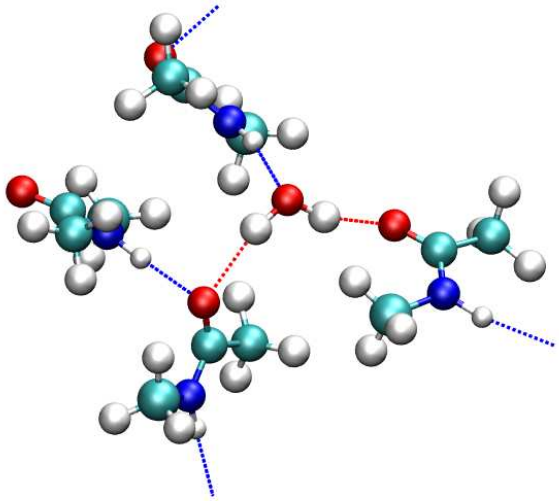

Figure 6.21: A typical hydrogen bonded water configuration from the $X_{NMA} = 0.9$ simulation. Dashed lines indicate hydrogen bonds donated by NMA (blue) and water (red).

found no evidence for free water hydrogens in the spectra implying that all water molecules donate both hydrogens to form hydrogen bonds with two different NMA molecules as has also been found in these simulations. The percentages of NMA molecules accepting hydrogen bonds from water molecules are in reasonable agreement with the published values of Zhang *et al* using the OPLS-AA and SPC forcefields, although their distribution for NMA molecules accepting a single hydrogen bond from water peaks at $X_{NMA} = 0.35$ rather than at $X_{NMA} = 0.45$[71].

Most investigations of aqueous NMA have studied the dilute regime where water molecules significantly outnumber NMA. *Ab initio* simulations of NMA n-H_2O clusters predict that two water molecules hydrogen bond to the lone pair electrons on the carbonyl oxygen while a third water accepts a hydrogen bond from the amide group[139]. This hydration structure is indeed found in the simulations as illustrated by snapshot (a) shown in Figure 6.22, however, for all concentrations it is more common for NMA carbonyl oxygens to accept one rather than two hydrogen bonds from waters as seen in Figure 6.19(b). This tendency leads to the prevalence of configurations such as that shown in (b) of Figure 6.22. In fact, *ab initio* simulations have shown that this configuration in which the carbonyl oxygen accepts only one hydrogen bond from water is more favourable than (a)[67]. It was found that when an NMA molecule accepts a single hydrogen bond from a water molecule the formation of a second hydrogen bond by donation of the amide group to water strengthens the original bond (positive cooperativity) while the acceptance of a second hydrogen bond from water by the carbonyl group weakens the original (negative cooperativity). The proportion of free

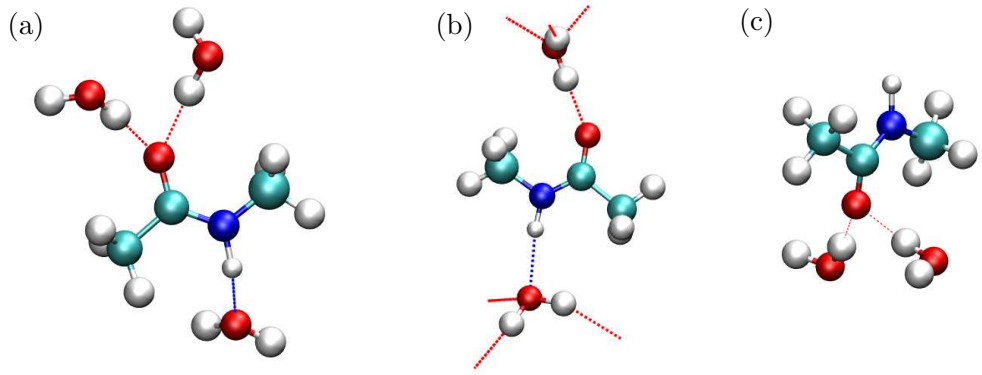

Figure 6.22: Typical hydrogen bonded NMA configurations from the $X_{NMA} = 0.1$ simulation. Dashed lines indicate hydrogen bonds donated by NMA (blue) and water (red).

amide groups increases with water mole fraction so in the dilute regime configurations such Figure 6.22(c) where an NMA molecule accepts one or two hydrogen bonds from water molecules account for approximately 30% of the NMA molecules.

When the NMA mole fraction increases above $X_{NMA} = 0.2$ more water molecules accept a single hydrogen bond from other waters than accept two. By $X_{NMA} = 0.3$ the same transition has occurred for water donors, therefore, above this concentration water molecules form linear rather than tetrahedrally coordinated structures. This is in spite of the fact that at this concentration water tends to form large almost system size clusters as will be shown in Section 6.6.2 but is in agreement with the results of Section 6.6.3 which will show an immediate decrease in the tetrahedral order parameter for "bulk" water molecules at even low NMA concentrations. The proportion of water molecules donating a single hydrogen bond to other waters/NMA reaches a maximum at $X_{NMA} = 0.5/0.6$. $X_{NMA} = 0.5$ similarly marks a maximum in the proportion of water molecules accepting a single hydrogen bond from NMA molecules as can be seen from Figure 6.19(d). The percentage of water molecules which donate two hydrogen bonds in a $X_{NMA} = 0.9$ mixture is only slightly less than the proportion which donate two hydrogen bonds in pure water. In contrast, there is a large decrease in the percentage of water molecules accepting two hydrogen bonds, from just under 60% in pure water to 14% in $X_{NMA} = 0.9$ aqueous NMA. Compare this to the increase in the proportion of NMA molecules accepting two hydrogen bonds which rises from 9% in liquid NMA to 46% in $X_{NMA} = 0.1$ from which it is clearly seen that water molecules prefer to donate to the NMA carbonyl oxygen rather than to their own species. This is in agreement with an experimental study of NMA-water complexes in carbon tetrachloride[57].

6.4.4 Discussion of hydrogen bonding in aqueous NMA

The average number of hydrogen bonds per molecule is higher in the aqueous solutions than predicted by supposing an ideal combination of the pure components and shows that interspecies hydrogen bonding is a significant factor in the intermolecular structure of the mixtures. There are more "free" amide hydrogens in the $X_{NMA} = 0.1$ NMA solution than are present in pure NMA indicating that both NMA and water oxygens prefer to accept water hydrogens when forming hydrogen bonds. This is logical since the OH and $O_W H$ hydrogen bonds are the longest and therefore presumably the weakest hydrogen bonds present in the mixture. Assuming the relative strengths of the hydrogen bonds are in proportion to their lengths the strongest is the OH_W bond with a length of about 1.7Å, followed by the $O_W H_W$ bond at 1.8Å, the OH bond at 1.9Å and finally the $O_W H$ bond whose length varies from 1.9Å in concentrated NMA to 2.0Å in dilute solution. Just over a quarter of NMA molecules in the pure liquid do not accept hydrogen bonds through their carbonyl groups and it is to these free lone pairs of electrons that water molecules principally attach upon hydration. Water molecules can also associate with NMA hydrogen bonded chains without breaking the chain itself as each singly hydrogen bonded carbonyl oxygen has a second lone pair of electrons with which to form a second hydrogen bond. This indicates that the intermolecular hydrogen bonded structure found in liquid NMA may not be significantly disrupted by the addition of small amounts of water. Conversely, the tetrahedral coordination of bulk water is broken down by the addition of even small proportions of NMA. Once the NMA mole fraction has reached 0.3 most waters are two rather than four coordinated with other waters as water prefers to donate its hydrogens to NMA molecules rather than to its own species.

6.5 Interspecies methyl group interactions

In Section 6.2.6 the effect of hydration upon methyl group interactions with the NMA oxygen were discussed and it was concluded that the observed changes were consistent with the break-up of NMA hydrogen bonded chains. In addition to weak hydrogen bonding interactions between NMA molecules there is now also the possibility that associations will form between methyl hydrogens and water oxygens. In fact, on the basis of DFT simulations and measurements of IR spectra Zhang *et al* [71, 72, 89] report the presence of weak or improper hydrogen-bonded contacts in aqueous NMA solutions and find that the carbonyl methyl group is favoured in forming weak C-H...O_W contacts. In this section the classical simulations are investigated to see whether or not the CHARMM22 and F3C forcefield produces comparable results. Figure 6.23 shows joint radial-angular distributions for interactions of the two distinct types of

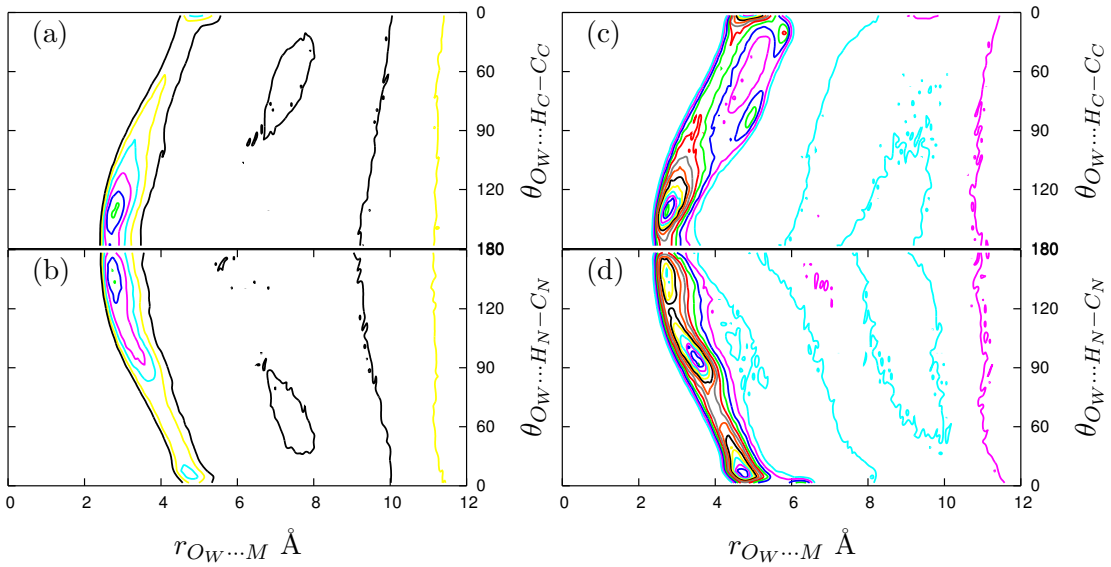

Figure 6.23: Contour plots of radial angular distributions (a) $g_2(r_{O_W\cdots H_C}, \theta_{O_W\cdots H_C-C_C})$ and (b) $g_2(r_{O_W\cdots H_N}, \theta_{O_W\cdots H_N-C_N})$ in $X_{NMA} = 0.1$ solution. Contour levels begin at $g_2(r,\theta) = 1$ (black), increasing in 0.2 increments. (c) and (d) show the same distributions for $X_{NMA} = 0.8$ and in these plots the $g_2(r,\theta) = 1$ contour is turquoise.

methyl hydrogen with the water oxygen atom in dilute $X_{NMA} = 0.1$ and concentrated $X_{NMA} = 0.8$ solutions.

The general shape of both of these radial angular distributions is the same as for interactions between the NMA oxygen and methyl groups. However, it is immediately apparent that while the specific features of $g_2(r_{OH_C}, \theta_{OH_CC_C})$ and $g_2(r_{OH_N}, \theta_{OH_NC_N})$ were quite different in liquid NMA indicating that the amide and carbonyl methyl groups interact differently with the NMA oxygen, in Figure 6.23(a) and (b) the two distributions are very similar indicating that when in dilute solution both methyl groups interact in much the same way with the water oxygen. In fact, both types of interaction resemble that of the carbonyl methyl group with the NMA oxygen. In each case there is a peak at $g_2(r = 2.9\text{Å}, \theta = 135°)$. By analogy with the situation in pure NMA the contributions to this peak come from interactions between methyl groups and water molecules hydrogen bonded to either the amide or carbonyl group of NMA. As before the three-fold degeneracy of methyl hydrogens, the flexibility of the NMA CHARMM model which permits methyl rotation, the flexibility of the hydrogen bonded interaction itself and the contributions of molecules in neighbouring chains result in a continuum of $\theta_{O_WMC_M}$ angles. At higher NMA concentrations the range of separation distances increases. Figure 6.23(d) shows that at $X_{NMA} = 0.8$ $g_2(r_{O_WH_N}, \theta_{O_WH_NC_N})$ which was very similar to $g_2(r_{OH_C}, \theta_{OH_CC_C})$ in the dilute regime now has features which are much

more similar to those of the amide methyl groups interactions with the NMA oxygen. From this concentration dependent trend it is possible to conclude that in general the carbonyl methyl hydrogens are more likely to form associations with water oxygens but as the proportion of water in the system increases this proclivity is swamped.

6.6 Water-water interactions

Thus far in this chapter the effect of hydration upon NMA chains has been investigated and interspecies interactions between the two components of the aqueous NMA system have been explored. Now attention is paid to the solvent and the effect of an increasing concentration of peptide groups upon the hydrogen bonded water network is studied.

6.6.1 $g(r_{O_W H_W})$: hydrogen bonding

First of all consider the effect which the addition of NMA has on the radial distribution function between the water oxygens and hydrogens which is shown in Figure 6.24.

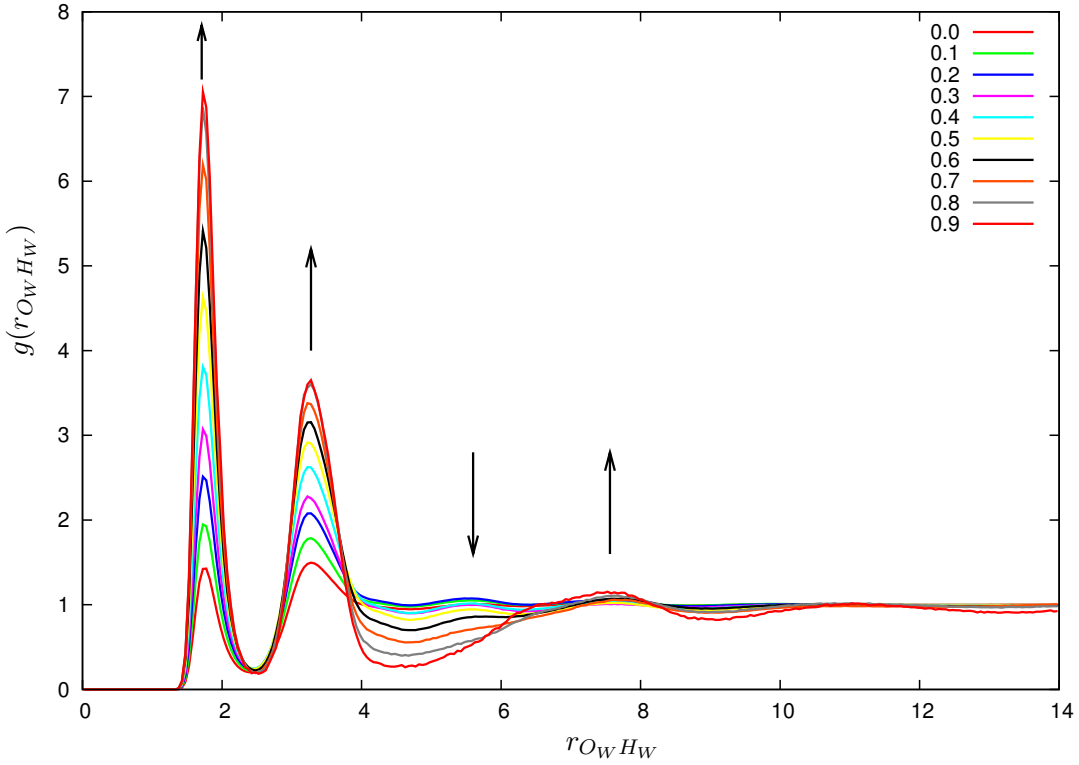

Figure 6.24: $g(r_{O_W H_W})$ radial distribution for interactions between the water oxygens and hydrogens. Arrows indicate the changes on increasing NMA mole fraction.

From this distribution it is apparent that some water molecules continue to be hydrogen bonded to each other even at the highest NMA concentration where the simulated

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
F3C	O_WH_W	1.775	2.425	1.9
0.1	O_WH_W	1.725	2.425	1.7
0.2	O_WH_W	1.725	2.425	1.5
0.3	O_WO_W	1.725	2.425	1.3
0.4	O_WH_W	1.725	2.475	1.2
0.5	O_WH_W	1.725	2.475	1.0
0.6	O_WH_W	1.725	2.475	0.8
0.7	O_WH_W	1.725	2.475	0.6
0.8	O_WH_W	1.725	2.475	0.4
0.9	O_WH_W	1.725	2.425	0.2

Table 6.9: Extrema positions and coordination numbers for the 1^{st} peak in $g(r_{O_WH_W})$

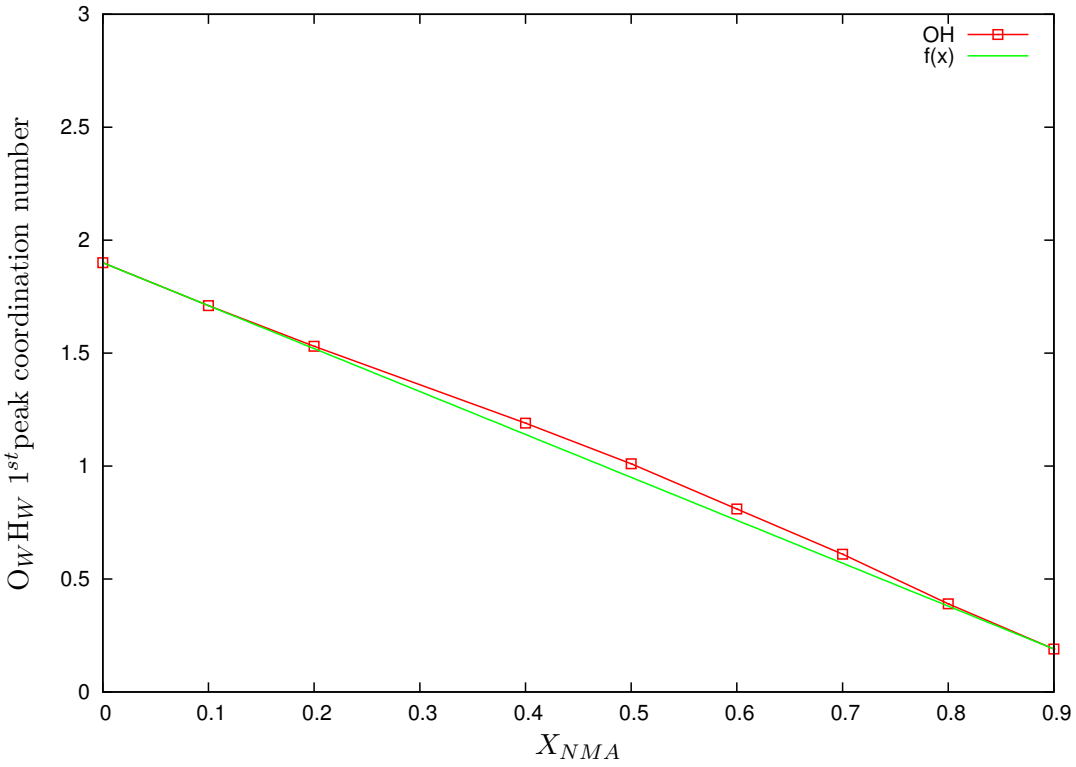

Figure 6.25: Coordination numbers for the 1^{st}peak in $g(r_{O_WH_W})$ together with a linear interpolation between the endpoints.

system contains only 23 waters compared to 207 of the larger NMA molecules. The vast difference between the local density for these configurations and the bulk water density in the NMA-rich regime causes the large increase in the $g(r)$ peak heights indicated by the arrows in Figure 6.24. Coordination numbers for the first peak of $g(r_{O_WH_W})$ are listed in Table 6.9 from which it can be seen that over the concentration range there is a ten-fold decrease from 1.90 in pure F3C water to 0.19 for the $X_{NMA} = 0.9$ solution.

As Figure 6.25 illustrates there is slightly more interaction between water molecules at intermediate NMA concentrations than one would expect from a linear relationship of the coordination number with NMA mole fraction. The indistinct peak at about 5Å due to second and third neighbour waters decreases with increasing NMA mole fraction while a new peak appears at about 7.5Å as indicated by an arrow in Figure 6.24. This new peak is due to waters separated by an NMA molecule.

6.6.2 Hydrogen bonded clusters

In F3C water hydrogen bonds form a system size network. This is shown in Figure 6.26

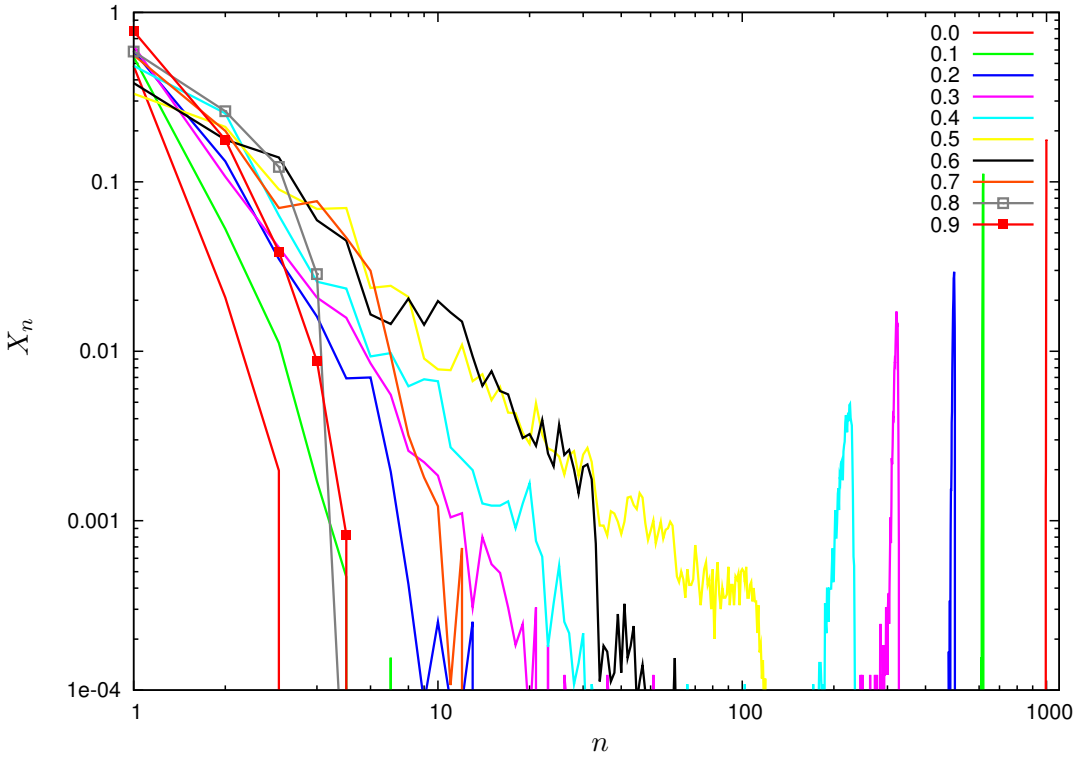

Figure 6.26: Hydrogen bonded water cluster population distributions as a function of cluster size for water molecules in aqueous NMA solutions. Note that the maximum possible cluster size depends on the total number of water molecules in the simulation which varies as a function of concentration. The spikes in the cluster distributions at large n for concentrations less than $X_{NMA} = 0.4$ indicate system size water clusters.

where the mole fraction of water clusters $X_n = \frac{N_n}{N_{tot}}$ is plotted against the cluster size, n, where N_n is the number of clusters of size n molecules and N_{tot} is the total number of water clusters identified in the simulation.

The spikes in the cluster populations for concentrations less than $X_{NMA} = 0.4$ are due to the formation of system size hydrogen bonded water clusters. The jaggedness

of the distributions simply reflects poor statistics in the sampling of infrequent cluster sizes and with longer simulation runs it is to be expected that these distributions would become smooth even for the medium size clusters. Table 6.10 lists the average percentage of the total number of water molecules contained in monomers, dimers, trimers and tetramers during the simulations. On average, in pure F3C water only a

X_{NMA}	$n=1$	$n=2$	$n=3$	$n=4$
F3C	9.78E-2	8.4E-3	1.2E-3	0.0
0.1	0.23	4.4E-2	1.4E-2	2.8E-3
0.2	0.56	0.25	9.96E-2	6.08E-4
0.3	1.26	0.43	0.24	0.16
0.4	2.23	2.34	0.88	0.48
0.5	4.53	5.77	3.70	3.79
0.6	8.86	8.21	9.67	5.50
0.7	27.73	19.78	10.40	15.23
0.8	36.91	32.81	23.09	7.17
0.9	60.30	27.60	9.06	2.72

Table 6.10: Percentages of water molecules contained in clusters of size $n = 1, 2, 3$ or 4

tiny proportion of water molecules are disconnected from the system size network. This is clearly shown in Figure 6.27 where the percentage of water molecules as a cumulative function of the cluster size appears flat until the cluster size equals the system size, at which point there is a spike indicating that almost 100% of the water molecules are contained in system size clusters. Up to an NMA mole fraction of 0.4 the spikes in Figure 6.27 indicate that the vast proportion of the molecules are contained in system size clusters. At $X_{NMA} = 0.5$ the shape of the distribution changes corresponding to a more even distribution of molecules across all possible cluster sizes. In the most concentrated NMA solution only 3% of the water molecules are contained in clusters of four or more waters indicating that the waters are dispersed separately through the NMA intermolecular structure rather forming self-associated clusters. This corrobates previous results which found that water molecules prefer to donate to NMA oxygens rather than to their own species.

6.6.3 $g(r_{O_WO_W})$: hydrogen bonding and tetrahedral order

As was discussed in Chapter 4 the tetrahedral intermolecular structure of pure water is characterised by the second peak of $g(r_{O_WO_W})$ at approximately 4.5Å as shown in Figure 6.28. Since this peak arises from second nearest neighbour contacts, water molecules participating in hydrogen-bonded clusters of less than three molecules will not contribute to this peak. As the mole fraction of water in the mixtures decreases this peak diminishes until it is barely present in the most concentrated solution. Therefore

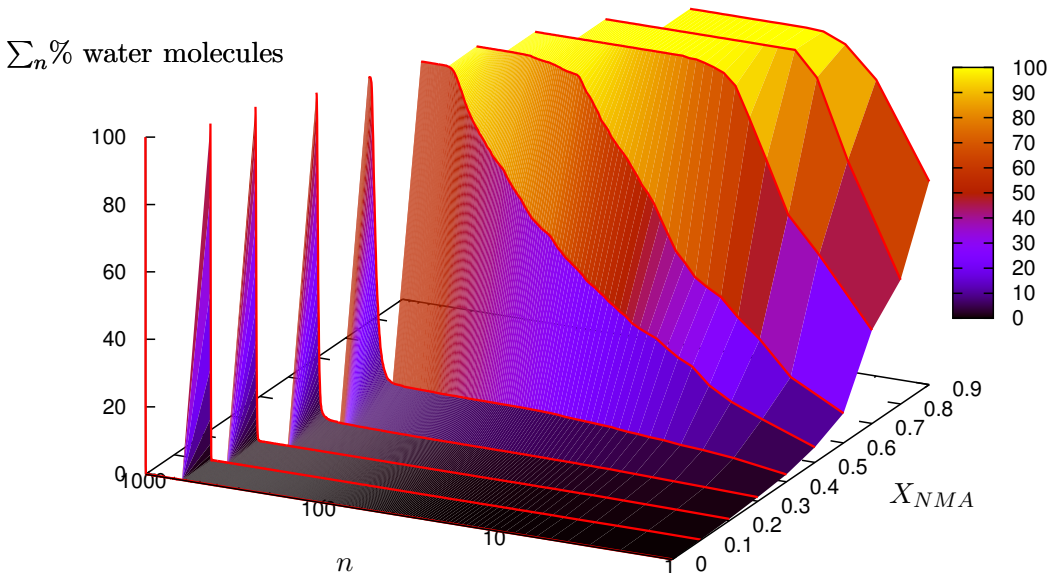

Figure 6.27: The cumulative % of water molecules included in clusters as a function of the number of molecules in the cluster n and the concentration X_{NMA}. At low NMA mole fractions only a tiny percentage of water molecules are contained in small clusters, almost all are contained in system size clusters, hence the spikes in distributions for $X_{NMA} \leq 0.4$. At $X_{NMA} = 0.5$ the shape of the distribution changes, showing that molecules are more evenly distributed across all possible cluster sizes. Above $X_{NMA} = 0.7$ it can be seen that almost all the molecules are contained in clusters of less than ten molecules.

X_{NMA}	g(r)	1^{st} peak position (Å)	1^{st} min position (Å)	n(r)
F3C	$O_W O_W$	2.775	3.425	4.9
0.1	$O_W O_W$	2.775	3.425	4.2
0.2	$O_W O_W$	2.775	3.425	3.6
0.3	$O_W O_W$	2.775	3.475	3.1
0.4	$O_W O_W$	2.775	3.475	2.6
0.5	$O_W O_W$	2.775	3.525	2.2
0.6	$O_W O_W$	2.775	3.525	1.7
0.7	$O_W O_W$	2.775	3.525	1.3
0.8	$O_W O_W$	2.775	3.675	0.8
0.9	$O_W O_W$	2.775	3.525	0.4

Table 6.11: Extrema positions and coordination numbers for the 1^{st} peak in $g(r_{O_W O_W})$.

in the concentrated solutions water molecules are isolated or form dimers rather than cluster together. Corrobative evidence for this comes from the clustering analysis

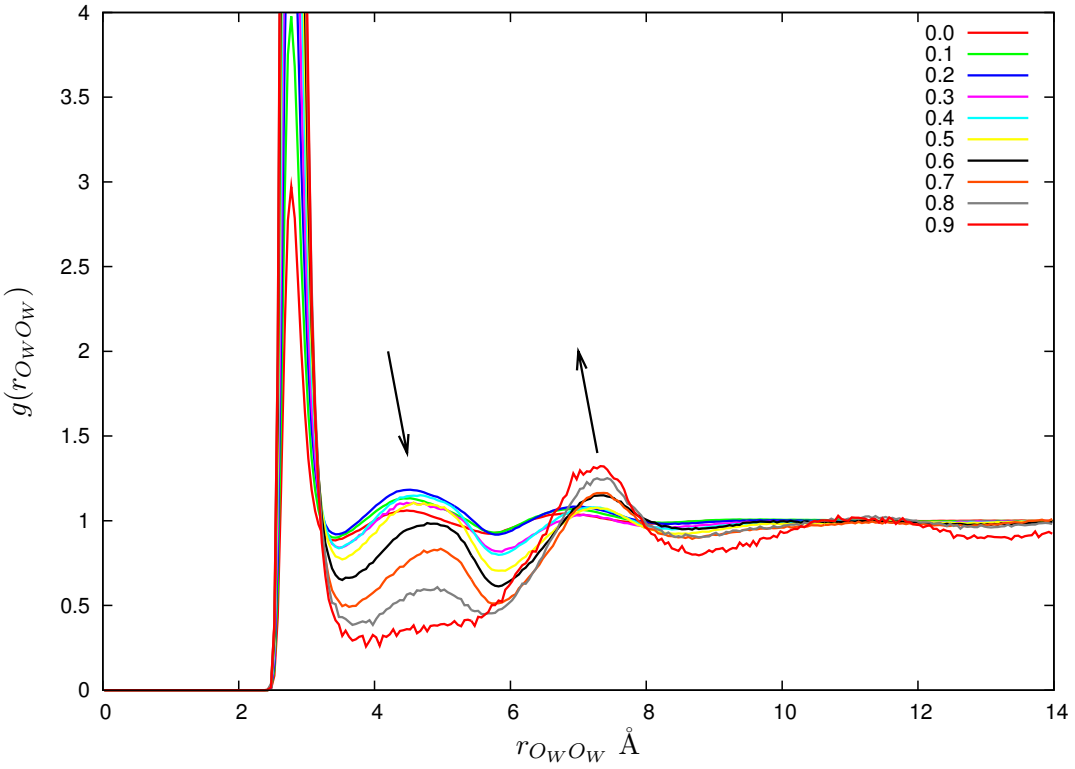

Figure 6.28: $g(r_{O_WO_W})$ for water and aqueous NMA mixtures. Arrows indicate the direction of changes with increasing NMA mole fraction.

presented in Section 6.6.2 which found that in the $X_{NMA} = 0.9$ solution approximately 90% of the water molecules are isolated or form half of a dimer. This still leaves 10% of water molecules associated in larger clusters although the very flat minimum in $g(r_{O_WH_W})$ at this concentration suggests that there is no tetrahedral order remaining. This idea will be investigated further in Section 6.6.4.

Significant proportions of water molecules are isolated in all the high concentration mixtures indicating that this may be a good model system in which to explore the structure and dynamics of water in peptide-rich environments such as those found in the interior of proteins[133]. As the 4.5Å $g(r_{O_WO_W})$ peak disappears with increasing NMA concentration a new peak forms at about 7Å, a slightly larger radial separation than the 3rd neighbour shell of pure water. This new peak arises from second neighbour contacts, however, in this case the bridging molecule is not water but NMA.

Table 6.11 lists the coordination numbers for the first peak of $g(r_{O_WO_W})$ showing the decrease in coordination from nearly five nearest neighbours in pure F3C water to only 0.4 at $X_{NMA} = 0.9$.

6.6.4 Tetrahedral order parameter, q

By calculating the tetrahedral order parameter q, defined in Section 3.4.4 it is possible to explore the orientational order of water in aqueous NMA solutions in more detail. The frequency distribution of q calculated for water oxygens from the positions of their

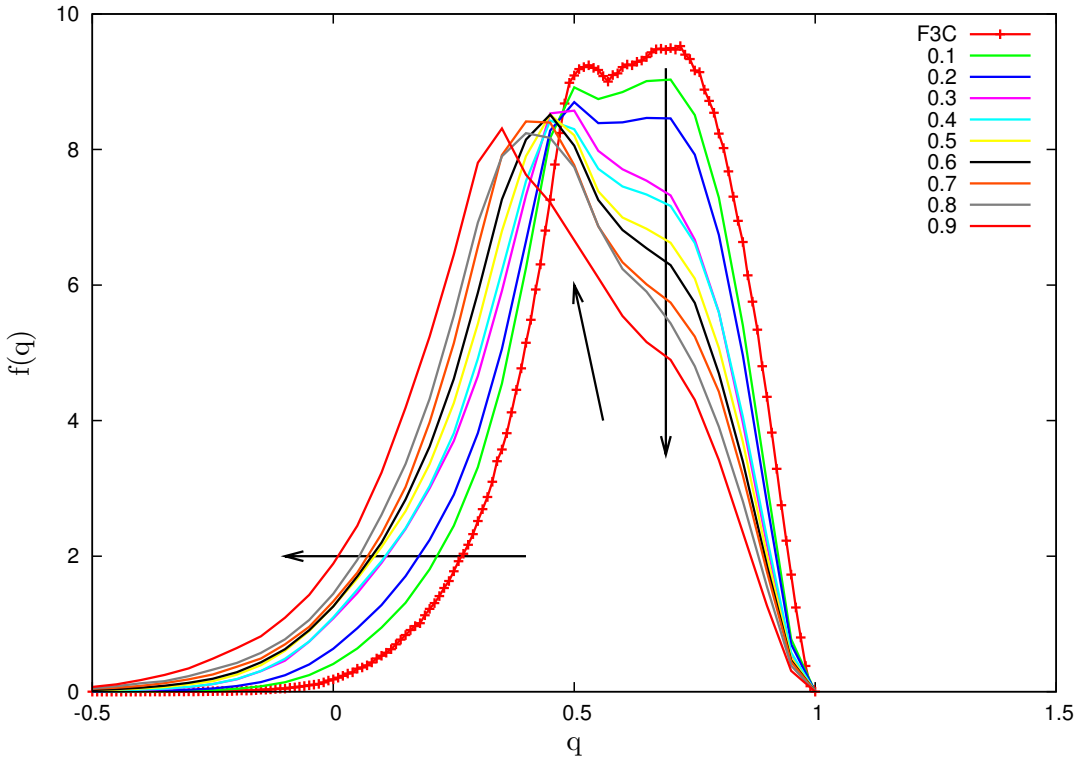

Figure 6.29: The distribution of the tetrahedral order parameter q, for water molecules calculated from the positions of their four nearest neighbours (either NMA or water). Arrows indicate changes with increasing NMA mole fraction. The bin width of the histograms used to collect frequency data was 0.05. The area under each curve sums to 100.

four nearest neighbours is shown in Figure 6.29. In this analysis "neighbours" are defined as donor or acceptor atoms of either species, specifically water oxygens, NMA oxygens or NMA nitrogens. The distribution for F3C water lies almost entirely in the range $q = 0$ to $q = 1$, the maximum of the distribution occurs at $q = 0.68$ although there is a smaller subsidiary peak at $q = 0.5$. Physically, these peaks correspond to two different configurations or groups of configurations of water molecules. Those contributing to the higher q peak have more tetrahedral order so can be thought of as relatively more structured than the set of configurations contributing to the lower peak at $q = 0.5$. On mixing with NMA the high q peak reduces in height and becomes a shoulder to the increasingly dominant "unstructured" peak. The location of this peak shifts to slightly lower q but remains approximately the same height indicating that configurations with this value of q are equally frequent at all concentrations. As the

NMA mole fraction increases there are an increasing number of random configurations ($q = 0$). The arrows in Figure 6.29 indicate the three main changes seen on adding an increasing proportion of NMA to water: a decrease in configurations with high tetrahedral order, the increasing dominance of the low q peak and a shift towards random configurations.

The average q of water molecules in F3C water is 0.61 and decreases to 0.43 in the $X_{NMA} = 0.9$ mixture. This decrease in $< q >$ is plotted in Figure 6.30, it is the

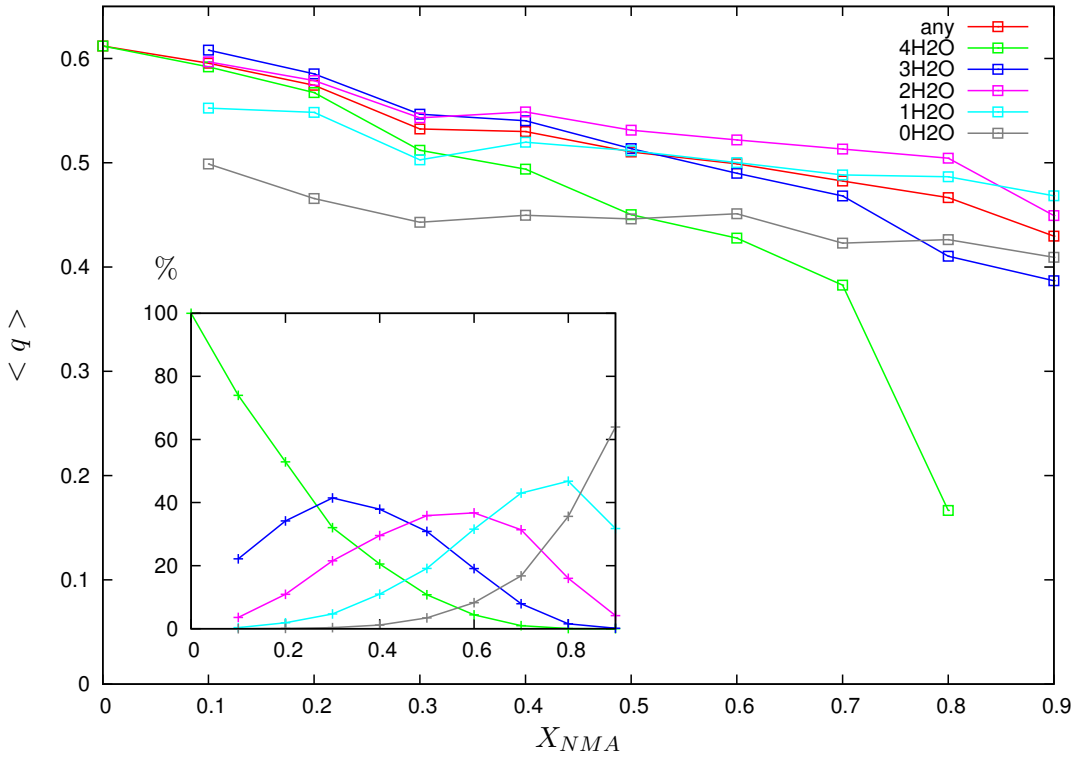

Figure 6.30: The mean value of the tetrahedral order parameter q for water molecules with specific neighbour combinations across the concentration range. red = neighbours of either species, green = all water, blue = 3 water + 1 NMA, pink = 2 water + 2 NMA, turquoise = 1 water + 3 NMA, grey = 4 NMA. The inset shows the population distribution of the five different neighbour combinations across the concentration range as the % of the total number of water molecules.

red line labelled "any" as q is calculated using the four nearest neighbours without distinguishing their species. By identifying the species of the four nearest neighbours it is possible to differentiate between specific configurations and calculate $< q >$ for populations of the five different nearest neighbour combinations: 1) four waters, 2) three waters and one NMA, 3) two waters, two NMA, 4) one water, three NMA and 5) four NMA molecules. Mean values of q are also plotted for these populations in Figure 6.30. The green curves show $< q >$ for waters with four water neighbours in

the main graph and the population of such waters as a percentage of the total number of waters in the inset. From the inset it is clear that the number of these "bulk" water molecules is depleted rapidly when NMA is added to the system. This is despite the results of Section 6.6.2 which show that water continues to form system size networks at concentrations up to and including $X_{NMA} = 0.4$. By $X_{NMA} = 0.4$ the population of water molecules with four water neighbours has dropped to 20%. The main graph shows that the local ordering around these "bulk" water molecules does not resemble that in liquid water. Instead $< q >$ for this population falls quite rapidly with increasing NMA concentration. At $X_{NMA} = 0.9$ none of the water molecules in the system have four nearest neighbours of the same species at any point during the simulation (as on average only a fraction of one percent of the molecules form clusters of five or more molecules).

$< q >$ for water molecules whose four nearest neighbours are all NMA molecules does not show nearly as much dependence upon concentration. At $X_{NMA} = 0.9$ the contribution from this population dominates $f(q)$ shown in Figure 6.29. As this distribution has quite a well defined peak this suggests that there is some favoured structure although from this analysis it does not appear to be particularly tetrahedral. In the NMA-rich regime a common structural motif was identified as an isolated water donating both hydrogens to NMA oxygens but accepting either none or only one hydrogen bond from an NMA amide group. This was illustrated in Figure 6.21. Therefore it is likely that the fourth NMA neighbour used to calculate q in this analysis is much further away from the water oxygen than the first three and probably not interacting very strongly with it. This would distort the calculated values of q towards the random case $q = 0$. From visualisation of the simulations the three NMA molecules associated with isolated waters appear to occupy three out of the four tetrahedral positions and a more appropriate choice of orientational order parameter could perhaps identify this. Restricting the analysis to include only water molecules which have four neighbours within a cutoff set equal to the location of the first minimum in $g(r_{O_W O_W})$ excludes configurations where the fourth neighbour is distant. As the NMA concentration increases, this reduces the number of molecules included in the analysis from 93% in F3C to 25% of the total possible number in $X_{NMA} = 0.9$. The result, not distinguishing between NMA and water neighbours, is shown in Figure 6.31. This clearly shows that, with the exception of the $X_{NMA} = 0.3$ simulation, waters with four closely associated neighbours are more tetrahedrally ordered in the aqueous NMA solutions than in pure water. Returning to Figure 6.30 it can be seen that there is evidence for this in the concentration dependence of $< q >$ for the different neighbour combinations. When $< q >$ is averaged over all neighbour configurations it remains higher at all concentrations than that when just "bulk" waters are considered.

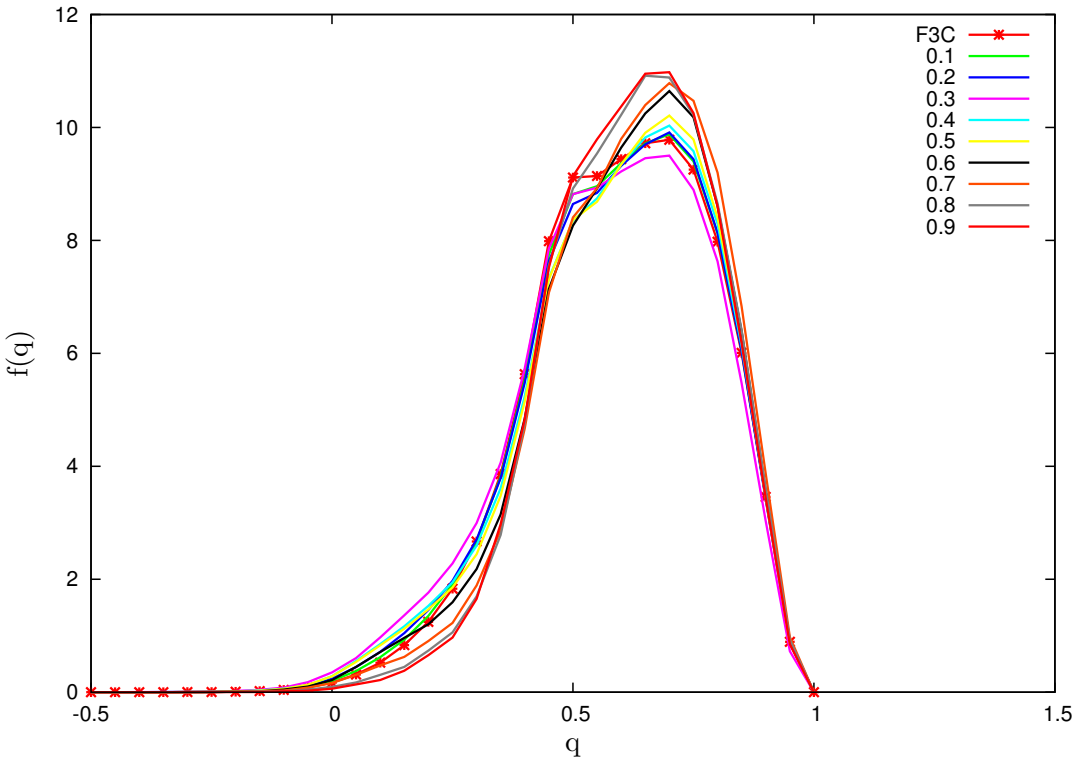

Figure 6.31: The distribution of the tetrahedral order parameter q, for water molecules calculated from the positions of their four nearest neighbours (either NMA or water) all of which are closer than 3.425Å.

Therefore the tetrahedral order must be maintained and even enhanced by water - NMA interactions. For almost the entire concentration range configurations with a combination of both water and NMA neighbours give a higher average $< q >$ than those where all four neighbours are water.

Continuing to restrict the analysis to close neighbours and plotting $f(q)$ for waters with four water neighbours in Figure 6.32(a) shows that as the NMA mole fraction increases the distribution flattens and indicates a shift towards more random configurations. Contrast this with Figure 6.32(b) for water molecules with two water and two NMA neighbours within the cutoff which shows the opposite trend. The high q peak becomes increasingly prominent and there is even a movement of the wing of the distribution away from random configurations. When no cutoff is applied and q is calculated the four nearest neighbours for all water molecules the resulting distributions no longer show the strong tendency towards more tetrahedral order. Nevertheless, $f(q)$ for four water neighbours shown in Figure 6.32(c) shows that the local order of these configurations is much more random than that of configurations in which water molecules have two water and two NMA neighbours. The plot of this distribution is shown in Figure 6.32(d) and

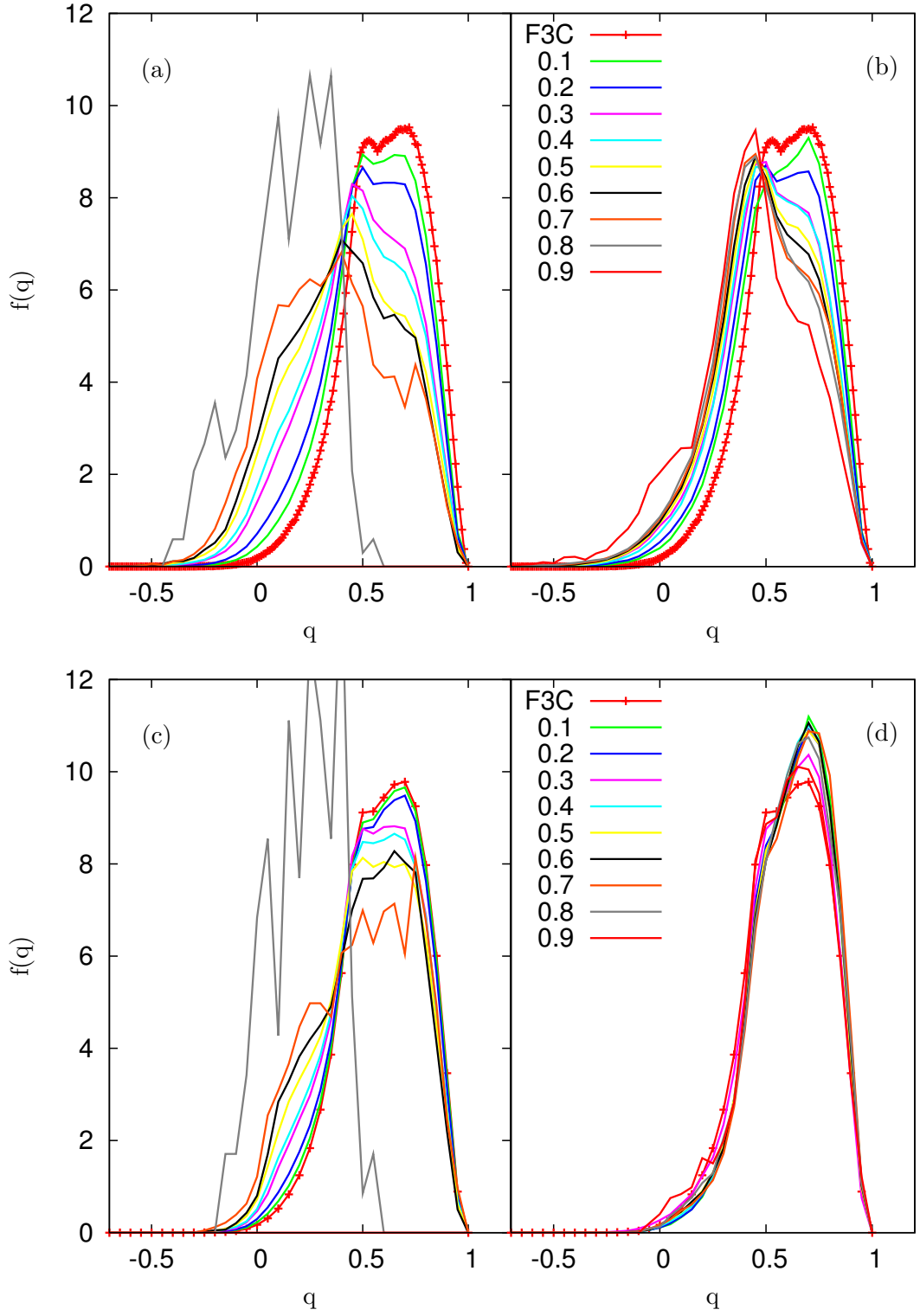

Figure 6.32: The distribution of the tetrahedral order parameter q, for (a) water molecules whose four nearest neighbours are all other waters and (b) water molecules with 2 water neighbours and 2 NMA neighbours. (c) and (d) show the same distributions as (a) and (b) respectively but with the analysis restricted to include only molecules whose four nearest neighbours are within 3.425Å. The frequency distributions in (a) and (c) for high NMA concentrations are very rough simply because the statistics are poor since very few water molecules have four water neighbours. None of the waters in the $X_{NMA} = 0.9$ mixture have four water neighbours at any point during the simulation.

it can be seen that the distribution remains narrow and peaked for all concentrations. A similar study of formamide-water mixtures also finds an enhancement of nearest neighbour structure among water molecules combined with loss of the global tetrahedral order as the concentration of formamide is increased[137].

6.6.5 Discussion of water structure in aqueous NMA solutions

As an increasing proportion of NMA is dissolved, system size water clusters give way to the formation of trimers, dimers and isolated molecules. However, even at low NMA concentrations when water molecules continue to form percolating networks it is clear that the topology of these networks is considerably different to that of pure water. Firstly, "bulk" water molecules surrounded by their own species fail to maintain the tetrahedral order of pure water. In fact, tetrahedral order is enhanced for water molecules with NMA neighbours. At concentrations between $X_{NMA} = 0.1$ and 0.2 water molecules make a transition from accepting two hydrogen bonds from other waters to accepting only one indicating a shift to predominantly linear structures. Below $X_{NMA} = 0.4$ water molecules continue to form system size networks with one donated and one accepted strong hydrogen bond formed with other waters instead of a four-coordinated tetrahedral structure. As the NMA concentration is increased the water molecules become increasingly isolated into short linear clusters. Above $X_{NMA} = 0.7$ the majority of waters are either completely isolated or form one half of a dimer. By hydrogen bonding to NMA molecules rather than to members of their own species water molecules can optimise the tetrahedral arrangement of their hydrogen bonds and this may explain the rapid disruption of the native intermolecular structure of water upon the addition of only a small amount of NMA.

6.7 Dynamics

The properties of hydrogen bonded molecular liquids include higher boiling points and slower dynamics than would be expected by only considering the size of their individual molecules. Both water and NMA fall into this category as do their mixtures in which both intra- and inter-species hydrogen bonding takes place. In Section 6.4.2 it was found that the average number of hydrogen bonds in the aqueous solutions is slightly higher than expected from an assumption of ideal mixing. The presence of excess hydrogen bonds might be expected to slow down the dynamics of the system and in this section the self-diffusion of both species will be studied to prove or disprove this hypothesis. Hydrogen bonds do not form permanently between molecules but are a transient feature, although key to the static intermolecular liquid structure. In this section the persistence of hydrogen bonded configurations in the aqueous mixtures

is also investigated. Naturally, self-diffusion and hydrogen bond lifetimes are closely related as in a system where hydrogen bonds persist for long times the movement of molecules is strongly impeded so self-diffusion will be slow and vice versa.

6.7.1 Self-diffusion

Due to their larger size NMA molecules self diffuse much more slowly than water molecules in their pure liquid phase so it is interesting to ask how the relative rates of self diffusion of the two species change on mixing. Figure 6.33 plots the mean

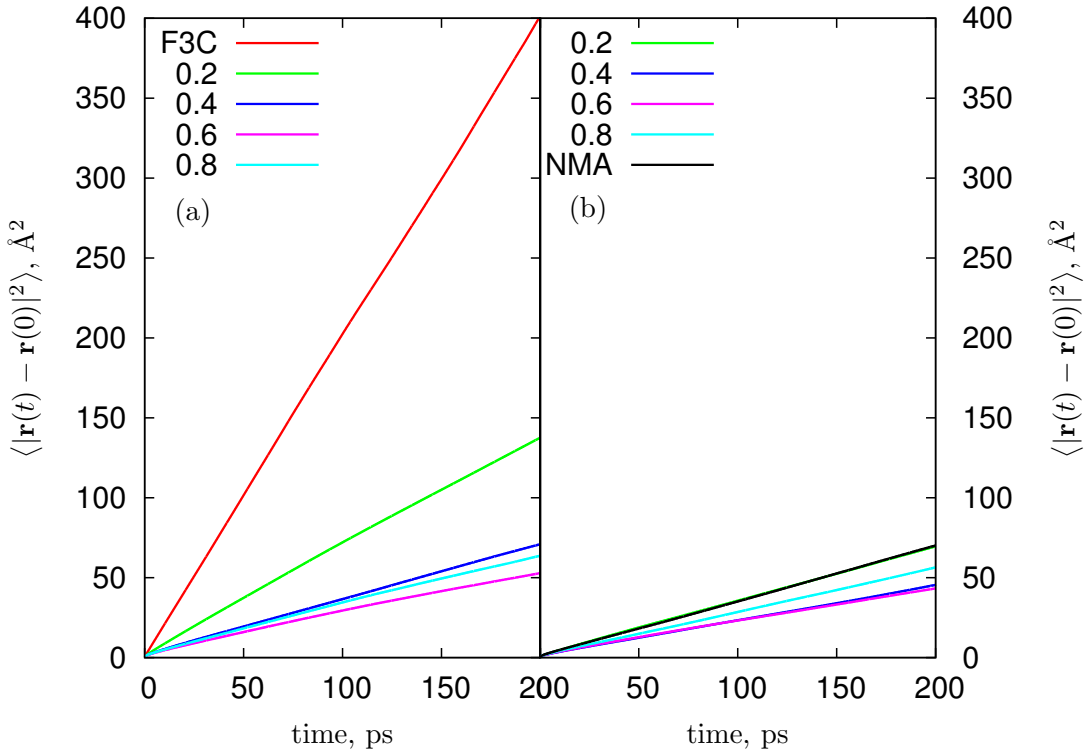

Figure 6.33: Centre of mass mean square displacement for (a) water and (b) NMA in a selection of aqueous NMA solutions.

square displacement of molecules in several of the simulations at differing concentrations from which it can be seen that there is an order of magnitude difference between the respective square distances travelled by water and NMA molecules in their pure liquids during the same time period. However, even the presence of a small proportion of NMA significantly impedes the motion of water as the mean square distance travelled in 200ps drops from 400 to less than 150Å^2 in $X_{NMA} = 0.2$ even though the mixture is numerically 80% water. The clustering analysis shows that water forms system size hydrogen bonded networks at concentrations up to and including $X_{NMA} = 0.4$ even though by this concentration the self diffusion coefficient has dropped to one-fifth of

its value in pure F3C water. Self diffusion coefficients for water and NMA in all the aqueous solutions are plotted in Figure 6.34. After the disappearance of system size

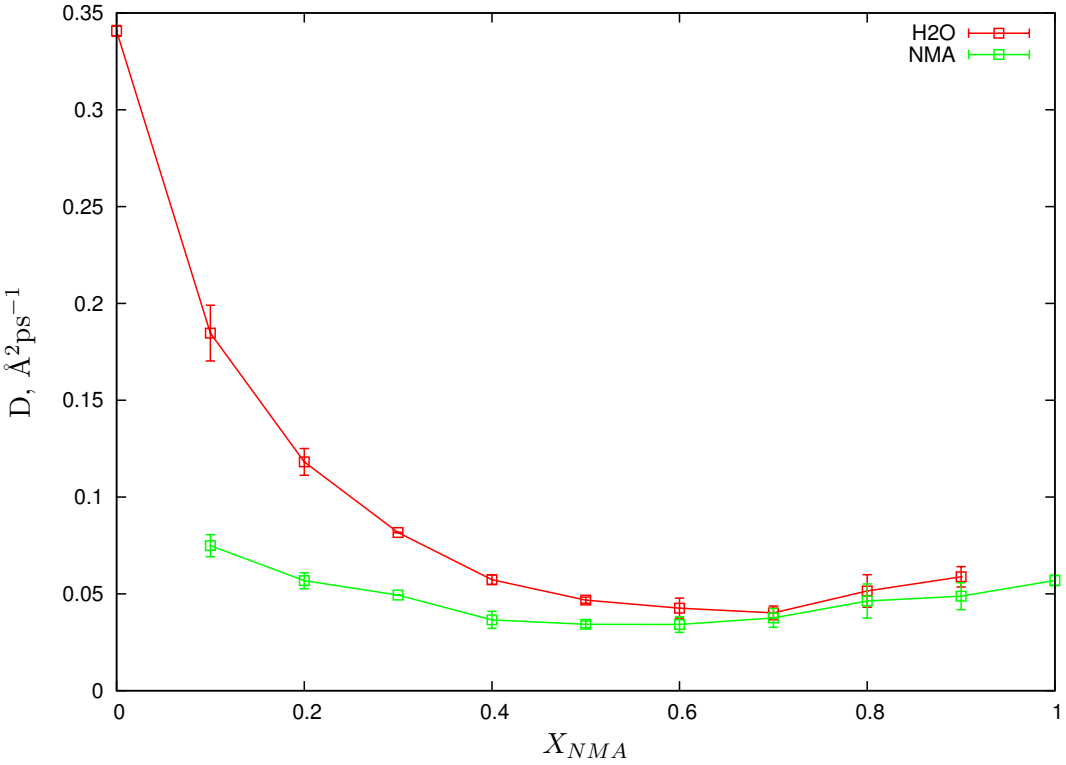

Figure 6.34: Self diffusion coefficients for water and NMA in aqueous NMA solutions. Each data point is the mean of a set of values calculated from the mean square displacements of four 250ps simulations. The error bars indicate the standard deviation of the set.

water clusters the rate of water self diffusion changes only slightly with increasing NMA mole fraction. At all concentrations water diffuses faster than NMA as one would expect since it is a much smaller molecule. However at concentrations greater than $X_{NMA} = 0.4$ the self diffusion coefficients of the two species are separated by no more than $0.02\text{Å}^2\text{ps}^{-1}$. The minimum in the self diffusion curves for both species at mid-range concentrations indicates that both water and NMA molecules find their motion in the binary mixture impeded by the presence of molecules of a different species. This minimum can be explained by considering the static hydrogen bonding analysis of Section 6.4 which found that the average number of hydrogen bonds per molecule in the system displayed a maximum at mid-range concentrations. A maxima in the number of hydrogen bonds corresponds to a maximal number of constraints upon the movement of individual molecules which can easily be equated to a minima in their self-diffusion as seen in Figure 6.34. At both extremes the self diffusion of the minority phase seems to track that of the prime constituent although in the dilute region the self-diffusion of NMA is much slower than that of water. A recent study of the

dynamical properties of formamide - water mixtures also found that the self diffusion of water drops upon the addition of the amide[137]. In this case, however, the self diffusion of formamide increases with increasing water mole fraction and the diffusion coefficients do not exhibit a minimum. This may be because formamide is smaller than NMA and is thus more mobile. Additionally, in formamide hydrogen atoms replace the NMA methyl groups and consequently the two hydrogen bond acceptor sites on the carbonyl oxygen are balanced by the two hydrogens bonded to the nitrogen which are both capable of donation. Therefore there is no excess of free acceptor sites and the topology of the formamide hydrogen bonded structure will be different to that of NMA, presumably more of a network than a chain-like structure. When water is added to NMA it links together neighbouring NMA chains creating a more connected network and consequently the system diffuses more slowly. This would not occur in formamide mixtures where water molecules simply substitute into the existing formamide network, as the topology does not change significantly it follows that self diffusion coefficients depend linearly on concentration as found by Elola and Ladanyi.

The next section investigates hydrogen bond dynamics: as these motions are stongly coupled to the diffusion it is to be expected that the hydrogen bond dynamics will be very fast in the dilute region where both water and NMA self diffusion is high but slow markedly by a concentration of about $X_{NMA} = 0.4$.

6.7.2 Hydrogen bond dynamics

There are four different types of hydrogen bonds in aqueous NMA solutions: in previous sections it was found that each type had a different mean length and hydrogen bond angle and Figures 6.35 and 6.36 show that there are clear differences in their lifetimes. Figure 6.35 shows the history independent autocorrelation of each different type of hydrogen bond according to Equation 3.25 given in Section 3.4.6. As the self-diffusion of the system slows with increasing NMA concentration it can be seen that hydrogen bonds become correlated over longer timescales. The results for the four different hydrogen bond types listed in Table 6.12 and plotted in Figure 6.36 show that the most persistant hydrogen bonded associations are those in which water molecules donate to the NMA carbonyl oxygen. These relaxation times do not represent true hydrogen bond lifetimes as no count is kept of how often bonds break and reform between the two times measured in the autocorrelation. However, they do give a quantitative measure of the timescale for which hydrogen bonded associations persist in the aqueous solutions. In order of decreasing duration, the water - (O=C)NMA hydrogen bonds are followed by water - water hydrogen bonds, then NMA - NMA and finally hydrogen bonds in which NMA molecules donate to the water oxygen. This is the same order as obtained by

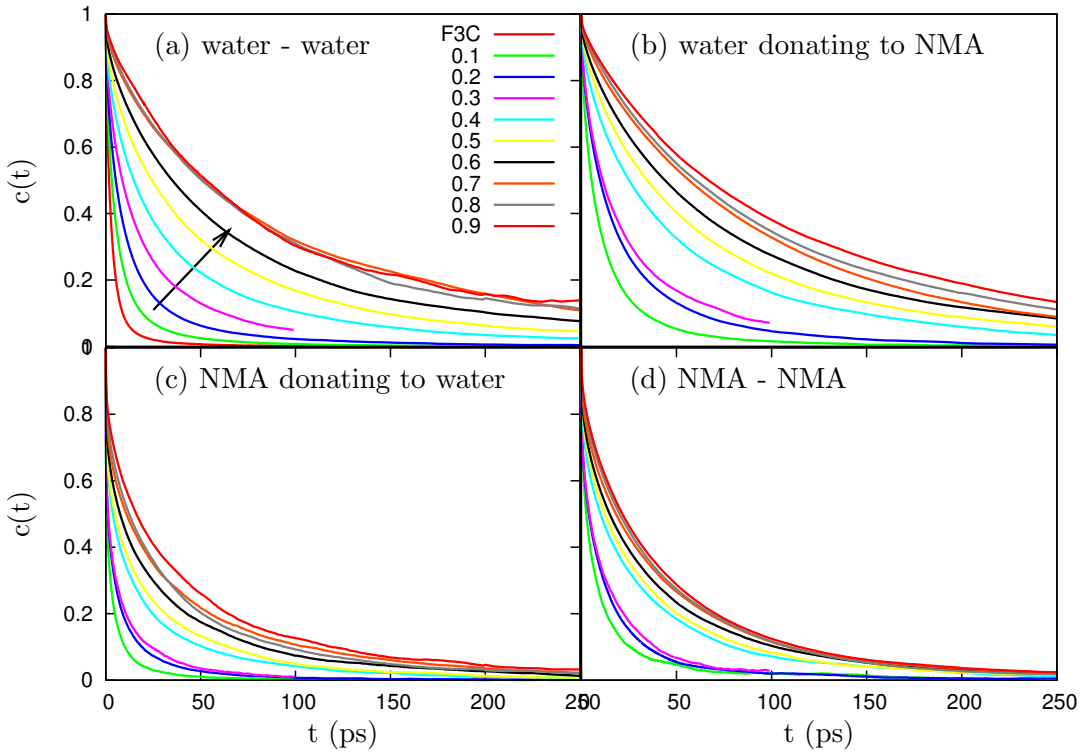

Figure 6.35: History independent autocorrelation functions for (a) water - water hydrogen bonds, (b) water donating to NMA, (c) NMA donating to water and (d) NMA - NMA. The arrow shows the direction of change with increasing NMA mole fraction. ($X_{NMA} = 0.3$ hydrogen bonds were only correlated over 100ps for a 250ps simulation. The other concentrations were correlated for time intervals up to 250ps using 1ns simulations.)

X_{NMA}	H_2O-H_2O	H_2O donating to NMA	NMA donating to H_2O	NMA-NMA
F3C	1.7	-	-	-
0.1	3.0	5.2	0.5	2.5
0.2	5.2	9.8	1.0	3.9
0.3	8.7	11.2	1.1	10.0
0.4	14.6	25.0	3.7	10.0
0.5	24.0	34.5	4.9	12.4
0.6	35.0	43.0	7.5	14.8
0.7	51.0	56.0	10.8	18.2
0.8	50.0	60.0	12.2	20.0
0.9	52.0	66.0	16.0	21.6
1.0	-	-	-	23.0

Table 6.12: Time taken (ps) for the history independent hydrogen bond autocorrelations plotted in Figure 6.35 to fall to half their initial value. These half lives are plotted as a function of concentration in Figure 6.36.

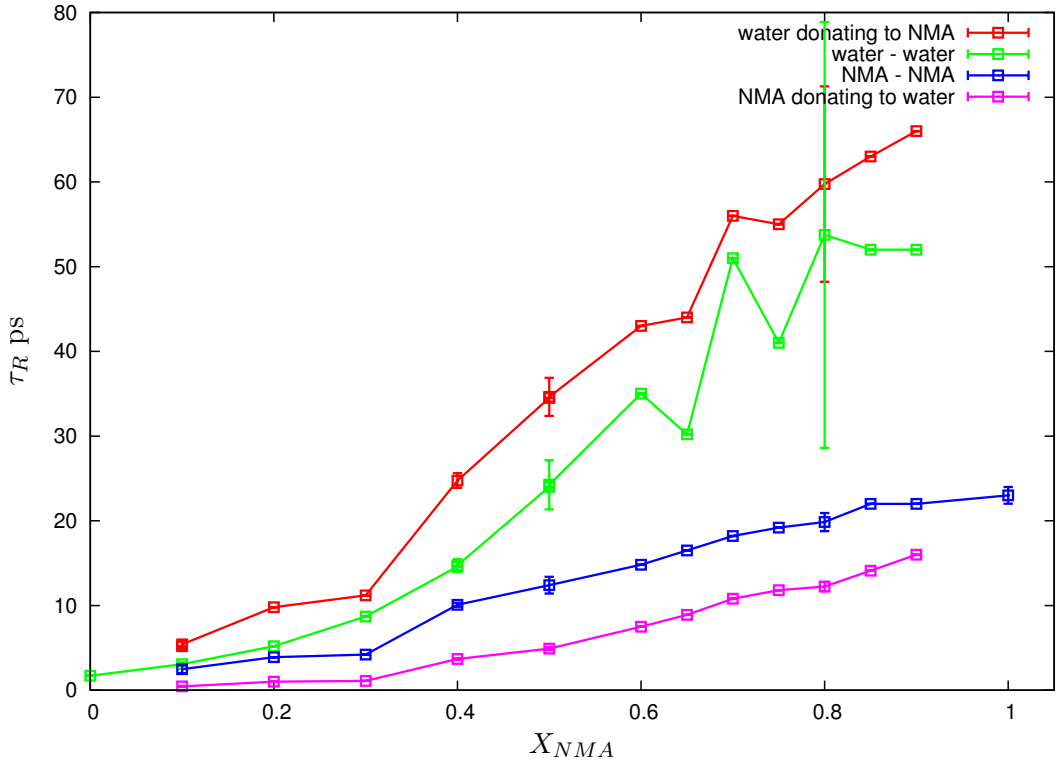

Figure 6.36: The time taken for history independent autocorrelation functions for intra- and inter-species hydrogen bonds in aqueous NMA solutions to drop to half their value at $t = 0$. Errors obtained by averaging the results of four 250ps simulations are shown for several of the concentrations. There are few water molecules in the high concentration mixtures but their hydrogen bonded associations are very persistent. Longer simulations would be required to accurately characterise water hydrogen bond timescales at this end of the concentration range.

considering the lengths of these four hydrogen bonds as defined by the location of the first peak in the relevant acceptor - hydrogen radial distribution function. The hydrogen bond donated by water to NMA is the shortest at 1.73Å, this is the same as the length of a water - water hydrogen bond in concentrated NMA but shorter than in pure F3C water where the hydrogen bond is 1.8Å. The NMA - NMA hydrogen bond is 1.9Åand the hydrogen bond donated by NMA to water is the longest at 2.0Å. There are very few studies of hydrogen bond energetics in aqueous NMA solutions and those which do exist focus on dimers or small clusters which cannot necessarily be expected to yield the same results as more sterically constrained larger systems such as those simulated in this work. Buck and Karplus determined the total pairwise interaction energies for CHARMM NMA and TIP3P dimers as well as for an NMA molecule complexed with two TIP3P molecules (one bound to each of the amide and carbonyl groups) as a function of the distance between the acceptor and hydrogen atoms (r) and the hydrogen bond angle. In order of decreasing strength, the calculated potential minima were -7.96 kcal/mol ($r = 1.85$Å) for the NMA-NMA dimer, -7.24 kcal/mol ($r = 1.80$Å) TIP3P

water donating to NMA, -6.64 kcal/mol ($r = 1.70$Å)TIP3P - TIP3P and -5.93 kcal/mol (1.90Å) NMA donating to TIP3P[78]. These results, which are in agreement with *ab initio* calculations[140], show that the relationship between hydrogen bond length and hydrogen bond energy is not straightforward. In the empirical CHARMM22 forcefield used in both this work and by Buck and Karplus the total interaction energy is the sum of van der Waals and electrostatic contributions. At short distances the van der Waals interaction is repulsive and this is the reason that the NMA - NMA hydrogen bond is calculated to be stronger than the TIP3P - TIP3P hydrogen bond, despite its greater length. Chapter 4 compares the TIP3P and F3C models and shows that there are differences between the two potentials. Similarly, Section 6.1 of this chapter compares the results of simulations of aqueous NMA using the F3C and the TIP3P model and shows that, as before, the two models do not give exactly the same results. Therefore, differences between this work and that of Buck and Karplus are likely to be due to the different model used for water as well as the difference in the system size.

In the previous section it was found that the self-diffusion of water fell rapidly as soon as even a small proportion of NMA was added. It is clear from the hydrogen bond dynamics that hydrogen bonds donated by water molecules to NMA are extremely persistent and this provides an explanation for the slow diffusion of water in the aqueous mixtures. From the static analysis it was concluded that there are a large number of free or partially hydrogen bonded carbonyl groups in liquid NMA and when water is added it preferentially hydrogen bonds to these groups rather than to its own oxygen atoms. This suggests the following picture of NMA hydration. In the pure liquid, NMA forms hydrogen bonded chains which diffuse relatively slowly due to both the size of the molecule and the intermolecular hydrogen bonds between molecules. When a small amount of water is added it preferentially donates hydrogens to the spare lone electron pairs on carbonyl oxygen atoms at the end of NMA chain and in mid chain sections where the NMA oxygen only forms a single hydrogen bond to another NMA. These water donated hydrogen bonds are strong and persistent so the water molecules become attached to the NMA chains for long time periods (tens of picoseconds) thus diffuse through the mixture at the same rate as the NMA chain itself. This explains why the self-diffusion coefficient of water drops to almost the same value as that of NMA in the concentrated mixtures.

The persistence of $\mathrm{H}_W \cdots O$ hydrogen bonds also explains the duration of water - water associations. At first there appears to be no reason why the timescale over which these associations persist should increase with NMA mole fraction, after all, the length of the water - water hydrogen bond is not significantly affected by an increase in peptide groups. In the NMA-rich region water molecules are dispersed throughout

the solution, even down to fairly dilute concentrations the intermolecular structure of water clusters is found to be chain-like rather than tetrahedral. Particularly at high concentrations almost all of the waters have at least one hydrogen bond to an NMA molecule, most likely a strong and persistent hydrogen bond to an NMA carbonyl oxygen. Consequently, the motion of the water molecule is impeded by its donation of at least one hydrogen bond to NMA. If a hydrogen bond forms between two water molecules whose motion is already restricted in this way then it is likely that the water - water hydrogen bond will last for as long as the two water molecules remain in the same vicinity, which is at least in part controlled by the persistence of the water - NMA hydrogen bonds. If peptide - water hydrogen bonds persist for tens of picoseconds even in simple molecular liquids such as aqueous NMA then it is hardly surprising that individual water molecules form an important structural element of larger, more rigid biomolecules.

The persistence of hydrogen bonded associations between NMA molecules appears to approach a plateau at high NMA concentrations. At these concentrations the self-diffusion of NMA and consequently water is presumably dominated by the lifetime of NMA - NMA hydrogen bonds. As water is added the average number of hydrogen bonds in the mixture increases as water molecules form hydrogen bonded bridges between NMA chains which further impedes movement thus slowing self-diffusion slightly. Both self-diffusion coefficients remain low until approximately $X_{NMA} = 0.4$ which is the first concentration where system size water clusters appear. At concentrations more dilute than $X_{NMA} = 0.4$ more than half the NMA molecules are isolated monomers so at these concentrations NMA - NMA interactions are much less important. Relative to NMA, water molecules self-diffuse rapidly due to their much smaller size. At dilute NMA concentrations the dynamics of the mixture will be dominated by those of water and from Figure 6.34 it can be seen that NMA monomers (75% of the molecules at $X_{NMA} = 0.1$) are in a sense carried along by the motion of water molecules in the aqueous mixture at a greater speed than in a liquid of their own species. Nonetheless, as it is larger NMA moves slower than water, and this has an impact on the water dynamics despite there being many more water than NMA molecules. In the dilute regime all the NMAs are hydrated, usually involving three hydrogen bonds, two the carbonyl oxygen and one donated by the amide hydrogen. While the hydrogen bond between the amide hydrogen and water is weak and less longlived than hydrogen bonds between waters themselves, the hydrogen bonds formed between water molecules and the carbonyl oxygen are twice as persistent as water - water hydrogen bonds in the dilute region. Therefore not only is the self-diffusion of water molecules sterically hindered by the presence of the much larger NMA molecules but there is a second impediment to movement in the formation of longlived hydrogen bonds.

6.8 Discussion

The static and dynamic results presented in this chapter for aqueous NMA solutions characterise the hydration process of this important model peptide or, conversely, investigate the disruption of the hydrogen bonded network of bulk water upon the addition of peptide groups.

In Chapter 5 the structure of liquid NMA was characterised as consisting of chains of molecules hydrogen bonded through the carbonyl and amide groups. The population distribution of these chains at 308K covers a large range of up to 100 molecules and is biased towards larger cluster sizes than expected for a distribution of linear chains formed by polycondensation reactions. Approximately one quarter of the carbonyl oxygen atoms in liquid NMA do not form hydrogen bonds and it is to these atoms that water molecules preferentially donate hydrogen bonds when they are added to NMA. The NMA chains are predominantly linear allowing water molecules to also donate to mid-chain carbonyl oxygens which only accept a single hydrogen bond from another NMA. Thus at high NMA concentrations the intermolecular structure of the pure liquid is not broken up but instead water molecules are dispersed through the mixture forming hydrogen bonded bridges between chains in agreement with the results of Czarnecki and Haufa[61]. The prevalence of this structural motif is seen in the increase of the low-r wing of $g(r_{OO})$ as the water mole fraction is increased in the aqueous solutions. The hydrogen bond formed when water donates to the NMA carbonyl oxygen is the second strongest[78], shortest and most long lived of the four types of hydrogen bonded interaction formed in the aqueous solutions. Therefore these bridge structures are prevalent, persistent, correlate the motions of neighbouring NMA chains and contribute to a slowing down of NMA self-diffusion. Due to inter-species hydrogen bonding water molecules are essentially attached to NMA chains in the NMA-rich region and consequently water self-diffusion mirrors that of NMA. It is only at $X_{NMA} = 0.4$ that system size water clusters appear and the self-diffusion of water begins to increase.

In pure water almost all the molecules are continuously connected to a system size hydrogen bonded network. This is not to imply that the network is static, on the contrary, a typical hydrogen bonded association between a specific pair of molecules persists for only a few picoseconds. Water molecules diffuse at $0.3\text{Å}^2\text{ps}^{-1}$ in the simulations presented in Chapter 4 continually swapping hydrogen bonding partners while maintaining an approximately tetrahedral arrangement of three or four hydrogen bonds around their oxygen atom. Water molecules prefer to donate their hydrogens to NMA molecules with the result that even at $X_{NMA} = 0.2$ there are more water molecules which only accept a single hydrogen bond from another water than accept

two. By $X_{NMA} = 0.3$ the same is true for water donors and consequently the system size water clusters at $X_{NMA} = 0.3$ and 0.4 are more chain-like than tetrahedral. Evidently the addition of a small amount of NMA has a significant impact on the intermolecular water structure and this is also apparent in the dynamics where water self-diffusion is nearly halved by the addition of 10% NMA. Not only is the movement of water sterically constrained by the presence of the much larger NMA molecules but hydrogen bonded associations between water and the NMA carbonyl group are persistent even at dilute concentrations. Hydrogen bonds donated by NMA to water are much less important as water prefers to accept hydrogen bonds from other waters resulting in a larger proportion of free amide groups in the aqueous solutions than are present in liquid NMA.

The hydrogen bonding interaction between the amide and carbonyl groups of the peptide bond plays the dominant role in determining not only the intermolecular structure of aqueous NMA but also the most prevalent secondary structure elements found in proteins: alpha helices and beta sheets. The aqueous NMA system is therefore a minimal model for the hydration of these protein backbone structures and characteristic features of this system such as the propensity of water molecules to donate rather than accept hydrogen bonds to peptide groups and the formation of water bridge structures between NMA chains might be expected to feature in protein - water interactions. Surveys of the bound water molecules in crystallographic protein structures confirm their tendency to donate rather than accept hydrogen bonds[141] and also reveal water bridges connecting adjacent peptide backbones and even mediating in backbone turns[142]. The formation of water bridges in the NMA-rich region of aqueous NMA solutions hinders the relative motion of neighbouring chains and contributes to a slowing down of the system dynamics. Water might be expected to have a similar effect upon both the intra- and inter-molecular dynamics of larger peptides and biological macromolecules.

Chapter 7

Quasielastic Neutron Scattering

Computer simulation is a useful method which gives access to atomic level information on molecular systems. However, any computational technique is only as good as its input data. Classical molecular mechanics forcefields such as those used to simulate water and NMA in the previous chapters are fitted to experimental data or *ab initio* computational results. Such empirical forcefields are therefore tuned to reproduce certain experimental results. However, the quality of the computer model can be tested by its agreement with experiment for properties to which it is not tuned. Disagreement with experimental results may not always reveal inadequacies of the simulation model, on the contrary, in some cases discrepancies have led to reinterpretation of experimental data[143]. Clearly, computer simulation and experiment are complementary techniques which can be combined to yield a level of understanding greater than the sum of its parts.

Neutron scattering is a powerful tool which can be used to explore the structure and dynamics of condensed matter on length scales ranging from the atomic to the macromolecular. It is particularly appropriate to combine neutron experiments with classical molecular dynamics simulations as the two techniques explore similar length and time scales. This chapter gives a brief introduction to quasielastic neutron scattering (QENS) which is a technique used to probe the stochastic dynamics of atoms and molecules. The experimental procedures and data analysis methods are outlined followed by the results of QENS experiments on liquid NMA.

7.1 Theory

In neutron scattering experiments a beam of neutrons is scattered by a sample placed in its path. By measuring the scattering angles and energies of the scattered neutrons information on the positions and dynamics of the nuclei in the sample can be obtained.

7.1.1 Why use neutrons?

Since neutrons are uncharged they experience no Coulomb interaction. Therefore, in order for a neutron to interact with matter it must approach atomic nuclei closely enough to be within the range of the strong nuclear force. This means that neutron scattering has several advantages over X-ray techniques where scattering is caused by interaction with electrons.

- light atoms (e.g. hydrogen) are easily detected despite the presence of heavier nuclei.
- neighbouring elements in the periodic table or isotopes of the same element have quite different scattering cross-sections and so are readily distinguished.
- unlike X-rays, neutrons interact weakly with matter and therefore are highly penetrating but non-destructive probes.
- neutrons are spin-$\frac{1}{2}$ particles so have a magnetic moment and because of this can be used to investigate magnetisation on the atomic scale.

The energies of thermal neutrons vary from μeV to eV which is the same energy range as that of many atomic and electronic processes including translational, vibrational and rotational modes. A neutron can be considered as either an electrically neutral particle with mass m_n, or as a plane wave with wavelength λ and wave vector $\mathbf{k}$ ($|\mathbf{k}| = \frac{2\pi}{\lambda}$). Such a plane wave will have momentum, $\mathbf{p} = \hbar\mathbf{k}$ and energy, $E = \frac{\hbar^2|\mathbf{k}|}{2m_n}$. Neutrons are produced at ISIS (the neutron facility at the Rutherford Appleton Laboratory, Didcot, UK) with wavelengths in the range 0.05 to 20Å . Experiments can therefore be performed across a range of lengthscales from probes of the hydrogen atom to studies of macromolecules.

7.1.2 Elastic and inelastic scattering

In any scattering experiment there are two parameters which quantify the difference between the initial and final states of the scattered particle:

- the energy transfer $\hbar\omega$
- the momentum transfer $\hbar\mathbf{Q}$.

Measuring the scattering angle between the direction of the scattered neutron and the incident beam is equivalent to measuring the momentum transfer $\hbar\mathbf{Q}$ since the two quantities are related by

$$\hbar Q = \frac{4\pi\hbar}{\lambda} sin\left(\frac{\theta}{2}\right) \tag{7.1}$$

where Q is the magnitude of the scattering vector $\mathbf{Q} = \mathbf{k}_{final} - \mathbf{k}_{initial}$ and λ is the neutron wavelength. If the neutron does not exchange energy with the sample then the scattering is elastic and $\hbar\mathbf{Q}$ is the only useful parameter. When the scattering is inelastic it then becomes interesting to measure the energy transfer which can be done by calculating the energy of the incident and scattered neutrons from measurements of either their wavelength or velocity. Energy transfer is defined as $\hbar\omega = E_{initial} - E_{final}$ so that interactions in which the measured energy of the scattered neutron is less than that of the incident neutrons correspond to positive energy transfers.

7.1.3 Experimental cross-sections

The general aim of a neutron scattering experiment is to measure the number of scattered neutrons as a function of $\mathbf{Q}$ and ω. The actual experimental measurement is of the effective cross-section presented by an atomic nucleus to an incoming neutron defined by

$$\sigma = \frac{I_{scatter}}{I_{incident}} \tag{7.2}$$

where $I_{scatter}$ is the number of scattering events occurring per second and $I_{incident}$ is the number of incident neutrons per second per unit area. By placing a number of detectors around the target the differential cross section $\frac{\partial\sigma}{\partial\Omega}$ is measured which gives the probability that a neutron leaves the sample in the solid angle element $d\Omega$ about the direction $\mathbf{\Omega}$. If in addition the final energy of each neutron hitting the detectors is determined then the result is the double differential cross section

$$\frac{\partial^2\sigma}{\partial\mathbf{\Omega_f}\partial E_f} \tag{7.3}$$

which gives the number of neutrons scattered per unit time into a small solid angle $d\Omega$ in the direction θ, ϕ, with final energy between E_f and $E_f + dE_f$.

The formalism relating the experimentally measured cross-sections to the structure and dynamics of the scattering system was developed by Van Hove[144] and can be found in the relevant textbooks[145][146]. Put very simply, experimental measurements are made of the intensity of the scattered neutron beam which is proportional to a function known as the dynamic structure factor $S(\mathbf{Q},\omega)$. A connection needs to be made from momentum and energy to real-space and time. This is done by defining the Intermediate function, $I(\mathbf{Q},t)$. $S(\mathbf{Q},\omega)$ is then the Fourier transform in time of $I(\mathbf{Q},t)$. By Fourier transforming $I(\mathbf{Q},t)$ in space one obtains $G(\mathbf{r},t)$, the time-dependent pair-correlation function of the scattering system. $G(\mathbf{r},0) = \delta(\mathbf{r}) + g(\mathbf{r})$ where $g(\mathbf{r})$ is the static pair distribution function which for isotropic systems gives the radial distribution function seen in previous chapters. The theory is general and applies to any system - solid, liquid

or gas - where the scattering can be divided into coherent and incoherent contributions.

7.1.4 Coherent and incoherent scattering

The nuclear scattering length, b, quantifies the degree of scattering of an incoming neutron by an atomic nucleus. The scattering length of a nucleus depends on its spin state and as most isotopes have several spin states, even samples composed of a single isotope will contain nuclei with a variety of scattering lengths. The Van Hove formalism qualitatively described in the previous section includes a sum over pairs of nuclei i and j. The product of the scattering lengths of these pairs can be averaged over the possible nuclear spin states to give two separate contributions to the scattering law:

$$\sum_{i,j} \overline{b_i b_j} = \sum_{i,j} (\bar{b})^2 + \sum_i \left(\overline{b^2} - (\bar{b})^2 \right) \tag{7.4}$$

The first term represents coherent scattering. It involves a sum over both i and j and therefore depends on correlations between different nuclei. This gives rise to interference effects from which the sample structure can be deduced. The second term gives the incoherent scattering contribution. This involves correlations between the same nuclei i at different times. Consequently, incoherent scattering holds information on the motion of individual particles.

7.1.5 Hydrogen

The coherent and incoherent cross-sections of different elements can be measured experimentally and it has been found that the incoherent cross-section of hydrogen is very much larger than its coherent cross-section. In fact, its incoherent cross-section is far larger than that of any other element. Consequently neutron scattering of hydrogenous substances is dominated by the incoherent scattering of the hydrogen atoms. Experimentally this is very useful as to a good approximation substances which contain a large proportion of hydrogen (for example, many biological materials) can be considered as single component, incoherent scattering systems. The total scattering cross section of a hydrogen atom is 82.02 barns where 1 barn $= 1 \times 10^{-24}\text{cm}^2$[147]. Nearly 98% (80.26 barns) of this is due to incoherent scattering. The scattering cross sections of other elements and even the hydrogen isotope deuterium are much smaller.

7.1.6 Quasielastic neutron scattering

The measured scattered intensity is proportional to the dynamic structure factor $S(\mathbf{Q}, \omega)$ which contains information on the structure and dynamics of the scattering system. In the case of hydrogenous samples such as NMA the coherent contribution is negligible as can be seen from Table 7.1 and so it is the incoherent part of $S(\mathbf{Q}, \omega)$

atom	atomic cross sections			N	NMA total cross sections		
	coh	inc	total scatt		coh	inc	total scatt
H	1.7568	80.26	82.02	7	12.2976	561.82	574.14
C	5.551	0.001	5.552	3	16.653	0.003	16.656
N	11.01	0.5	11.51	1	11.01	0.5	11.51
O	4.232	0.0008	4.232	1	4.232	0.0008	4.232
total					44.1926	562.3238	606.538
%					7.3%	92.7%	100%

Table 7.1: Coherent, incoherent and total scattering cross sections for thermal neutrons from data provided by the NIST Center for Neutron Research[147]. For NMA incoherent scattering accounts for more than 90% of the total scattering cross section. 99.9% of the incoherent scattering from NMA is due to scattering from its seven hydrogen nuclei.

which is measured, corresponding to the motion of individual protons. In a quasielastic experiment small energy transfers are measured which correspond to the energies of the translational, rotational and vibrational motions of the atoms. Since these three motions have very different time scales their contributions can be decoupled and written as the product of the respective intermediate scattering functions.

$$I(\mathbf{Q},t) = I^{vibr}(\mathbf{Q},t) \times I^{rot}(\mathbf{Q},t) \times I^{trans}(\mathbf{Q},t) \tag{7.5}$$

Upon Fourier transforming the multiplication becomes a convolution in energy space.

$$S(\mathbf{Q},\omega) = S^{vibr}(\mathbf{Q},\omega) \bigotimes S^{rot}(\mathbf{Q},\omega) \bigotimes S^{trans}(\mathbf{Q},\omega) \tag{7.6}$$

The IRIS spectrometer observes motions on timescales between 1 and 75ps which are much longer than the timescale of vibrational motions of the protons. Therefore the vibrational contribution reduces to the Debye-Waller factor $e^{-Q^2\langle r^2\rangle/3}$, where $\langle r^2\rangle$ is the mean square displacement of nuclei due to thermal vibrations.

The incoherent structure factor consists of an elastic and a quasielastic part

$$S_{inc}(\mathbf{Q},\omega) = e^{-Q^2\langle r^2\rangle/3}\left[A_\circ(\mathbf{Q})\delta(\omega) + S^{qel}_{inc}(\mathbf{Q},\omega)\right] \tag{7.7}$$

$A_\circ(\mathbf{Q})$ is known as the Elastic Incoherent Structure Factor (EISF) and represents the space-Fourier transform of the final distribution of the scattering particles. As before the structure factor is attenuated by the Debye-Waller factor. In the simplest case, the intermediate scattering function $I(\mathbf{Q},t)$ decays exponentially with time. As the Fourier transform of an exponential is a Lorentzian, in this case the quasielastic component can

be described by a single Lorentzian function $L(\omega)$

$$S_{inc}^{qel}(\mathbf{Q},\omega) = [1 - A_\circ(\mathbf{Q})]\, L(\omega) = [1 - A_\circ(\mathbf{Q})]\, \frac{1}{\pi} \frac{\Gamma}{\Gamma^2 + \omega^2} \tag{7.8}$$

where the half width at half maximum of the Lorentzian, Γ, is related to the characteristic time of the motion.

7.1.7 Liquid NMA: a jump model for diffusion

The random jump diffusion model involves two processes: oscillatory motion at a specific site of duration τ, followed by an instantaneous jump to a neighbouring site. $\Gamma(Q)$ in this model is given by

$$\Gamma(Q) = \hbar \frac{DQ^2}{1 + DQ^2\tau} \tag{7.9}$$

where D is the translational self-diffusion coefficient. In the limit as Q tends to zero, corresponding to large distances

$$\Gamma(Q) \sim \hbar DQ^2 \tag{7.10}$$

which is the form expected if the structure factor is derived by assuming that normal Fickian diffusion is taking place[145]. In the limit as $Q \sim \infty$

$$\Gamma(Q) \sim \frac{\hbar}{\tau} \tag{7.11}$$

Therefore this model predicts a crossover in behaviour. At small Q, corresponding to long distances, the details of the jump process are irrelevant and the broadening of the Lorentzian is linear in Q^2 as predicted by continous Fickian diffusion. The crossover to a constant at high Q indicates the granularity of the motion on the atomic length scale. In the long time limit self-diffusion coefficients can be related to the mean square displacement of the particles via the Einstein relation

$$D = \lim_{t\to\infty} \frac{1}{6t} \left\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \right\rangle \tag{7.12}$$

and using this relation a characteristic length scale l corresponding to the jump distance can be defined as

$$l = \sqrt{6D\tau} \tag{7.13}$$

7.2 Experimental method

The quasielastic neutron scattering (QENS) experiments were carried out on IRIS, an instrument at the ISIS facility designed for quasielastic and low-energy, high resolution inelastic spectroscopy. It is a time-of-flight, indirect geometry, crystal analyser spectrometer[148]. In fact, IRIS consists of two spectrometers. The "primary" spectrometer views a liquid hydrogen moderator which is cooled to 25K and thus accesses a flux of long-wavelength (cold) neutrons. Two disc choppers select the range of neutron energies incident upon the sample. Next a supermirror focuses the beam at the sample position. The "secondary" spectrometer is a vacuum vessel, 2m in diameter, containing two crystal analyser arrays, two scintillator detector banks and a diffraction detector bank at 2θ =170^o which is used for diffraction experiments. Only scattered neutrons with a wavelength of 6.66Å are directed towards the detectors by the crystal array. The wavelength is related to the energy of the neutron by

$$E = \frac{1}{2}mv^2 \ , \ p = mv = \frac{h}{\lambda} \tag{7.14}$$

Thus the energy of the scattered neutron can be calculated from its wavelength. Since the exact distance between the sample and the detector bank, L_2 is known, the time taken t_2, for a neutron to travel a distance L_2, can be calculated. In addition the total time-of-flight, t, (from production to detection of the neutrons) and the primary flight path L (distance from production site to sample) are measured so that the energy exchange within the sample can be calculated using

$$\hbar\omega = \Delta E = \frac{1}{2}m\left[\left(\frac{L_2}{t_2}\right)^2 - \left(\frac{L}{t-t_2}\right)^2\right] \tag{7.15}$$

The detectors count the number of scattered neutrons over a succession of brief time periods. Therefore the basic result is a histogram of number of neutrons detected per unit time (usually neutrons/μs) versus time. A simple unit conversion from time to energy can then be performed using Equation 7.15. The instrument resolution gives the accuracy with which energy transfers can be measured. The full width at half maximum (FWHM) of IRIS's resolution function at the elastic line is 17.5μeV. IRIS can observe motions over timescales between 1 and 75ps and over a momentum transfer range corresponding to 0.3 - 1.8Å^{-1} using the PG002 pyrolytic graphite analyser which was used in these experiments. IRIS is also equipped with several refrigerators, cryostats and heaters enabling the investigation of samples at temperatures ranging from 15-673K.

7.2.1 Sample preparation

Pure NMA was bought from Aldrich and injected into a cylindrical aluminium sample can. NMA is a crystalline solid at room temperature but was easily melted by placing in a warm cupboard for a few minutes. For the aqueous NMA experiment a 0.2 mole fraction dNMA solution was prepared using deuterated NMA and distilled water. This mixture is liquid at room temperature. The sample thickness was approximately 80μm which allows more than 90% transmission of the neutron beam thus avoiding the neccesity of correcting the results for the effects of multiple scattering.

7.2.2 Calibration

Runs with the empty sample can at 327K and with vanadium foil at 300K were performed. Sample cans are made from aluminium since its incoherent cross-section is negligible when using long wavelength neutrons. The empty can run is used to correct the results for background radiation including contributions from the can itself. Vanadium is an almost purely incoherent scatterer and is used to calibrate the angular response and energy resolution of the instrument.

7.2.3 Experimental runs

Two sets of experimental runs were performed. First a set of short runs (data collection times of about 10 minutes) spanning a temperature range of 18-370K were carried out. The results of this elastic window scan give an overview of the motional changes in NMA in its solid and liquid phases and directed the choice of specific temperatures for longer runs (3 or 4 hours) in the temperature range of methyl rotations (20, 60, 90,120, 140 and 160K) and in the liquid phase (306, 310, 325, 340, 355 and 370K). Only the results for liquid NMA are presented in this thesis. The 0.2:0.8 dNMA:H_2O sample was investigated at 250, 272, 293 and 308K. Data collection times were approximately 4 hours per temperature.

7.3 Data Analysis

Standard ISIS in-house software packages (OpenGenie and MODES) were used to carry out the data analysis. The spectra from each detector in the short runs were summed to give a total intensity. The IONIAN program in the MODES package[149] was used to do this and to convert from time-of-flight to energy transfer including corrections for background and variations in detector efficiency. The corrected counts were then summed over the energy transfer window $\pm 0.003\mu$eV. The set of longer runs were used to obtain the elastic line intensity as a function of momentum transfer, **Q**. In this experiment the spectra measured by the detectors were grouped in threes to

points of approximately equal momentum transfer. This was done so as to obtain better statistics. Since there were 51 detectors this grouping resulted in 17 groups corresponding to 17 values of $\mathbf{Q}$.

The count measured by the detectors $I(\mathbf{Q},\omega)$ is proportional to the convolution of the structure factor $S(\mathbf{Q},\omega)$ with the resolution function $R(\mathbf{Q},\omega)$ of the IRIS spectrometer. The MODES package uses Bayesian analysis to determine the number and widths of Lorentzian peaks necessary to fit the quasielastic scattering spectra. The model comprises a δ-function with amplitude $A(0)$ to fit the elastic peak, together with a maximum of three Lorentzians and a choice of sloping, flat or zero background intensity. For the NMA analysis the best fits were found using an elastic peak and a flat background together with a single Lorentzian. The background accounts for a weak inelastic contribution in the quasielastic region of the spectra. For temperatures between 306 and 340K the widths increase as a function of Q but at low Q values the width of the single Lorentzian was within or close to the instrument resolution therefore these values were excluded from further analysis.

7.4 Liquid diffusion in NMA

Above the melting point the widths of the fitted Lorentzians tend to zero at low Q and to a constant at high Q as illustrated in Figure 7.1. The linear section at longer length scales (corresponding to low Q) indicates that the molecules are undergoing translational diffusion while the crossover to a constant at high Q shows the granularity of the motion on the atomic length scale. This functional form is well described by the jump diffusion model outlined in Section 7.1.7 in which the line width of the Lorentzian fitted to the QENS spectra depends on Q^2 as

$$\Gamma(Q) = \hbar\frac{DQ^2}{1+DQ^2\tau} \tag{7.16}$$

where D is the translational self diffusion coefficient and the residence time τ is a measure of how long atoms spend at one position before jumping to a neighbouring site. The values of D and τ obtained by fitting the above equation to the Q dependent Lorentzian line widths obtained from fits to the scattering spectra are listed in Table 7.2. However, these self diffusion coefficients are twice or three times higher than those measured by pulsed field gradient nuclear magnetic resonance (PFG-NMR) by Chen, Groß and Ludemann[131] and calculated from the molecular dynamics simulations presented in Chapter 5, as can be seen from Table 7.2. PFG-NMR is a standard method used to accurately determine macroscopic self-diffusion coefficients therefore it is necessary to find an explanation for why the QENS spectra do not yield similar

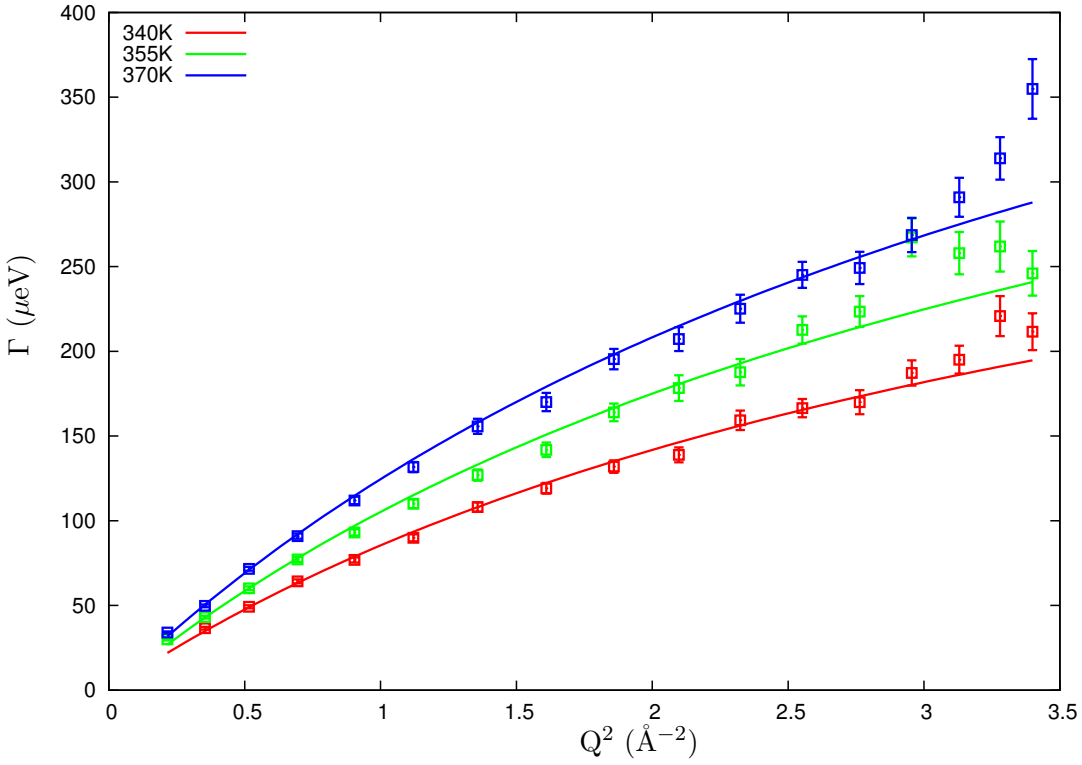

Figure 7.1: Half width at half maximum (Γ) of the fitted Lorentzians at 340, 355 and 370K plotted as a function of $\mathbf{Q}^2$.

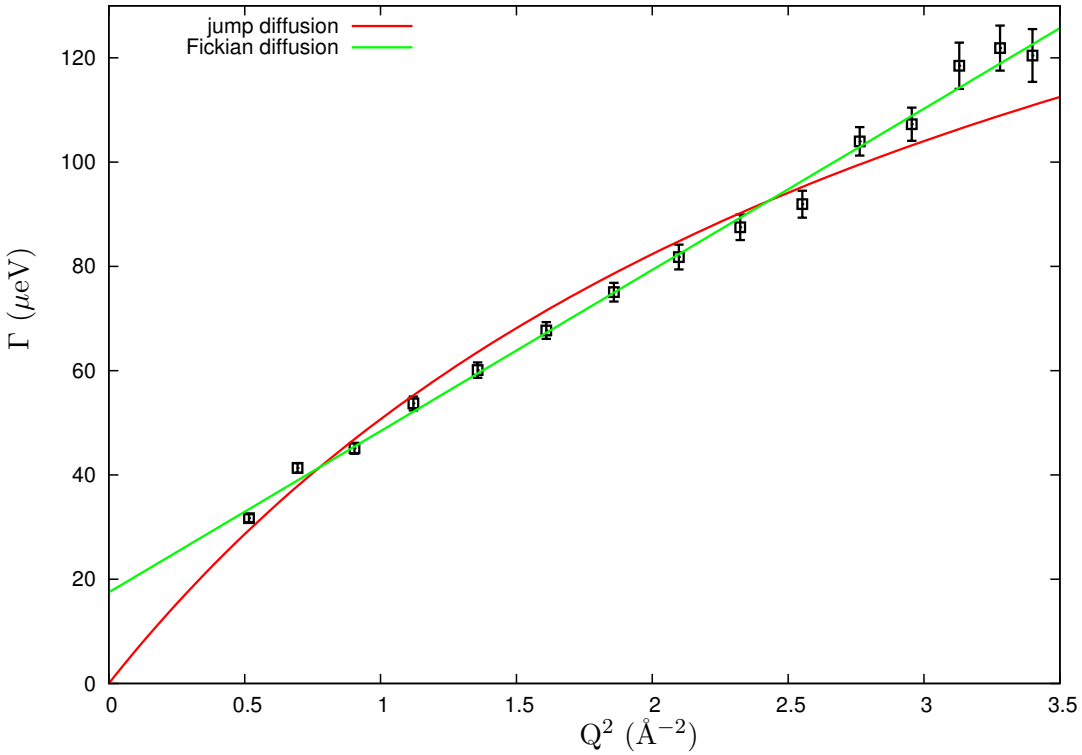

Figure 7.2: Half width at half maximum (Γ) of the fitted Lorentzians at 306K plotted as a function of $\mathbf{Q}^2$. The lines are fits to the Jump Diffusion model, Equation 7.16 and the expected Q^2 dependence for Fickian diffusion plus a constant energy offset as described by Equation 7.18.

results.

T (K)	Ref [131]	Ref [130]	MD	D^{jump}_{QENS}	±	%	τ^{jump}_{QENS}	±	%
305.5	0.0370								
306				0.100	0.005	5.0	3.0	0.3	9.9
308		0.0411	0.06						
310				0.109	0.004	4.0	2.8	0.2	8.1
313	0.0461	0.0458							
315				0.115	0.003	2.8	2.5	0.2	6.4
318		0.0510							
323		0.0593							
325				0.137	0.004	3.3	2.1	0.2	8.0
340				0.163	0.004	2.2	1.6	0.1	6.3
355				0.200	0.006	2.8	1.3	0.1	9.8
362	0.103								
370				0.235	0.005	2.1	1.0	0.1	7.3
373	0.147		0.17						
423	0.290								
454	0.416								
458			0.41						

Table 7.2: Self diffusion coefficients D in $\text{Å}^2\text{ps}^{-1}$ measured using PFG-NMR[131] and the capillary method[130] and also calculated from the classical molecular dynamics simulations presented in Chapter 5. The error in the diffusion coefficients using PFG-NMR is estimated to be ±5%. The simulation results are correct to within $0.01\text{Å}^2\text{ps}^{-1}$. D and residence times τ were derived from the QENS spectra by fitting to a jump diffusion model, Equation 7.16.

At lower temperatures the widths of the fitted Lorentzians do not follow the expected Q dependence of the jump diffusion model as well as at higher temperatures. This is illustrated in Figure 7.2 for NMA at 306K. The fit assuming the jump diffusion model described by Equation 7.16 is shown alongside a fit assuming Fickian diffusion which is the low Q limit of the jump diffusion model,

$$\Gamma = \hbar D Q^2 \tag{7.17}$$

It is apparent from Figure 7.2 that a constant must be added to this equation in order to obtain a good fit of the QENS widths to a straight line such that

$$\Gamma = \hbar D Q^2 + \Gamma_\circ \tag{7.18}$$

where the energy offset $\Gamma_\circ$, is constant in Q.

At 306K the value of this constant is 17.5 μeV and the gradient of the fit in Figure

T (K)	D_{QENS}^{Fick}	±	% error	$\Gamma_\circ$	±	% error
306	0.047	0.001	2.171	17.5	0.9	5.139
306	0.047	0.001	2.564	17.7	1.0	5.774
310	0.051	0.001	1.779	18.4	0.8	4.131
315	0.057	0.001	1.696	18.1	0.8	4.191
315	0.058	0.001	1.876	17.7	0.8	4.596
325	0.072	0.002	2.611	16.7	1.1	7.119
340	0.093	0.003	3.026	17.8	1.6	9.081
355	0.125	0.003	2.786	14.5	1.4	9.66
370	0.145	0.005	3.388	15.8	1.9	12.08
370	0.151	0.005	3.191	14.9	1.8	11.91

Table 7.3: Self diffusion coefficients obtained from fits of the QENS Lorentzian widths to Equation 7.18 which describes Fickian diffusion but also includes an energy offset $\Gamma_\circ$.

7.2 gives a self diffusion coefficient of 0.047 ± 0.001 $Å^2ps^{-1}$. This is much closer to the literature values: 0.0370 $Å^2ps^{-1}$ at 305.5K using PFG-NMR[131] and 0.0411 $Å^2ps^{-1}$ at 308.15K using the capillary method[130]. It also gives better agreement with the self diffusion coefficient calculated using the CHARMM22 forcefield which is $0.06\pm0.01Å^2ps^{-1}$.

The self-diffusion coefficients and values of $\Gamma_\circ$ derived from fits to Equation 7.18 are listed in Table 7.3 and the fits themselves are plotted in Figure 7.3.

For liquid NMA at temperatures between 306 and 340K the Q dependence of the Lorentzian widths is better described by Fickian diffusion plus a constant energy offset than by the jump diffusion model. For these temperatures the average value of $\Gamma_\circ$ is $\approx 17.8\mu$eV corresponding to a characteristic time of $\frac{\hbar}{\Gamma_\circ} = 37$ps.

Above 340K the widths tend towards a constant at high Q and the data is better fitted by the jump diffusion model. At these higher temperatures when Fickian diffusion is assumed the width of $\Gamma_\circ$ decreases with temperature.

Chen *et al* find that the temperature dependence of the self diffusion coefficients measured by PFG-NMR between 333 and 423K follows an Arrhenius relation with an activation energy of 18.4 ± 0.5 $kJmol^{-1}$ (190.7 meV)[131]. This Arrhenius relation,

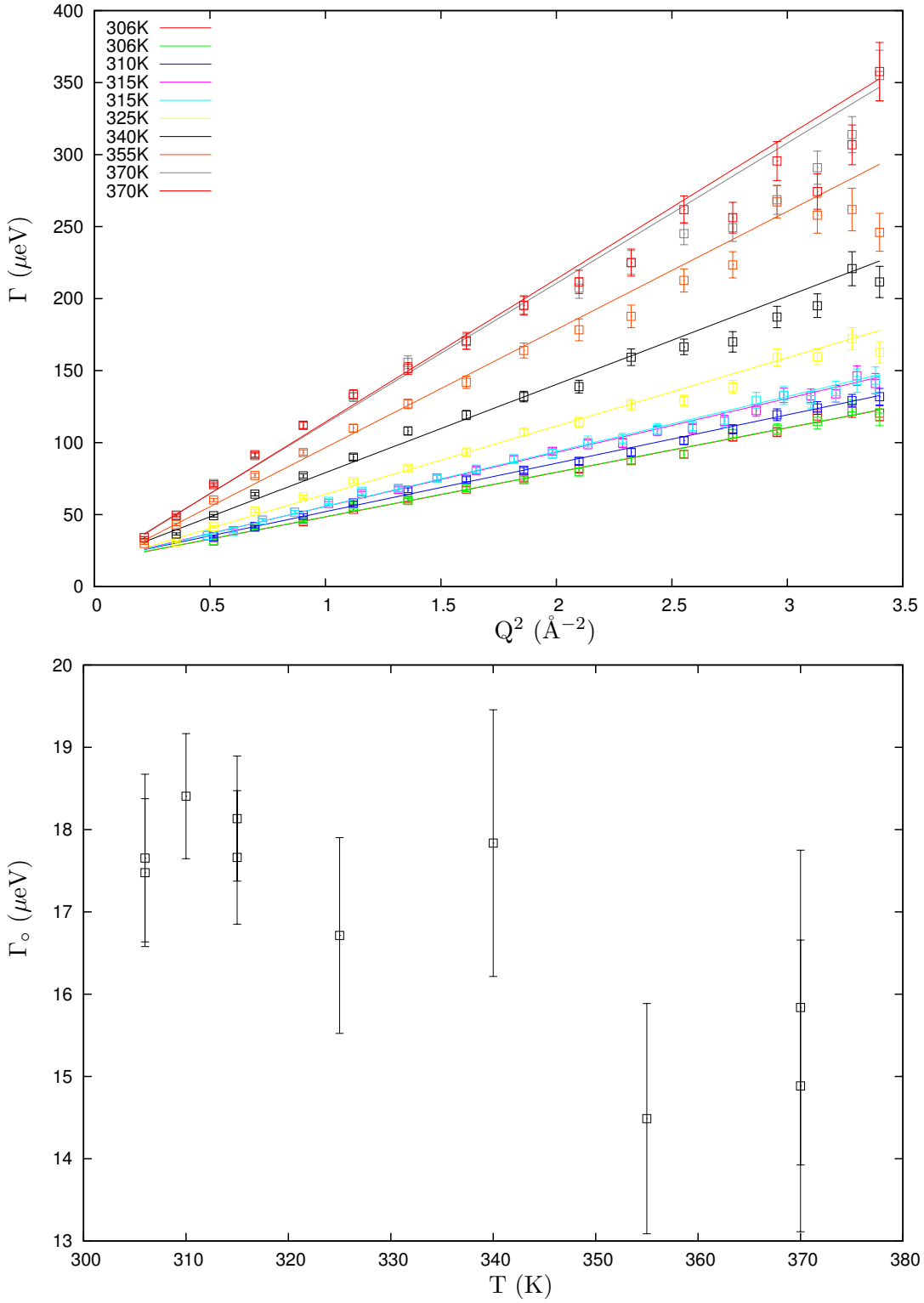

Figure 7.3: Fits to the QENS data using Equation 7.18. The constant energy offsets $\Gamma_\circ$ are plotted in the lower figure as a function of temperature.

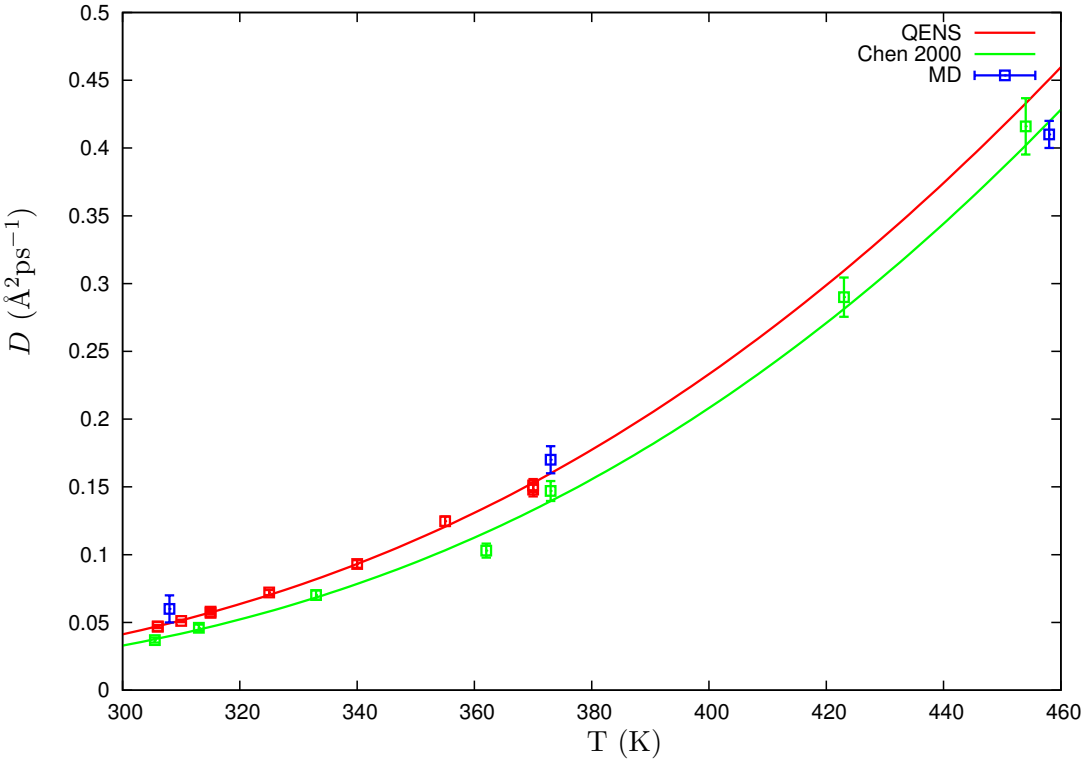

Figure 7.4: Self-diffusion coefficients as a function of temperature from fits to QENS data using Equation 7.18 (red), PFG-NMR[131] (green) and from the classical simulations (blue).

along with a similar fit to the QENS self-diffusion coefficients listed in Table 7.3, is plotted in Figure 7.4. The activation energy for the diffusive process derived from the QENS measurements is 179 ± 2 meV. This was calculated using the full temperature range of 306 to 370K, whereas Chen's value applies only between 333 and 423K. Consideration of the full temperature range of Chen's experiments yields an activation energy of 184 meV which is closer to the QENS value.

One of the first assumptions made in order to achieve a tractable theoretical model of the quasielastic spectra was that the translational, rotational and vibrational motions occur on well separated time scales and can be decoupled as in Equation 7.5. It was also assumed that rotational motions would occur on much faster timescales than the translational motion and thus that the QENS spectra would be adequately described by a model that included only Q-dependent translational diffusion. From these results it appears that this standard method of analysing QENS data cannot be applied to NMA as there is a Q-independent contribution, probably due to rotational motions at a similar timescale to the translational diffusive motion of NMA molecules. The rotational relaxation time of liquid NMA has been measured recently using OHD-OKE spectroscopy and ranges from 19ps at 303K to approximately 7ps at 453K[65]. These

timescales are slower than the equivalent rotation times in smaller molecular liquids such as water or hydrogen fluoride and fall within the 1 to 75ps window observable using IRIS. While the results presented here show that the QENS data is consistent with the NMR measurements when a Q-independent contribution is included further analysis is required in order to quantify the extent of the rotational - translational coupling.

7.4.1 Aqueous dNMA

An aqueous dNMA solution of concentration $X_{dNMA} = 0.2$ was also investigated. At 308K the QENS spectra were well represented by single Lorentzians convolved with the instrument resolution function and a flat background. The Q^2 dependence of the width of the Lorentzians fits that of the jump diffusion model to give a self diffusion coefficient of $D = 0.111 \pm 0.003 \text{Å}^2\text{ps}^{-1}$ and an interval between jumps of $\tau = 2.72 \pm 0.04$ps. The solution was made using deuterated NMA and distilled water therefore the motions observed are those of the hydrogens belonging to water molecules. The self diffusion of water in a $X_{NMA} = 0.2$ classical molecular dynamics simulation was calculated to be $0.118 \pm 0.007 \text{Å}^2\text{ps}^{-1}$ which is very similar to the QENS value. This result is meaningful in several ways. Firstly, it shows that assuming the translational and rotational motions to be well separated in timescale is appropriate for liquid water, even for water in an environment rich in peptide groups. Secondly, the agreement verifies that combining empirical forcefields based on the pure components does give meaningful predictions for liquid mixtures and confirms the results of the aqueous simulations which show a marked slowing in the water dynamics when NMA is added.

7.4.2 Discussion

The QENS results presented in this chapter are complementary to the MD results presented in the rest of the thesis and indeed comparisons between the two are both possible and illuminating. The most obvious point of contact is a comparison of self diffusion coefficients. The QENS spectra of liquid NMA can be modelled by assuming Fickian diffusion coupled with a Q-independent width corresponding to rotational motion. By taking this approach the QENS data have been shown to be consistent with the results of PFG-NMR experiments as well as with the classical molecular dynamics simulations. It must be stressed that this is not the normal method of analysing QENS data and that further analysis of the QENS spectra is required to fully determine the nature and temperature dependence of the rotational-translational coupling of motions in liquid NMA. Insight into this coupling could be achieved by exploiting the connection between simulation and experiment and determining rotational correlation times of NMA molecules from the existing simulation trajectories. Although this thesis presents

both computational and experimental results it has barely scratched the surface with regard to the potential for combining these two techniques. The work of Kneller and co-workers in developing the package *n*Moldyn for calculating neutron scattering intensities from MD simulations shows the greater depth which is possible and indicates future directions for the comparison of MD and QENS data[150].

Chapter 8

Conclusions

The molecule N-methylacetamide is a minimal model of the peptide linkage ubiquitous in biological macromolecules and its behaviour in aqueous solution captures some of the essential features of protein interactions in their natural environment. In this thesis classical molecular dynamics simulations of aqueous NMA solutions have been carried out across the full concentration range and the structure and dynamics of the mixtures have been characterised using a variety of analysis methods.

Water has been found to have a nontrivial effect on NMA - NMA interactions through the formation of water bridges which promote close contact between NMA oxygens. These structures which connect hydrogen bonded NMA chains are thought to contribute to a slowing down of the system dynamics in the NMA-rich region. Water preferentially donates its hydrogens to NMA carbonyl oxygens but is a poor acceptor of hydrogen bonds from NMA and thus the proportion of free amide groups increases with water mole fraction. NMA molecules hydrogen bond to form predominantly linear but also branched chains which show a small proportional increase as water is added and the size of NMA clusters decreases suggesting that chains preferentially break at linear sections.

While NMA cluster topology shows little qualitative change with concentration, that of water displays significant topological change upon the addition of even a small amount of NMA. Although water molecules in the aqueous solutions form system size clusters at concentrations up to and including $X_{NMA} = 0.4$, the topology of these clusters in the water-rich regime is fundamentally different to that of pure water. The local structural order around bulk water molecules contained within the system size clusters is much more random than the prevailing tetrahedral order found in pure water. Interestingly it is water molecules with a combination of NMA and water neighbours which optimise the tetrahedral arrangement of their hydrogen bonds. By forming strong and persistent hydrogen bonds with free hydrogen bonding sites on NMA carbonyl oxygens at chain

ends and mid-chain links, water molecules in the aqueous solutions appear to mimic the structure of NMA and form cross-linked networks enclosing the NMA hydrogen bonded chains.

It is not obvious that combining empirical forcefields derived for pure components or parameterised to a different water model will necessarily give realistic predictions for liquid mixtures, therefore the good agreement between simulation and QENS experiment for the self diffusion of water in $X_{NMA} = 0.2$ aqueous solution inspires confidence in the other dynamical simulation results. Classical molecular dynamics simulation has its limitations, for example, the dielectric constant of CHARMM22 NMA is much less than the experimental value[54]. The use of polarizable classical forcefields or more exact *ab initio* techniques gives a more accurate description of the real system; however, comparisons with these methods and experiment indicate that the classical forcefield description gives reasonably realistic results. For an investigation of medium range intermolecular structure and dynamics such as that carried out for this thesis classical molecular dynamics simulation will remain the most suitable technique until computing power permits the simulation of much larger systems by more exact quantum mechanical methods.

8.1 Future directions

Neutron scattering experiments on NMA were scheduled for 2005 but did not take place due to repair work on the SANDALS instrument at ISIS. Consequently the measurement of accurate experimental data on liquid and aqueous NMA structure for comparison with the radial distributions presented in this thesis remains outstanding.

The QENS results presented in Chapter 7 require further analysis in order to quantify the extent of the translational-rotational coupling in liquid NMA. It is expected that calculating the rotational correlation times of NMA molecules from the existing simulations will give insight into this coupling and aid the interpretation of the experimental results. Molecular dynamics is a powerful technique giving access to detailed information on atomic and molecular length scales which can easily be analysed statistically or visualised to give a qualitative feel for the system properties. The time and length scales accessible by simulation are the same as those probed by many experimental techniques including neutron scattering or IR spectroscopy. Therefore the level of detail available from simulation data can be a great aid to the interpretation of experimental results and simulations should be planned so as to optimise their comparability with experiment.

Water is increasingly recognised as an active component in biological systems and the investigation of aqueous NMA has revealed several features of water behaviour even in this simple system which are worthy of future study. In particular, characterising the tetrahedral order using an alternative orientational order parameter to q would be of interest due to the high proportion of water molecules which form three rather than four hydrogen bonds in the NMA-rich concentration range. The strong interest in supercooled water has led to sophisticated methods for analysing translational, orientational and dynamical order in these systems and the application of these techniques to aqueous NMA might be expected to give results relevant to the behaviour of protein hydration water.

As a model system for proteins and protein hydration NMA is obviously limited due to its small size. Nevertheless, the results presented in this thesis show that strong perturbation of water structure and dynamics is a fundamental property of the peptide group. The advantage of investigating such a small system is that the intermolecular interactions are simple enough to quantify upon an individual basis through all-atom molecular dynamics and enable direct comparison with experimental techniques such as QENS. NMA contains only one peptide link so a logical extension of the present work is to apply these techniques to larger molecules such as di-peptides or amino acids. Investigation of these molecules has already begun here in the School of Physics at the University of Edinburgh.

Appendix A

Ewald summation method

Electrostatic interactions between charged particles are long ranged forces and cannot be adequately described by a simple truncation as is possible for van der Waals interactions. This Appendix presents an outline of the Ewald summation method for the calculation of long range electrostatic contributions to the potential energy of a neutral system of point charges in a cubic box with periodic boundary conditions. More comprehensive treatments and discussion can be found from many sources including Allen and Tildesley[92] or Frenkel and Smit[94].

The Coulombic contribution to the potential energy of such a system of N charges q_i is

$$U_{coulomb} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(r_i) \tag{A.1}$$

where $\phi(r_i)$ is the electrostatic potential at the position of the i^{th} charge due to its interaction with all the other charges and their periodic images.

$$\phi(r_i) = \sum_{j}^{N} {\sum_{\mathbf{n}}}' \frac{q_j}{|\mathbf{r}_{ij} + \mathbf{n}L|} \tag{A.2}$$

The prime on the summation indicates that although each particle interacts with all its periodic images it does not interact with itself. L is the side length of the cubic box and $\mathbf{n}$ are a set of lattice vectors defining the displacement from the fundamental box of its periodic images. This summation does not converge quickly enough to enable its calculation in a simulation.

The solution is to split the potential into two parts which both converge quickly, one in Fourier space and one in real space and this is shown schematically in Figure A.1. On the left of the diagram is the real system depicted as a set of $\delta-$functions. This can be considered as the sum of the two right hand diagrams. In the first of the two right hand diagrams each point charge is screened by a Gaussian charge distribution whose total charge exactly equals that of the point charge. Therefore the potential felt at a distance from a point charge is due to just the fraction of the point charge which is not screened by the Gaussian distribution at that distance. This fractional unscreened charge falls quickly to zero as the distance increases and therefore the potential for this

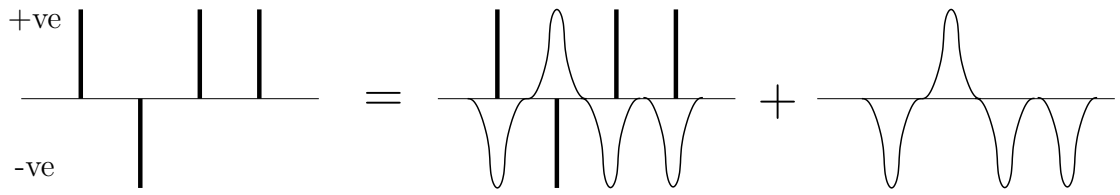

Figure A.1: Pictorial representation of the Ewald method showing how a set of Gaussian charge distributions can be introduced in order to separate the single poorly converging summation of point charge contributions into two components, each of which is rapidly convergent.

diagram can be summed in real space. A second set of smoothly varying Gaussians must then be introduced to compensate for the screening charges and these are shown in the final diagram of Figure A.1. The summation for this contribution converges rapidly in Fourier space.

Screening charge contributions : the Fourier space summation

The screening charge distributions are assumed to be Gaussians with width $\sqrt{2/\alpha}$.

$$\rho_{gauss}(r) = -q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha r^2\right) \tag{A.3}$$

where the parameter α can be chosen to optimise the speed of the Ewald calculation. The charge distribution due to the periodic distribution of a set of such Gaussians is therefore

$$\rho_{screen}(r) = \sum_{j=1}^{N} \sum_{\mathbf{n}} q_j \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha |\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)\,|^2\right) \tag{A.4}$$

Taking the Fourier transform of this charge distribution gives

$$\begin{aligned}
\rho_{screen}(\mathbf{k}) &= \int_V \mathrm{d}\mathbf{r}\, \exp\left(-i\mathbf{k}\cdot\mathbf{r}\right) \rho_{screen}(r) \\
&= \int_V \mathrm{d}\mathbf{r}\, \exp\left(-i\mathbf{k}\cdot\mathbf{r}\right) \sum_{j=1}^{N} \sum_{\mathbf{n}} q_j \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha |\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)\,|^2\right) \\
&= \int_{\text{allspace}} \mathrm{d}\mathbf{r}\, \exp\left(-i\mathbf{k}\cdot\mathbf{r}\right) \sum_{j=1}^{N} q_j \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha |\mathbf{r} - \mathbf{r}_j|^2\right) \\
&= \sum_{j=1}^{N} q_j \exp\left(-i\mathbf{k}\cdot\mathbf{r}_j\right) \int_{\text{allspace}} \mathrm{d}\mathbf{r}\, \exp\left(-i\mathbf{k}\cdot(\mathbf{r} - \mathbf{r}_j)\right) \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp\left(-\alpha |\mathbf{r} - \mathbf{r}_j|^2\right) \\
&= \sum_{j=1}^{N} q_j \exp\left(-i\mathbf{k}\cdot\mathbf{r}_j\right) \exp\left(-\frac{k^2}{4\alpha}\right)
\end{aligned} \tag{A.5}$$

as the Fourier transform of a Gaussian is also a Gaussian. Poisson's equation in real space describes the electrostatic potential due to a charge distribution

$$-\nabla^2\phi(\mathbf{r}) = 4\pi\rho(\mathbf{r}) \tag{A.6}$$

and the Fourier transform of this equation is simply

$$k^2\tilde{\phi}(k) = 4\pi\tilde{\rho}(k) \tag{A.7}$$

Substituting Equation A.5 into Equation A.7 gives

$$\tilde{\phi}_{screen}(k) = \frac{4\pi}{k^2}\sum_{j=1}^{N} q_j \exp\left(-i\mathbf{k}\cdot\mathbf{r}_j\right)\exp\left(-\frac{k^2}{4\alpha}\right) \tag{A.8}$$

This expression is undefined for $\mathbf{k} = 0$. Fortunately this term in the Fourier sum is equal to zero if the periodic system is considered to be surrounded by a continuum with an infinite dielectric constant. Finally the contribution to the potential energy due to the set of Gaussian screened charges can be calculated by taking the Fourier transform of Equation A.8 and inserting it into the expression for the electrostatic potential energy

$$\begin{aligned} U_{screen} &= \frac{1}{2}\sum_{i}^{N} q_i\phi_{screen}(r_i) \\ &= \frac{1}{2}\sum_{\mathbf{k}\neq 0}\sum_{i,j=1}^{N}\frac{4\pi q_i q_j}{Vk^2}\exp\left(-i\mathbf{k}\cdot\mathbf{r}_j\right)\exp\left(-\frac{k^2}{4\alpha}\right) \end{aligned} \tag{A.9}$$

Self-interaction correction

Each point charge should interact with all other charges in the fundamental box and with all the periodic images, including its own. It should not, however, interact with itself, or with the Gaussian charge distribution that has been introduced to screen it. Equation A.9 includes such a contribution and this therefore must be calculated and subtracted from the final sum. The electrostatic potential due to a single Gaussian charge distribution is

$$-\nabla^2\phi_{gauss}(\mathbf{r}) = 4\pi\rho_{gauss}(\mathbf{r}) \tag{A.10}$$

The Gaussian distribution is spherically symmetric, integrating by using this property gives

$$\begin{aligned} \phi_{gauss}(r) &= 2q_i\left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}}\int_0^r \mathrm{d}r\,\exp(-\alpha r^2) \\ &= \frac{q_i}{r}\mathrm{erf}\left(\sqrt{\alpha}r\right) \\ \text{where}\quad \mathrm{erf}(x) &\equiv \frac{2}{\sqrt{\pi}}\int_0^x \exp(-u^2)\mathrm{d}u = \frac{2}{\sqrt{\pi}}\left(x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \ldots\right) \end{aligned} \tag{A.11}$$

The point charge is located at the very centre of the Gaussian charge distribution so the correction is equal to $\phi_{gauss}(r = 0)$. Using the Taylor expansion of the error function

$\mathrm{erf}(x)$ to calculate the limit of Equation A.12 at $r = 0$ gives

$$\phi_{gauss}(r=0) = 2q_i \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} \tag{A.12}$$

The correction required to remove the effect of the self-interaction included in Equation A.9 is therefore

$$\begin{aligned} U_{self} &= \frac{1}{2}\sum_i^N q_i \phi_{self}(r_i) \\ &= \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} \sum_i^N q_i^2 \end{aligned} \tag{A.13}$$

Point charges + screened Gaussians: the real space summation

The potential due to an individual point charge together with its screening Gaussian distribution can be written as

$$\begin{aligned} \phi_{short-range}(r) &= \frac{q_i}{r} - \frac{q_i}{r}\mathrm{erf}\left(\sqrt{\alpha}r\right) \\ &= \frac{q_i}{r}\mathrm{erfc}\left(\sqrt{\alpha}r\right) \end{aligned} \tag{A.14}$$

where $\mathrm{erfc}(x) \equiv 1 - \mathrm{erf}(x)$ is the complementary error function. The contribution of these screened interactions to the Coulomb potential energy is therefore

$$U_{short-range} = \frac{1}{2}\sum_{i\neq j}^N q_i q_j \frac{\mathrm{erfc}\left(\sqrt{\alpha}r_{ij}\right)}{r_{ij}} \tag{A.15}$$

Total electrostatic potential energy

The complete Ewald summation for the total electrostatic potential energy is thus the sum of Equations A.9, A.13 and A.15

$$U_{coulomb} = U_{screen} - U_{self} + U_{short-range} \tag{A.16}$$

$$= \frac{1}{2}\sum_{\mathbf{k}\neq 0}\sum_{i,j=1}^N \frac{4\pi q_i q_j}{Vk^2}\exp\left(-i\mathbf{k}\cdot\mathbf{r}_j\right)\exp\left(-\frac{k^2}{4\alpha}\right) - \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}}\sum_i^N q_i^2 + \frac{1}{2}\sum_{i\neq j}^N q_i q_j \frac{\mathrm{erfc}\left(\sqrt{\alpha}r_{ij}\right)}{r_{ij}}$$

Appendix B

Forcefield parameters

TIP3P water

Atom Number	Atom Type	Mass (amu)	Charge (e)
1	OT	15.9994	-0.834
2	HT	1.008	0.417
3	HT	1.008	0.417

Table B.1: Atomic masses and charges used in the CHARMM TIP3P model[76]

Atom 1	Atom 2	Separation (Å)
1	2	0.9572
1	3	0.9572
2	3	1.5139

Table B.2: Constraint distances for the CHARMM22 rigid TIP3P model in angstroms (1Å $= 1\times10^{-10}$m)

Atom Type	ϵ_i	$\frac{Rmin_i}{2}$
OT	-0.1521	1.7682
HT	-0.0460	0.2245

Table B.3: CHARMM potential parameters for non-bonded Van der Waals interactions used in the TIP3P model[76]. ϵ and R^{min} represent the well depth and the radial coordinate of the minimum of the potential well, measured in units of kcal mol^{-1} and Å respectively.

Atom Type 1	Atom Type 2	A (kcal Å^{12} mol^{-1})	B (kcal Å^{6} mol^{-1})
OT	OT	581923.3674	595.0144340
OT	HT	327.9038928	10.4743008
HT	HT	0.000003088	0.000753817

Table B.4: Van der Waals potential parameters used in the CHARMM TIP3P model[76]

SPC/E water

Atom Number	Atom Type	Mass (amu)	Charge (e)
1	OW	15.9994	-0.8476
2	HW	1.008	0.4238
3	HW	1.008	0.4238

Table B.5: Atomic masses and charges used in the SPC/E model[102]

Atom Type 1	Atom Type 2	A (kcal Å^{12} mol^{-1})	B (kcal Å^{6} mol^{-1})
OW	OW	628937.5774	625.0374

Table B.6: Potential parameters for non-bonded Van der Waals interactions used in the SPC/E model. The original paper (reference [102]) quotes values A = 0.37122 (kJ/mol)$^{(1/6)}$ nm and B = 0.3428 (kJ/mol)$^{(1/12)}$ according to the equation: $V_{LJ} = -\left(\frac{A}{r}\right)^6 + \left(\frac{B}{r}\right)^1$ 2 Note that the r^{-12} and r^{-6} terms in this equation are in the opposite order to the potentials used in the other forcefields.

F3C water

Number	Type	Mass (amu)	Charge (e)
1	OW	15.9994	-0.82
2	HW	1.008	0.41
3	HW	1.008	0.41

Table B.7: Atomic masses and charges used in the F3C model[107] (same as SPC[103])

Atom 1	Atom 2	k	r_o
1	2	500	1.0
1	3	500	1.0

Table B.8: Harmonic bond potential parameters used in the F3C model. The units of k and r_o are kcal mol^{-1} Å^{-2} and Å respectively.

Atom 1	Atom 2	Atom 3	k_θ	θ_o
2	1	3	120.0	109.47

Table B.9: Harmonic angle potential parameters used in the F3C model. The units of k_θ and θ_o are kcal mol^{-1} rad^{-2} and degrees respectively.

Type	ϵ_i	$\frac{Rmin_i}{2}$
OW	-0.1848	1.7766
HW	-0.0100	0.4500

Table B.10: CHARMM potential parameters for non-bonded Van der Waals interactions used in the F3C model.

Atom Type 1	Atom Type 2	A (kcal Å^{12} mol^{-1})	B (kcal Å^{6} mol^{-1})
OW	OW	748406.943	743.78
OW	HW	45.975235	2.8117
HW	HW	0.002824	0.0106

Table B.11: Van der Waals potential parameters used in the F3C model.

CHARMM22 NMA

Atom Number	Atom Type	Mass (amu)	Charge (e)
1	CC	12.011	-0.27
2	C	12.011	0.51
3	CN	12.011	-0.11
4	N	14.007	-0.47
5	O	15.999	-0.51
6	HC	1.0080	0.09
7	HC	1.0080	0.09
8	HC	1.0080	0.09
9	H	1.0080	0.31
10	HN	1.0080	0.09
11	HN	1.0080	0.09
12	HN	1.0080	0.09

Table B.12: Atomic masses and charges used in the NMA model[76]

Atom 1	Atom 2	k	r_o
1	6	644	1.111
1	7	644	1.111
1	8	644	1.111
3	10	644	1.111
3	11	644	1.111
3	12	644	1.111
1	2	500	1.490
2	5	1240	1.230
2	4	740	1.345
3	4	640	1.430
4	9	880	0.997

Table B.13: Harmonic bond potential parameters used in the NMA model[76]. The units of k and r_o are kcal mol^{-1} Å^{-2} and Å respectively.

Atom 1	Atom 2	Atom 3	k_θ	θ_o
6	1	2	66	109.5
7	1	2	66	109.5
8	1	2	66	109.5
1	2	4	160	116.5
1	2	5	160	121.0
5	2	4	160	122.5
2	4	9	68	123.0
2	4	3	100	120.0
9	4	3	70	117.0
4	3	10	103	109.5
4	3	11	103	109.5
4	3	12	103	109.5
6	1	7	71	108.4
6	1	8	71	108.4
7	1	8	71	108.4
10	3	11	71	108.4
10	3	12	71	108.4
11	3	12	71	108.4

Table B.14: Harmonic angle potential parameters used in the NMA model[76]. The units of k_θ and θ_o are kcal mol^{-1} rad^{-2} and degrees respectively.

Atom 1	Atom 3	k_{UB}	s_o
6	2	30.0	2.163
7	2	30.0	2.163
8	2	30.0	2.163
6	7	5.4	1.802
6	8	5.4	1.802
7	8	5.4	1.802
10	11	5.4	1.802
10	12	5.4	1.802
11	12	5.4	1.802

Table B.15: Harmonic Urey-Bradley angle potential parameters used in the NMA model[76]. The units of k_{UB} and s_o are kcal mol^{-1} Å^{-2} and Å respectively.

Atom Type	ϵ_i	$\frac{R_{min}i}{2}$
CC	-0.080	2.0600
C	-0.110	2.0000
CN	-0.080	2.0600
N	-0.200	1.8500
O	-0.120	1.7000
HC	-0.022	1.3200
H	-0.046	0.2245
HN	-0.022	1.3200

Table B.16: CHARMM potential parameters for non-bonded Van der Waals interactions used in the NMA model[76].

Atom 1	Atom 2	Atom 3	Atom 4	k_χ	δ	n	1-4 elec	1-4 vdW
1	2	4	9	2.5	180.0	2	1	1
1	2	4	3	1.6	0.0	1	1	0.05
1	2	4	3	2.5	180.0	2	1	0.05
5	2	4	9	2.5	180.0	2	1	1
5	2	4	3	2.5	180.0	2	1	0.1
10	3	4	2	0.0	0.0	3	1	1
11	3	4	2	0.0	0.0	3	1	1
12	3	4	2	0.0	0.0	3	1	1
10	3	4	9	0.0	0.0	3	1	1
11	3	4	9	0.0	0.0	3	1	1
12	3	4	9	0.0	0.0	3	1	1
4	2	1	6	0.0	0.0	3	1	1
4	2	1	7	0.0	0.0	3	1	1
4	2	1	8	0.0	0.0	3	1	1
5	2	1	6	0.0	0.0	3	1	1
5	2	1	7	0.0	0.0	3	1	1
5	2	1	8	0.0	0.0	3	1	1

Table B.17: Dihedral potential parameters used in the NMA model[76]. The units of k_χ and δ are kcal mol^{-1} and degrees respectively while n, the multiplicity, is an integer.

Atom 1	Atom 2	Atom 3	Atom 4	k_ψ	ψ	1-4 elec	1-4 vdW
4	2	3	9	40.0	0.0	0	0
2	1	4	5	240.0	0.0	0	0

Table B.18: Improper dihedral potential parameters used in the NMA model[76]. The units of k_ψ and ψ_o are kcal mol^{-1} rad^{-2} and degrees respectively.

Atom Type 1	Atom Type 2	A (kcal Å^{12} mol^{-1})	B (kcal Å^{6} mol^{-1})
OW	OW	628937.5774	625.0374
OW	CC	1133673.7922	710.9395
OW	C	1100254.7514	758.4220
OW	CN	1133673.7922	710.9395
OW	N	912213.4253	801.9030
OW	O	425631.2285	482.0900
OW	HC	45438.1369	103.0699
OW	H	348.4984	10.8544
OW	HN	45438.1369	103.0699

Table B.19: Van der Waals potential parameters: CHARMM NMA and SPC/E water

Atom Type 1	Atom Type 2	A (kcal Å^{12} mol^{-1})	B (kcal Å^{6} mol^{-1})
CC	CC	1913623.8690	782.5341
CC	C	1881714.5326	840.2868
CC	CN	1913623.8690	782.5341
CC	N	1615031.8077	903.9627
CC	O	782329.5878	553.7232
CC	HC	93273.1700	125.1084
CC	H	1225.7950	17.2465
CC	HN	93273.1700	125.1084
C	C	1845493.7600	901.1200
C	CN	1881714.5326	840.2868
C	N	1573041.1891	966.0636
C	O	756323.6039	589.5590
C	HC	88219.5194	131.7547
C	H	1044.3935	17.2385
C	HN	88219.5194	131.7547
CN	CN	1913623.8690	782.5341
CN	N	1615031.8077	903.9627
CN	O	782329.5878	553.7232
CN	HC	93273.1700	125.1084
CN	H	1225.7950	17.2465
CN	HN	93273.1700	125.1084
N	N	1316590.4012	1026.2906
N	O	620648.7107	620.1628
N	HC	68302.6392	134.6207
N	H	609.3337	15.2899
N	HN	68302.6392	134.6207
O	O	286370.4820	370.7531
O	HC	29572.3120	77.9603
O	H	191.7717	7.5493
O	HN	29572.3120	77.9603
HC	HC	2521.5629	14.8962
HC	H	5.8620	0.8637
HC	HN	2521.5629	14.8962
H	H	3.0883E-06	7.5382E-04
H	HN	5.8620	0.8637
HN	HN	2521.5629	14.8962

Table B.20: Van der Waals potential parameters used in the NMA model[76]

Interspecies van der Waals parameters

Atom Type 1	Atom Type 2	A (kcal Å^{12} mol^{-1})	B (kcal Å^{6} mol^{-1})
OT	CC	1092810.9238	694.3962
OT	C	1060152.3794	740.6186
OT	CN	1092810.9238	694.3962
OT	N	877991.7526	782.6454
OT	O	409170.9918	470.2297
OT	HC	43525.2196	100.3549
OT	H	327.9039	10.4743
OT	HN	43525.2196	100.3549
HT	CC	1225.7950	17.2465
HT	C	1044.3935	17.2385
HT	CN	1225.7950	17.2465
HT	N	609.3337	15.2899
HT	O	191.7717	7.5493
HT	HC	5.8620	0.8637
HT	H	3.0883E-06	7.5382E-04
HT	HN	5.8620	0.8637

Table B.21: Van der Waals potential parameters: CHARMM NMA and TIP3P water

Atom Type 1	Atom Type 2	A (kcal Å^{12} mol^{-1})	B (kcal Å^{6} mol^{-1})
OW	CC	18698.9500	95.3645
OW	C	18363.0671	102.3355
OW	CN	18698.9500	95.3645
OW	N	15510.0884	109.2119
OW	O	7233.5761	65.6413
OW	HC	678.7719	13.1575
OW	H	0.0238	0.0936
OW	HN	678.7719	13.1575
HW	CC	1.1487	0.3605
HW	C	1.1281	0.3869
HW	CN	1.1487	0.3605
HW	N	0.9528	0.4128
HW	O	0.4444	0.2481
HW	HC	0.0417	0.0497
HW	H	1.4593E-06	3.5382E-04
HW	HN	0.0417	0.0497

Table B.22: Van der Waals potential parameters: CHARMM NMA and F3C water

Appendix C

Simulation protocol

In this Appendix the simulation protocols used for this thesis are presented in a way which would enable another researcher to repeat the simulations.

Water simulations

Initialisation of configuration

The initial configurations for the simulations were generated by placing water molcules of the correct geometry (either TIP3P or SPC) in random orientations on a simple cubic lattice at the correct experimental density (994.11kgm^{-3} at 308.15K[91]).

Equilibration

This unnatural configuration was then run at a high "temperature" of 500K for 5ps while the density was held constant. This allows the molecules to move faster and thus lose their initial configuration more quickly than would occur at the lower temperature where results are collected. The system was then "cooled" in a series of steps and allowed to equilibrate at 308.15K. The equilibration of the system was checked by monitoring the energy convergence. Since the simulation was started from a lattice configuration the mean square displacement was checked to ensure that the initial configuration had "melted" and molecules were diffusing as in a liquid.

Verification

Simulations of the TIP3P, SPC/E and F3C models are widely reported and the results from the simulations are in reasonable agreement with those in the literature for radial distributions, coordination numbers, tetrahedral order parameters and self diffusion coefficients.

NMA

Initialisation and equilibration

Atomic coordinates for the trans-NMA molecule were obtained from the web [151]. Only trans-NMA molecules were included in the initial configurations as cis-NMA

molecules make up a very small percentage of naturally occuring NMA (2.7%[57], 1.5% and unaffected by solvation in water[152]). Experimental values of the density of pure liquid NMA at atmospheric pressure were found in the literature: 946.24 kgcm^{-3}[91] at 308.15K (35^{o}C) and 894 kgcm^{-3}[68] at 373.15K (100^{o}C).

Cubic lattices of 216 NMA molecules were created at these two densities, with box lengths of 30.256Å and 30.834Å respectively. The two systems were both thermalised under an NVT ensemble at 500K and 1bar for 25ps before their temperatures were reduced to the appropriate value and equilibrated at that temperature. The resulting configurations were then used as the starting point for further simulations.

A third state point was chosen at 458.15K (185^{o}C), unfortunately I could not find an experimental value for the density of NMA at this temperature. Therefore the higher density simulation box was allowed to evolve under an NPT ensemble with the temperature fixed at 458.15K for 75ps. The resulting box length was 31.388Å corresponding to a density of 847.47 kgm^{-3}.

The "natural" density of the CHARMM22 NMA forcefield

NPT simulations at 308.15K and 373.15K were run to test by how much the simulation densities deviated from the experimental values. After 75ps both systems had become more dense. The density of the system at 308.15K became 963.87 kgm^{-3}, within 2% of its original (and the experimental) value. An NPT simulation was also run at 308.15K but in this case the starting configuration was the lower density box. The final density of this second simulation run was 973.38 kgm^{-3}, within 3% of the experimental value. NPT simulations at 373.15K were also run for 75ps from both the starting configurations. In this case the final densities were very similar, 912.25 kgm^{-3} for the box whose initial density was set at the correct experimental density of 893.96 kgm^{-3} and 913.57 kgm^{-3} from the other set up for the experimental density at 308.15K (946.24 kgm^{-3}). Both these results are within 3% of the experimental value. Hence these results suggest that the density of the simulation at 458.15K should be within a few percent of its correct experimental value.

Optimisation of control parameters

The value used to truncate the potentials ought to be as large as possible but no larger than half the width of the simulation box otherwise it causes difficulties with the periodic boundary conditions. The Verlet shell width was optimised by running a series of short simulations using a variety of values and choosing the value for which the simulation ran most quickly.

Checks on finite size effects

While the size of simulation boxes is limited by the amount of available computing power the use of periodic boundary conditions may introduce systematic errors into the calculations. Therefore it is important to compare the results of simulations of large and small boxes in order to test whether or not finite size effects are present in the smaller (computationally less expensive) simulations. Equilibrated 30Å boxes

temperature (K)	308.15	373.15
density (kgm^{-3})	946.24	893.96
boxlength (Å)	60.5122210720	61.6686377668
number of molecules	1728	1728
pressure (bar)	1	1
ensemble	NVT	NVT
thermostat relaxation (ps)	0.05	0.05
timestep (ps)	0.0005	0.0005
equilibration (ps)	2.5	2.5
data collection (ps)	5	5
trajectory data freq (steps)	200	200
rdf data freq (steps)	200	200
cutoff (Å)	20	20
verlet shell (Å)	0.8	0.8

Table C.1: Large simulation control parameters

temperature (K)	308.15	373.15	458.15
density (kgm^{-3})	946.24	893.96	847.47
boxlength (Å)	30.2561105360	30.8343188834	31.3880777074
number of molecules	216	216	216
pressure (bar)	1	1	1
ensemble	NVT	NVT	NVT
thermostat relaxation (ps)	0.05	0.05	0.05
timestep (ps)	0.0005	0.0005	0.0005
equilibration (ps)	25	25	25
data collection (ps)	1000	50	50
trajectory data freq (steps)	200	200	200
rdf data freq (steps)	200	200	200
cutoff (Å)	14	14	14
verlet shell (Å)	0.8	0.8	0.8

Table C.2: Simulation control parameters

were replicated in three dimensions to create 60Å cubes which were then run using the parameters listed in Table C.1. The results for radial distributions, hydrogen bonding, clustering and dynamics agreed with those of the 30Å simulations.

Production runs

After equilibration, simulations were run for each of the three state points using the parameters listed in Table C.2.

X_{NMA}	ρ_{expt}[91] kg m^{-3}	N_{NMA}	N_{H_2O}	box length Å	molarity (M)
0.0	994.11	0	1000	31.1041861668	-
0.1002	997.15	69	620	29.9996476420	4.24
0.1999	997.54	125	500	31.1424445630	6.87
0.3006	992.74	140	327	29.9890406062	8.63
0.4012	985.26	160	240	29.9990549012	9.85
0.5000	977.75	175	175	30.0289133673	10.73
0.6018	970.28	185	123	29.9756501588	11.41
0.7019	963.35	194	83	30.0073584114	11.93
0.8070	956.69	200	48	29.9535446184	12.36
0.9015	951.17	207	23	30.0509137747	12.67
1.0	946.24	216	0	30.2594601494	-

Table C.3: Experimental densities, number of molecules, simulation box sizes and molarities.

Aqueous NMA

Initialisation

Initial configurations for each of the concentrations were obtained by placing the correct numbers of *trans*-NMA and water molecules randomly on an approximately 30Å lattice as for the simulations of the pure species. Densities at 308.15K were obtained from reference [91] and are listed in Table C.3 along with the numbers of water and NMA molecules and the simulation box length. Molarity is a commonly used measure of concentration defined as the number of moles of solute (in this case NMA) per litre of solution. The NMA mole fraction and the solution densities were used to calculate the molarities of the aqueous mixtures and these values are also listed in this table.

Equilibration and production runs

The initial configurations were thermalised at 500K and the temperature reduced in steps before equilibration at 308K as for the water and NMA systems. The same control parameters were used for the production runs as for the simulation of NMA at 308K listed in Table C.2.

Check on variation between ensembles

The effect of using the NVT ensemble was investigated by running a series of simulations on a $X_{NMA} = 0.2$ mixture with a range of thermostat relaxation times between $\tau_R = 0.01 - 2$ and also in the NVE ensemble. Both static and dynamic results were within a few percent for all the simulations which is approximately the same as the variation seen on running separate simulations using the same control parameters and ensemble.

Data Collection

Different data collection strategies were employed to study static and dynamic properties. First the equilibrated systems were evolved for 1ns while collecting trajectory and statistical data at 100fs intervals. Following this run two shorter simulations were carried out, one lasting for 4ps and the other for 100ps. Data was collected at 2fs and 10fs intervals respectively so that the short and long timescale dynamics could be characterised. Subsequently it was decided not to calculate the history dependent hydrogen bond lifetimes which required the short sampling intervals. Diffusion and history independent hydrogen bond autocorrelations were calculated using both the short and long sampling intervals to check that the results for different sampling intervals coincided. As the longer 1ns runs provided better statistics these were used to produce the graphical results in Chapters 4, 5 and 6.

Appendix D

Publications

Allison, S.K.; Bates, S.P.; Crain, J.; Martyna, G.J.: Solution structure of the aqueous model peptide *N*-methylacetamide *J. Phys. Chem. B* 110, 21319-21326 (2006)

Whitfield, T.W.; Martyna, G.J.; Allison, S.K.; Bates, S.P.; Vass, H.; Crain, J.: Structure and Hydrogen Bonding in Neat *N*-Methylacetamide: Classical Molecular Dynamics and Raman Spectroscopy Studies of a Liquid of Peptidic Fragments *J. Phys. Chem. B* 110, 3624-3637 (2006)

Whitfield, T. W.; Martyna, G. J.; Allison, S.K. ; Bates, S. P.; Crain, J. : Liquid NMA : A surprisingly realistic model for hydrogen bonding motifs in proteins *Chem. Phys. Lett.* 414(1-3), 210-214 (2005)

Bibliography

[1] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, "Molecular dynamics simulations of the complete satellite tobacco mosaic virus," *Structure*, vol. 14, pp. 437–449, 2006.

[2] Y. Levy and J. N. Onuchic, "Water mediation in protein folding and molecular recognition," *Ann. Rev. Biophys. Biomol. Struct.*, vol. 35, pp. 389–415, 2006.

[3] N. Prabhu and K. Sharp, "Protein-solvent interactions," *Chem. Rev.*, vol. 106, pp. 1616–1623, 2006.

[4] P. Ball, *H2O: A biography of water*. The Orion Publishing Group Ltd, 2000.

[5] F. Franks, *Water: A Comprehensive Treatise*. Plenum Press, 1972.

[6] M. Chaplin, "http://www.lsbu.ac.uk/water/," (date last accessed: 16/06/07).

[7] L. Pauling, "The structure and entropy of ice and of other crystals with some randomness of atomic arrangement," *J. Am. Chem. Soc.*, vol. 57, pp. 2680–2684, 1935.

[8] E. D. Issacs, A. Shukla, P. M. Platzman, D. R. Hamann, B. Barbiellini, and C. A. Tulk, "Covalency of the hydrogen bond in ice: A direct x-ray measurement," *Phys. Rev. Lett.*, vol. 82, pp. 600–604, 1999.

[9] Y. Mo, J. Gao, and S. D. Peyerimhoff, "Energy decomposition analysis of intermolecular interactions using a block-localized wave function approach," *J. Chem. Phys.*, vol. 112, pp. 5530–5538, 2000.

[10] T. Steiner, "The hydrogen bond in the solid state," *Agnew. Chem. Int. Ed.*, vol. 41, pp. 48–76, 2002.

[11] F. Jensen, *Introduction to Computational Chemistry*. Wiley, 1999.

[12] T. Steiner and G. R. Desiraju, "Distinction between the weak hydrogen bond and the van der Waals interaction," *Chem. Commun.*, pp. 891–892, 1998.

[13] F. W. Starr, J. K. Nielsen, and H. E. Stanley, "Hydrogen-bond dynamics for the extended simple point-charge model of water," *Phys. Rev. E*, vol. 62, pp. 579–587, 2000.

[14] R. Kumar, J. R. Schmidt, and J. L. Skinner, "Hydrogen bonding definitions and dynamics in liquid water," *J. Chem. Phys.*, vol. 126, p. 204107, 2007.

[15] F. N. Keutsch and R. J. Saykally, "Water clusters: Untangling the mysteries of the liquid, one molecule at a time," *P.N.A.S.*, vol. 98, pp. 10533–10540, 2001.

[16] C. Levinthal, "How to fold graciously," *Mossbaur Spectroscopy in Biological Systems Proceedings, Univ. of Illinois Bulletin*, vol. 67, pp. 22–24, 1969.

[17] A. R. Fersht and V. Daggett, "Folding and binding: implementing the game plan," *Curr. Op. Struct. Biol.*, vol. 17, pp. 1–2, 2007.

[18] S. Walter and J. Buchner, "Molecular chaperones - cellular machines for protein folding," *Agnew. Chem. Int. Ed.*, vol. 41, pp. 1098–1113, 2002.

[19] A. E. Mirsky and L. Pauling, "On the structure of native, denatured and coagulated proteins," *Proc. Natl. Acad. Sci.*, vol. 22, pp. 439–447, 1936.

[20] J. A. Schellman, "The stability of hydrogen-bonded peptide structures in aqueous solution," *C. R. Trav. Lab. Carlsberg (Chim.)*, vol. 29, pp. 230–259, 1955.

[21] W. Kauzmann, "Some factors in the interpretation of protein denaturation," *Adv. Prot. Chem.*, vol. 14, pp. 1–63, 1959.

[22] I. M. Klotz and J. S. Franzen, "Hydrogen bonds between model peptides in solution," *J. Am. Chem. Soc.*, vol. 84, pp. 3461–3466, 1962.

[23] K. Dill, "Dominant forces in protein folding," *Biochemistry*, vol. 29, pp. 7133–7155, 1990.

[24] C. N. Pace, B. A. Shirley, M. McNutt, and K. Gajiwala, "Forces contributing to the conformational stability of proteins," *FASEB*, vol. 10, pp. 75–83, 1996.

[25] J. M. Richardson, M. M. Lopez, and G. I. Makhatadze, "Enthalpy of helix-coil transition: Missing link in rationalizing the thermodynamics of helix-forming propensities of the amino acid residues," *Proc. Natl. Acad. Sci.*, vol. 102, pp. 1413–1418, 2005.

[26] N. Ben-Tal, D. Sitkoff, I. A. Topol, A.-S. Yang, S. K. Burt, and B. Honig, "Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution," *J. Phys. Chem. B*, vol. 101, pp. 450–457, 1997.

[27] A. Moglich, K. Joder, and T. Kiefhaber, "End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation," *Proc. Natl. Acad. Sci.*, vol. 103, pp. 12394–12399, 2006.

[28] G. D. Rose, P. J. Fleming, J. R. Banavar, and A. Maritan, "A backbone-based theory of protein folding," *Proc. Natl. Acad. Sci.*, vol. 103, pp. 16623–16633, 2006.

[29] D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, "Protein hydration in solution: experimental observation by x-ray and neutron scattering," *Proc. Natl. Acad. Sci.*, vol. 95, pp. 2267–2272, 1998.

[30] F. Merzel and J. C. Smith, "Is the first hydration shell of lysozyme of higher density than bulk water?," *Proc. Natl. Acad. Sci.*, vol. 99, pp. 5378–5383, 2002.

[31] K. Modig, E. Liepinsh, G. Otting, and B. Halle, "Dynamics of protein and peptide hydration," *J. Am. Chem. Soc.*, vol. 126, pp. 102–114, 2004.

[32] B. Halle, "Protein hydration dynamics in solution: a critical survey," *Phil Trans. R. Soc. Lond. B*, vol. 359, pp. 1207–1224, 2004.

[33] F. Gabel, D. Bicout, U. Lehnert, M. Tehei, M. Weik, and G. Zaccai, "Protein dynamics studies by neutron scattering," *Quarterly Reviews of Biophysics*, vol. 35, pp. 327–367, 2002.

[34] S. Dixit, *Molecular models of hydration in methanol-water mixtures*. PhD thesis, University of Edinburgh, 2002.

[35] L. Dougan, *The Aqueous Environment: Lessons from Small Molecules*. PhD thesis, University of Edinburgh, 2005.

[36] C. Nieto-Draghi, R. Hargreaves, and S. P. Bates, "Structure and dynamics of water in aqueous methanol," *J. Phys. Cond. Mat.*, vol. 17, pp. S3265–S3272, 2005.

[37] S. K. Allison, J. P. Fox, R. Hargreaves, and S. P. Bates, "Clustering and micro-immiscibility in alcohol-water mixtures: evidence from molecular dynamics simulations," *Phys. Rev. B*, vol. 71, p. 024201, 2005.

[38] J. L. Katz and B. Post, "The crystal structure and polymorphism of *N*-methyl acetamide," *Acta. Cryst.*, vol. 13, p. 624, 1960.

[39] F. Hamzaoui and F. Baert, "X-ray study of static disorder in *N*-methylacetamide," *Acta. Cryst. C*, vol. 50, pp. 757–759, 1994.

[40] J. Eckert, M. Barthes, W. T. Klooster, A. Albinati, R. Aznar, and T. F. Koetzle, "No evidence for proton transfer along the N-H...O hydrogen bond in *N*-methylacetamide: neutron single crystal structure at 250 and 276K," *J. Phys. Chem. B*, vol. 105, pp. 19–24, 2001.

[41] T. Miyazawa, T. Shimanouchi, and S.-i. Mizushima, "Characteristic infrared bands of monosubstituted amides," *J. Am. Chem. Phys.*, vol. 24, pp. 408–418, 1956.

[42] F. Fillaux and J. Tomkinson, "Spectroscopic studies of the monosubstituted amides. VI. Neutron inelastic scattering spectra of *N*-methyl acetamide," *Chem. Phys.*, vol. 26, pp. 295–300, 1977.

[43] G. J. Kearley, M. R. Johnson, M. Pazanet, and E. Suard, "Structure and vibrational dynamics of the strongly hydrogen-bonded model peptide: *N*-methyl acetamide," *J. Chem. Phys.*, vol. 115, pp. 2614–2620, 2001.

[44] S. Rols, H. N. Bordallo, K. W. Herwig, and M. Barthes, "Dynamical disorder and reorientation of the CH3 groups in *N*-methylacetamide," *Physica B*, vol. 350, pp. e587–e589, 2004.

[45] H. N. Bordallo, D. N. Argyriou, M. Barthes, W. Kalceff, S. Rols, K. W. Herwig, C. Fehr, F. Juranyi, and T. Seydel, "Hydrogen in *N*-methylacetamide: Positions and dynamics of the hydrogen atoms using neutron scattering," *J. Phys. Chem. B*, vol. web release date: 16-June-2007, p. DOI: 10.1021/jp068528z, 2007.

[46] H. Akiyama, M. Torii, "Cooperative effect in hydrogen bonding of *N*-methylacetamide in carbon tetrachloride solution confirmed by NMR and IR spectroscopies," *Spectrochim. Acta. A Mol. Biomol. Spectrosc.*, vol. 56, pp. 137–144, 2000.

[47] W. A. Herrebout, K. Clou, and H. O. Desseyn, "Vibrational spectroscopy of *N*-methylacetamide revisited," *J. Phys. Chem. A*, vol. 105, pp. 4865–4881, 2001.

[48] R. Ludwig, F. Weinhold, and T. C. Farrar, "Structure of liquid *N*-methylacetamide: Temperature dependence of NMR chemical shifts and quadrupole coupling constants," *J. Phys. Chem. A*, vol. 101, pp. 8861–8870, 1997.

[49] D. G. Regan, B. E. Chapman, and P. W. Kuchel, "PGSE NMR diffusion study of the self-association of *N*-methylacetamide in carbon tetrachloride," *Magn. Reson. Chem.*, vol. 40, pp. S115–S121, 2002.

[50] X. G. Chen, R. Schweitzer-Stenner, S. Krimm, N. G. Mirkin, and S. A. Asher, "*N*-methylacetamide and its hydrogen-bonded water molecules are vibrationally coupled," *J. Am. Chem. Soc.*, vol. 116, pp. 11141–11142, 1994.

[51] I. Noda, Y. Liu, and Y. Ozaki, "Two-dimensional correlation spectroscopy study of temperature-dependent spectral variations of *N*-methylacetamide in the pure liquid state. 2. Two-dimensional raman and infrared-raman heterospectral analysis," *J. Phys. Chem.*, vol. 100, pp. 8674–8680, 1996.

[52] R. Schweitzer-Stenner, "Visible and UV-resonance Raman spectroscopy of model peptides," *J. Raman Spectrosc.*, vol. 32, pp. 711–732, 2001.

[53] R. Schweitzer-Stenner, G. Sieler, N. G. Mirkin, and S. Krimm, "Intermolecular coupling in liquid and crystalline states of trans-*N*-methylacetamide investigated by polarized Raman and FT-IR spectroscopies," *J. Phys. Chem. A*, vol. 102, pp. 118–127, 1998.

[54] T. W. Whitfield, G. J. Martyna, S. K. Allison, S. P. Bates, H. Vass, and J. Crain, "Structure and hydrogen bonding in neat *N*-methylacetamide: Classical molecular dynamics and Raman spectroscopy studies of a liquid of peptidic fragments," *J. Phys. Chem. B*, vol. 110, pp. 3624–3637, 2006.

[55] J. Edler and P. Hamm, "Spectral response of crystalline acetanilide and *N*-methylacetamide: Vibrational self-trapping in hydrogen-bonded crystals," *Phys. Rev. B*, vol. 69, pp. 214301–214309, 2004.

[56] A. L. Furer, N. A. Zhikhareva, and L. I. Maklakov, "Study of the structure of *N*-methylacetamide associates by vibrational spectroscopy," *J. Appl. Spectrosc.*, vol. 34, pp. 546–551, 1981.

[57] T. Koddermann and R. Ludwig, "*N*-methylacetamide/water clusters in a hydrophobic solvent," *Phys. Chem. Chem. Phys.*, vol. 6, pp. 1867–1873, 2004.

[58] L. M. Kutznetsova, V. L. Furer, and L. I. Maklakov, "Infrared intensities of *N*-methylacetamide associates," *J. Mol. Struct.*, vol. 380, pp. 23–29, 1996.

[59] R. Ludwig, R. O., and R. Winter, "Quantum cluster equilibrium theory of liquids: Temperature dependence of hydrogen bonding in liquid *N*-methylacetamide studied by IR spectra," *J. Phys. Chem. B*, vol. 102, pp. 9312–9318, 1998.

[60] S. Woutersen, Y. Mu, G. Stock, and P. Hamm, "Hydrogen-bond lifetime measured by time-resolved 2D-IR spectroscopy: *N*-methylacetamide in methanol," *Chem. Phys.*, vol. 266, pp. 137–147, 2001.

[61] M. A. Czarnecki and K. Z. Haufa, "Effect of temperature and concentration on the structure of *N*-methylacetamide-water complexes: Near-infrared spectroscopic study," *J. Phys. Chem. A*, vol. 109, pp. 1015–1021, 2005.

[62] H. Huang, S. Malkov, M. Coleman, and P. Painter, "Two-dimensional correlation infrared spectroscopic study of *N*-methylacetamide as a function of temperature," *J. Phys. Chem. A*, vol. 107, pp. 7697–7703, 2003.

[63] Y. Liu, Y. Ozaki, and I. Noda, "Two-dimensional Fourier-transform near-infrared correlation spectroscopy study of dissociation of hydrogen-bonded *N*-methylacetamide in the pure liquid state," *J. Phys. Chem.*, vol. 100, pp. 7326–7332, 1996.

[64] I. Noda, Y. Liu, and Y. Ozaki, "Two-dimensional correlation spectroscopy study of temperature-dependent spectral variations of *N*-methylacetamide in the pure liquid state. 1. Two-dimensional infrared analysis," *J. Phys. Chem.*, vol. 100, pp. 8665–8673, 1996.

[65] N. T. Hunt and K. Wynne, "The effect of temperature and solvation on the ultrafast dynamics of *N*-methylacetamide," *Chem. Phys. Lett.*, vol. 431, pp. 155–159, 2006.

[66] M. Akiyama, "Study on hydration enthalpy of *N*-methylacetamide in water," *Spectrochim. Acta. A Mol. Biomol. Struct.*, vol. 58, pp. 1943–1950, 2002.

[67] H. Guo and M. Karplus, "Ab initio studies of hydrogen bonding of *N*-methylacetamide: Structure, cooperativity and internal rotational barriers," *J. Phys. Chem.*, vol. 96, pp. 7273–7287, 1992.

[68] W. L. Jorgensen and C. J. Swenson, "Optimized intermolecular potential functions for amide and peptides. structure and properties of liquid amides," *J. Am. Chem. Soc.*, vol. 107, pp. 569–578, 1985.

[69] S. Trabelsi and S. Nasr, "Intermolecular association in liquid *N*-methylacetamide as studied by x-ray scattering," *J. Chem. Phys.*, vol. 121, pp. 6380–6385, 2004.

[70] M. Huelsekopf and R. Ludwig, "Correlations between structural, NMR and IR spectroscopic properties of *N*-methylacetamide," *Magn. Reson. Chem.*, vol. 39, pp. 127–134, 2001.

[71] R. Zhang, H. Li, Y. Lei, and S. Han, "Different weak C-H...O contacts in *N*-methylacetamide-water system: Molecular dynamics simulations and NMR experimental study," *J. Phys. Chem. B*, vol. 108, pp. 12596–12601, 2004.

[72] R. Zhang, H. Li, Y. Lei, and S. Han, "Structures and interactions in *N*-methylacetamide-water mixtures studied by IR spectra and density functional theory," *J. Mol. Struct.*, vol. 693, pp. 17–25, 2004.

[73] R. Vargas, J. Garza, R. A. Friesner, H. Stern, B. P. Hay, and D. A. Dixon, "Strength of the N-H...O=C and C-H...O=C bonds in formamide and *N*-methylacetamide dimers," *J. Phys. Chem. A*, vol. 105, pp. 4963–4968, 2001.

[74] T. W. Whitfield, J. Crain, and G. J. Martyna, "Structural properties of liquid *N*-methylacetamide via ab initio, path integral, and classical molecular dynamics," *J. Chem. Phys.*, vol. 124, p. 094503, 2006.

[75] J. B. O. Mitchell and S. L. Price, "On the relative strengths of the amide...amide and amide...water hydrogen bonds," *Chem. Phys. Lett.*, vol. 180, pp. 517–523, 1991.

[76] A. D. MacKerell Jr. et al, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Phys. Chem. B*, vol. 102, pp. 3586–3616, 1998.

[77] S. K. Gregurick, G. M. Chaban, and R. B. Gerber, "Ab initio and improved empirical potentials for the calculation of the anharmonic vibrational states and intramolecular mode coupling of *N*-methylacetamide," *J. Phys. Chem. A*, vol. 106, pp. 8696–8707, 2002.

[78] M. Buck and M. Karplus, "Hydrogen bond energetics: A simulation and statistical analysis of *N*-methyl acetamide (NMA), water, and human lysozyme," *J. Phys. Chem. B*, vol. 105, pp. 11000–11015, 2001.

[79] C. H. Langley and N. L. Allinger, "Molecular mechanics (MM4) and ab initio study of amide-amide and amide-water dimers," *J. Phys. Chem. A*, vol. 107, pp. 5208–5216, 2003.

[80] H. Torii, "Vibrational interactions in the amide I subspace of the oligomers and hydration clusters of *N*-methylacetamide," *J. Phys. Chem. A*, vol. 108, pp. 7272–7280, 2004.

[81] S. Ham, J.-H. Kim, H. Lee, and M. Cho, "Correlation between electronic and molecular structure distortions and vibrational properties. Ii. Amide I modes of NMAnD2O complexes," *J. Chem. Phys.*, vol. 118, pp. 3491–3498, 2003.

[82] H. Torii, T. Tatsumi, and M. Tasumi, "Effects of hydration on the structure, vibrational wavenumbers, vibrational force field and resonance Raman intensities of *N*-methylacetamide," *J. Raman. Spectrosc.*, vol. 29, pp. 537–546, 1998.

[83] J. Gao and M. Freindorf, "Hybrid ab initio QM/MM simulation of *N*-methylacetamide in aqueous solution," *J. Phys. Chem. A*, vol. 101, pp. 3182–3188, 1997.

[84] B. Mennucci and J. Martinez, "How to model solvation of peptides? Insights from a quantum-mechanical and molecular dynamics study of *N*-methylacetamide. 1. Geometries, infrared, and ultraviolet spectra in water," *J. Phys. Chem. B*, vol. 109, pp. 9818–9829, 2005.

[85] B. Mennucci and J. Martinez, "How to model solvation of peptides? Insights from a quantum mechanical and molecular dynamics study of *N*-methylacetamide. 2. 15N and 17O nuclear shielding in water and in acetone," *J. Phys. Chem. B*, vol. 109, pp. 9830–9838, 2005.

[86] Y. Mantz, H. Gerard, R. Iftimie, and G. J. Martyna, "Isomerization of a peptidic fragment studied theoretically in vacuum and in explicit water solvent at finite temperature," *J. Am. Chem. Soc.*, vol. 126, pp. 4080–4081, 2004.

[87] Y. Mantz, H. Gerard, R. Iftimie, and G. J. Martyna, "Ab initio and empirical model MD simulation studies of solvent effects on the properties of *N*-methylacetamide along a cis-trans isomerization pathway," *J. Phys. Chem. B*, vol. 110, pp. 13523–13538, 2006.

[88] K. Kwac and M. Cho, "Molecular dynamics simulation study of *N*-methylacetamide in water. I. Amide I mode frequency fluctuation," *J. Chem. Phys.*, vol. 119, pp. 2247–2255, 2003.

[89] R. Zhang, H. Li, Y. Lei, and S. Han, "All-atom molecular dynamic simulations and relative NMR spectra study of weak C-H...O contacts in amide-water systems," *J. Phys. Chem. B*, vol. 109, pp. 7482–7487, 2005.

[90] P. Assarsson and F. R. Eirich, "Properties of amides in aqueous solution: 1. A. Viscosity and density changes of amide-water systems B. Analysis of volume deficiencies of mixtures based on molecular size differences (mixing of hard spheres)," *J. Phys. Chem.*, vol. 72, pp. 2710–2719, 1968.

[91] P. J. Victor and D. K. Hazra, "Excess molar volumes, viscosity deviations, and isentropic compressibility changes in binary mixtures of *N*-methylacetamide + 2-methoxyethanol and *N*-methylacetamide + water at (308.15, 313.15, and 318.15)K," *J. Chem. Eng. Data*, vol. 47, pp. 79–82, 2002.

[92] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. Oxford University Press, 2001.

[93] F. Ercolessi, "A molecular dynamics primer." (http://www.fisica.uniud.it/ ercolessi/md/md/), 1997 (date accessed: 19/04/07).

[94] D. Frenkel and B. Smit, *Understanding Molecular Simulation (2nd Ed.)*. Academic Press, 2002.

[95] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2004.

[96] W. Smith, T. R. Forester, and I. T. Todorov, *The DL_POLY_2.0 User Manual (http://www.cse.scitech.ac.uk/ccg/software/DL_POLY/index.shtml)*. Daresbury Laboratory, United Kingdom.

[97] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, pp. 1695–1697, 1985.

[98] S. Nose, "A unified formulation of the constant temperature molecular dynamics methods," *J. Chem. Phys.*, vol. 81, pp. 511–519, 1984.

[99] H. J. C. Berendsen, J. P. M. Postma, W. F. van Guntersen, A. DiNola, and J. R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, vol. 81, pp. 3684–3690, 1984.

[100] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes," *J. Comp. Phys.*, vol. 23, pp. 327–341, 1977.

[101] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, pp. 926–935, 1983.

[102] H. Berendsen, J. Grigera, and T. Straatsma, "The missing term in effective pair potentials," *J. Phys. Chem.*, vol. 91, pp. 6269–6271, 1987.

[103] H. Berendsen, J. Postma, W. van Gunsteren, and J. Hermans, *Intermolecular forces*, ch. Interaction models for water in relation to protein hydration, pp. 331–342. D. Reidel Publishing, 1981.

[104] A. G. Csaszar, G. Czako, T. Furtenbacher, J. Tennyson, V. Szalay, S. Shirin, N. F. Zobov, and O. L. Polyansky, "On equilibrium structures of the water molecule," *J. Chem. Phys.*, vol. 122, p. 214305, 2005.

[105] P. L. Silvestrelli and M. Parrinello, "Structural, electronic and bonding properties of liquid water from first principles," *J. Chem. Phys.*, vol. 111, pp. 3572–3580, 1999.

[106] M. Tarek and D. J. Tobias, "The dynamics of protein hydration water: A quantitative comparison of molecular dynamics simulations and neutron-scattering experiments," *Biophysical J.*, vol. 79, pp. 3244–3257, 2000.

[107] M. Levitt, M. Hirshberg, R. Sharon, K. E. Laidig, and V. Daggett, "Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution," *J. Phys. Chem. B*, vol. 101, pp. 5051–5061, 1997.

[108] S. J. Grabowski, "Hydrogen bonding strength - measures based on geometric and topological parameters," *J. Phys. Org. Chem.*, vol. 17, pp. 18–31, 2004.

[109] J. R. Errington and P. G. Debenedetti, "Relationship between structural order and the anomalies of liquid water," *Nature*, vol. 409, pp. 318–321, 2001.

[110] M. Ferrario, M. Haughney, I. R. McDonald, and M. L. Klein, "Molecular-dynamics simulation of aqueous mixtures: Methanol, acetone and ammonia," *J. Chem. Phys.*, vol. 93, pp. 5156–5166, 1990.

[111] D. C. Rapaport, "Hydrogen bonds in water: Network organization and lifetimes," *Mol. Phys.*, vol. 50, pp. 1151–1162, 1983.

[112] A. Luzar, "Resolving the hydrogen bond dynamics conundrum," *J. Chem. Phys.*, vol. 113, pp. 10663–10675, 2000.

[113] P. Raiteri, A. Laio, and M. Parrinello, "Correlations among hydrogen bonds in liquid water," *Phys. Rev. Lett.*, vol. 93, p. 087801, 2004.

[114] A. Soper, "The radial distribution functions of water and ice from 220 to 673K and at pressures up to 400 MPa," *Chem. Phys.*, vol. 258, pp. 121–137, 2000.

[115] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, "Towards an assessment of the accuracy of density functional theory for first principles simulations of water," *J. Chem. Phys.*, vol. 120, pp. 300–311, 2004.

[116] Y. A. Mantz, B. Chen, and G. J. Martyna, "Structural correlations and motifs in liquid water at selected temperatures: *Ab Initio* and empirical model predictions," *J. Phys. Chem. B*, vol. 110, pp. 3540–3554, 2006.

[117] H.-S. Lee and M. E. Tuckerman, "Structure of liquid water at ambient temperature from *ab initio* molecular dynamics performed in the complete basis set limit," *J. Chem. Phys.*, vol. 125, p. 154507, 2006.

[118] Y. Wu, J. L. Tepper, and G. A. Voth, "Flexible simple point-charge water model with improved liquid-state properties," *J. Chem. Phys.*, vol. 124, p. 024503, 2006.

[119] I. M. Svishchev and P. G. Kusalik, "Structure in liquid water: A study of spatial density distributions," *J. Chem. Phys.*, vol. 99, pp. 3049–3058, 1993.

[120] J. Morgan and B. E. Warren, "X-ray analysis of the structure of water," *J. Chem. Phys.*, vol. 6, pp. 666–673, 1938.

[121] H. S. Frank and A. S. Quist, "Pauling's model and the thermodynamic properties of water," *J. Chem. Phys.*, vol. 34, pp. 604–611, 1961.

[122] A. H. Narten, M. D. Danford, and H. A. Levy, "X-ray diffraction study of liquid water in the temperature range 4-200C," *Discuss. Faraday Soc.*, vol. 43, pp. 97–107, 1967.

[123] J. A. Pople, "Molecular association in liquids II: Atheory of the structure of water," *Proc. Royal Soc. Lon.*, vol. 205, pp. 163–178, 1951.

[124] F. Sciortino, A. Geiger, and H. E. Stanley, "Isochoric differential scattering functions in liquid water: The fith neighbour as a network defect," *Phys. Rev. Lett.*, vol. 65, pp. 3452–3455, 1990.

[125] H. E. Stanley, D. V. Buldyrev, G. Franzese, N. Giovambattista, and F. W. Starr, "Static and dynamics heterogeneities in water," *Phil. Trans. R. Soc. A*, vol. 363, pp. 509–523, 2005.

[126] S. Melchionna, G. Briganti, P. Londei, and P. Cammarano, "Water induced effects on the thermal response of a protein," *Rhys. Rev. Lett.*, vol. 92, p. 158101, 2204.

[127] C. Vega, E. Sanz, and J. L. F. Abascal, "The melting temperature of the most common models of water," *J. Chem. Phys.*, vol. 122, p. 114507, 2005.

[128] A. Luzar and D. Chandler, "Hydrogen-bond kinetics in liquid water," *Nature*, vol. 379, pp. 55–57, 1996.

[129] M. Holz, S. R. Heil, and A. Sacco, "Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements," *Phys. Chem. Chem. Phys.*, vol. 2, pp. 4740–4742, 2000.

[130] W. D. Williams, J. A. Ellard, and L. R. Dawson, "Solvents having high dielectric constants. VI. diffusion in *N*-methylacetamide," *J. Am. Chem. Soc.*, vol. 79, pp. 4652–4654, 1957.

[131] L. Chen, T. Gross, and H.-D. Ludemann, "T,p-dependence of self-diffusion in the lower *N*-methylsubstituted amides," *Z. Phys. Chem.*, vol. 214, pp. 239–251, 2000.

[132] H. Flyvbjerg and H. G. Petersen, "Error estimates on averages of correlated data," *J. Chem. Phys.*, vol. 91, pp. 461–466, 1989.

[133] M. A. Williams, J. M. Goodfellow, and J. M. Thornton, "Buried waters and internal cavities in monomeric proteins," *Prot. Sci.*, vol. 3, pp. 1224–1235, 1994.

[134] R. Schweitzer-Stenner, G. Sieler, and H. Christiansen, "Competition between peptide-peptide and peptide-solvent hydrogen bonding probed by polarized Raman spectroscopy on *N*-methylacetamide," *Asian J. Phys.*, vol. 7, pp. 287–312, 1998.

[135] R. L. A. Timmer and H. J. Bakker, "Water as a molecular hinge in amidelike structures," *J. Chem. Phys.*, vol. 126, p. 154507, 2007.

[136] S. Song, S. A. Asher, S. Krimm, and K. D. Shaw, "Ultraviolet resonance Raman studies of trans and cis peptides: Photochemical consequences of the twisted π^* excited state," *J. Am. Chem. Soc.*, vol. 113, pp. 1155–1163, 1991.

[137] M. D. Elola and B. M. Ladanyi, "Computational study of structural and dynamical properties of formamide-water mixtures," *J. Chem. Phys.*, vol. 125, p. 184506, 2006.

[138] J. Zielkiewicz, "Preferential solvation of *N*-methylformamide, *N*,*N*-dimethylformamide and *N*-methylacetamide by water and alcohols in the binary and ternary mixtures," *Phys. Chem. Chem. Phys.*, vol. 2, pp. 2925–2932, 2000.

[139] W. L. Jorgensen and C. J. Swenson, "Optimized intermolecular potential functions for amides and peptides. hydration of amides," *J. Am. Chem. Soc.*, vol. 107, pp. 1489–1496, 1985.

[140] H. Guo and M. Karplus, "Solvent influence on the stability of the peptide hydrogen bond: A supramolecular cooperative effect," *J. Phys. Chem.*, vol. 98, pp. 7104–7105, 1994.

[141] M. Teeter, Martha, "Water-protein interactions: Theory and experiment," *Annu. Rev. Biophys. Biophys. Biochem.*, vol. 20, pp. 577–600, 1991.

[142] C. A. Bottoms, T. A. White, and J. J. Tanner, "Exploring structurally conserved solvent sites in protein families," *Proteins. Struct. Funct. Bioinf.*, vol. 64, pp. 404–421, 2006.

[143] A. K. Soper, F. Bruni, and M. A. Ricci, "Site-site pair correlation functions of water from 25 to 400 °c: Revised analysis of new and old diffraction data," *J. Chem. Phys.*, vol. 106, pp. 247–254, 1997.

[144] L. van Hove, "Correlations in space and time and Born approximation scattering in systems of interacting particles," *Phys. Rev.*, vol. 95, pp. 249–262, 1954.

[145] G. L. Squires, *Introduction to the theory of thermal neutron scattering.* CUP, 1978.

[146] M. Bee, *Quasielastic Neutron Scattering: Principles and Applications in Solid State Chemistry, Biology and Materials Science.* Adam Hilger, 1988.

[147] A. Munter, "NIST Center for Neutron Research," *Neutron News*, vol. 3, pp. 29–37, 1992.

[148] C. J. Carlile and M. A. Adams, "The design of the IRIS inelastic neutron spectrometer and improvements to its analysers," *Physica B*, vol. 182, pp. 431–440, 1992.

[149] W. S. Howells, *MODES Manual.* ISIS Pulsed Neutron and Muon Source, Rutherford Appleton Laboratory. unpublished.

[150] T. Rog, K. Murzyn, K. Hinsen, and G. R. Kneller, "nMoldyn : A program package for a neutron scattering oriented analysis of molecular dynamics simulations," *J. Comp. Chem.*, vol. 24, pp. 657–667, 2003.

[151] B. Cannon, D. Taylor, B. H. Dunford-Shore, W. Sulaman, B. Feng, F. Fabrizio, J. Holcomb, W. Wise, and T. Kazic, "Klotho: Biochemical compounds declarative database: http://www.biocheminfo.org/klotho/," 1994–present (date accessed: 18/07/04).

[152] A. Radzicka, L. Pedersen, and R. Wolfenden, "Influences of solvent water on protein folding: free energies of solvation of cis and trans peptides are nearly identical," *Biochem.*, vol. 27, pp. 4538–4541, 1988.