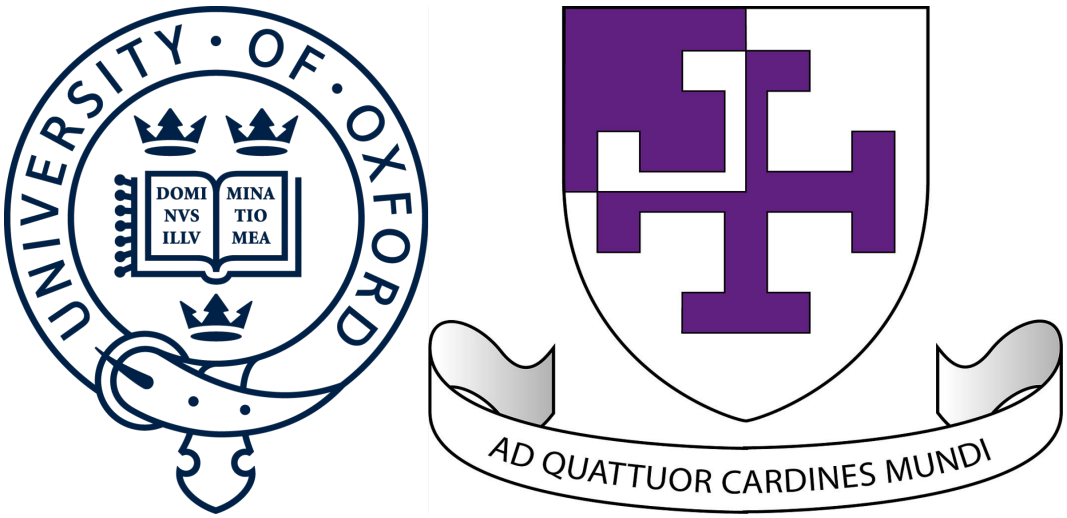
The Atomic Structure of Pharmaceuticals in Solution



Andrew J. Johnston

Department of Biochemistry

St Cross College, University of Oxford

This thesis was submitted to the Board of the Medical Sciences Division in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Trinity Term 2016

The Atomic Structure of Pharmaceuticals in Solution

Andrew J. Johnston, St Cross College, Doctor of Philosophy in Biochemistry, Trinity Term 2016

Investigations of the structure of indole, cocaine hydrochloride, alprazolam and clonidine hydrochloride in solution were undertaken using a combination of experimental, principally neutron diffraction, and computational techniques. This thesis aims to probe the interactions of the hydrophilic and hydrophobic groups of each these molecules with their solvent environment, drawing conclusions about how their interaction with solvent molecules and their conformation in solution are related to their biological function. Cocaine, alprazolam and clonidine are of particular interest because of their ability to cross the blood-brain barrier. Investigation of the solvation of indole, the functional group of the amino acid tryptophan, in methanol/water solutions highlighted the the role played by electrostatic interactions between the benzene ring of indole and its environment in dictating the location of tryptophan in the membrane bilayer. By studying the hydration of cocaine in aqueous solution, a possible mechanism explaining the ability of cocaine to cross the blood-brain barrier in its protonated form, which involves shielding of its hydrophilic groups through a water mediated internal hydrogen bond, was proposed. The preferentially solvated regions of the benzodiazepine alprazolam in methanol/water solutions indicated that these regions coincide with the location of bridging water molecules in the crystal structure of alprazolam bound to the BRD4 protein, which regulates transcription of oncogenes. A similar analysis of the solvation of clonidine in methanol/water solutions supported hypotheses relating to the mechanism of action of clonidine occuring *via* binding to adrenoceptors rather than imidazoline receptors. The work presented in this thesis demonstrates the importance of determining the structure of biomolecules in solution in gaining a comprehensive understanding of their behaviour *in vivo*.

Word count: 46934

Publications and Presentations

Below is a list of publications and presentations arising from the work presented in this thesis and associated collaborations.

Research Publications:

Solvation of Indole by Water and Methanol
A. Henao, **A. J. Johnston**, E. Guárdia, S. E. McLain and L. C. Pardo. *Phys. Chem. Chem. Phys.*, (Submitted).

The Solvation Structure of Alprazolam
A. Sridhar, **A. J. Johnston**, L. Varathan, S. E. McLain and P. C. Biggin. *Phys. Chem. Chem. Phys.*, (Under Review).

On the Atomic Structure of Cocaine in Solution
A. J. Johnston, S. Busch, L. C. Pardo, S. K. Callear, P. C. Biggin and S. E. McLain. *Phys. Chem. Chem. Phys.*, 18, 991-999 (2016).

Amphipathic Solvation of Indole: Implications for the Role of Tryptophan in Membrane Proteins
A. J. Johnston, Y. (Rosie) Zhang, S. Busch, L. C. Pardo, S. Imberti and S. E. McLain. *J. Phys. Chem. B*, 119 (19), 5979-5987 (2015).

Conformation and Interactions of Dopamine Hydrochloride in Solution
S. K. Callear, **A. J. Johnston**, S. E. McLain and S. Imberti. *J. Chem. Phys.*, 142 (1), (2015).

Presentations:

The Atomic Structure of Pharmaceuticals in Solution: Penetrating the Blood-brain Barrier
RMC-6 Conference, September 2015 (Oral Presentation), Budapest, Hungary.

The Atomic Structure of Pharmaceuticals in Solution
SANDALS 25$^{\text{th}}$ Anniversary Meeting, January 2015 (Oral Presentation), Abingdon, UK.

Cracking the Structure of Cocaine in Solution: Determining the Atomic Structure of Cocaine in Solution
American Conference on Neutron Scattering, June 2014 (Poster), Knoxville, Tennessee, USA.

Cracking the Structure of Cocaine: The Atomic Structure of Pharmaceuticals in Solution
Red Kite Meeting, January 2013 (Oral Presentation), Oxford, UK.

The Structure of Indole in Solution: Insights from Neutron Scattering and Computer Simulation
International Conference on Neutron Scattering, July 2013 (Poster), Edinburgh, UK.

Other Publications:

Cracking the Structure of Cocaine
A. J. Johnston. *St Cross College Alumni Magazine: Crossword*, 2014.

Cracking Cocaine's Secret to Sneaking into the Brain
J. Sudlow. *Chemistry World*, 2015.

Acknowledgements

I would like to thank all of the people who have been part of the McLain group during the course of my doctorate. This includes Sebastian Busch, who, besides showing me the ropes when I arrived, is responsible for the majority of the excellent plotting scripts the group now uses, and Richard Gillams, the person who has served in the group with me the longest, who offered help and verbose advice throughout. Thanks to Natasha Rhys for giving me this Latex thesis template! Fellow DPhil students include Nicola Steinke and Tom Dixon, who I thank for their frequent and perspicacious input. Undergraduate students who have worked in the group include Rosie Zhang, Luxmmi Varathan, Andrei Florea, Guillem Ramírez, Chris Sowden, Medeea Popescu and Alvaro Silva-Santisteban, all of whom contributed to my research. I would particularly like to recognise the work of Rosie, Luxmmi and Medeea on indole, alprazolam and clonidine, respectively.

I have also had the good fortune to work with a myriad of people outside the group, each of whom have made invaluable contributions to my research. Firstly, I acknowledge the support of Sam Callear and Silvia Imberti, instrument scientists at ISIS. I am particularly grateful to Sam for the opportunity to participate in her study of dopamine in aqueous solution. I also thank Luis Carlos Pardo of the Universitat Politècnica de Catalunya, who was generous enough to tailor his programme, ANGULA, for my research, and his colleague Andres Henao Aristizabal. In the final year of my DPhil I enjoyed a fruitful collaboration with Phil Biggin and Akshay Sridhar, who I thank for their MD expertise in particular. Thanks to Christina Redfield for her NMR expertise and for acting as my secondary supervisor. Thanks also to my thesis committee, Elspeth Garman and Tony Watts, for their time.

Finally, I thank Sylvia McLain for her relentless supervision, support, sagacity and for the chance to undertake a DPhil in the first place.

Andrew J. Johnston

Oxford, Summer 2016

Contents

Abstract iii

Publications and Presentations v

Acknowledgements vii

Abbreviations xiii

1 Introduction 1

1.1 Why Study the Structure of Pharmaceuticals in Solution? 1

1.2 Studying the Atomic Structure of Solutions 3

1.2.1 Theory of Liquid Structure 3

1.2.2 Simulation Techniques 5

1.2.3 Experimental Techniques 6

1.3 Molecules Able to Cross the Blood-brain Barrier 9

1.4 Aims of Thesis 12

1.5 Structure and Scope of Thesis 13

2 Theoretical Background 15

2.1 Neutron Diffraction 15

2.1.1 Fundamentals 15

2.1.2 The SANDALS Instrument 28

2.1.3 Measurements, Corrections and Processing 32

2.1.3.1 Measurements 32

2.1.3.2 Corrections and Processing 35

2.2 Empirical Potential Structure Refinement 40

2.3 ANGULA 47

2.3.1 Assigning the Coordinate System 48

2.3.2 Spatial Distribution 49

2.3.3 Whole Molecule Analysis 52

2.3.4 Orientation 53

2.3.5 Most Probable Position and Orientation 55

2.3.6 Intermolecular Association Plot 56

3 Indole in Methanol/water Solutions 59
3.1 Introduction . . . 59
3.2 Methods . . . 61
3.2.1 Neutron Diffraction Sample Preparation . . . 61
3.2.2 Neutron Diffraction Measurements . . . 63
3.2.3 Empirical Potential Structure Refinement . . . 64
3.2.4 Molecular Dynamics . . . 67
3.2.5 ANGULA . . . 68
3.2.6 NMR Sample Preparation and Measurement . . . 69
3.3 Results . . . 70
3.3.1 Diffraction Data . . . 70
3.3.2 Solvent Structure . . . 72
3.3.3 Indole Solvation . . . 76
3.3.3.1 Radial Distribution Functions . . . 76
3.3.3.2 Orientation . . . 83
3.3.3.3 Spatial Density Maps . . . 84
3.3.4 Simulation Checks . . . 89
3.3.4.1 The Effect of Dummy Atoms . . . 89
3.3.4.2 The Effect of Fitting . . . 89
3.3.4.3 Uncharged Systems . . . 91
3.3.5 Indole in Different Methanol/water Concentrations . . . 93
3.3.6 NMR . . . 95
3.3.6.1 Assignment in Chloroform . . . 95
3.3.6.2 Indole in Different Solvents . . . 96
3.4 Conclusions . . . 98

4 Cocaine in Aqueous Solution 103
4.1 Introduction . . . 103
4.2 Methods . . . 104
4.2.1 Neutron Diffraction Sample Preparation . . . 104
4.2.2 Neutron Diffraction Measurements . . . 105
4.2.3 X-ray Diffraction Measurements . . . 106
4.2.4 Empirical Potential Structural Refinement . . . 107
4.2.5 Molecular Dynamics . . . 109
4.2.6 ANGULA . . . 110
4.2.7 NMR Sample Preparation and Measurement . . . 111
4.3 Results . . . 112
4.3.1 Diffraction Data . . . 112
4.3.2 Cocaine-cocaine Interactions . . . 114
4.3.3 Cocaine Conformation . . . 119

4.3.4 Solvent Structure . . . 121
4.3.5 Cocaine Hydration . . . 122
4.3.5.1 Radial Distribution Functions . . . 122
4.3.5.2 Spatial Density Maps . . . 125
4.3.5.3 Whole Molecule Analysis . . . 127
4.3.5.4 Cocaine Hydration in MD Compared with EPSR . . . 129
4.3.6 NMR . . . 132
4.3.6.1 Assignment in Water and Methanol . . . 132
4.3.6.2 The Effect of Temperature . . . 133
4.3.6.3 Concentration Effects . . . 133
4.4 Conclusions . . . 137

5 Alprazolam in Methanol/water Solutions 141
5.1 Introduction . . . 141
5.2 Methods . . . 143
5.2.1 Neutron Diffraction Sample Preparation . . . 143
5.2.2 Neutron Diffraction Measurements . . . 143
5.2.3 Empirical Potential Structure Refinement . . . 144
5.2.4 Molecular Dynamics . . . 147
5.2.5 ANGULA . . . 147
5.3 Results . . . 148
5.3.1 Diffraction Data . . . 148
5.3.2 Alprazolam Conformation . . . 149
5.3.3 Solvent Structure . . . 151
5.3.4 Alprazolam Solvation . . . 154
5.3.4.1 Radial Distribution Functions . . . 154
5.3.4.2 Spatial Density Maps . . . 163
5.3.4.3 Whole Molecule Analysis . . . 167
5.3.4.4 The Effect of Temperature . . . 168
5.4 Conclusions . . . 169

6 Clonidine in Methanol/water Solutions 171
6.1 Introduction . . . 171
6.2 Methods . . . 173
6.2.1 Neutron Diffraction Sample Preparation . . . 173
6.2.2 Neutron Diffraction Measurements . . . 174
6.2.3 X-ray Diffraction Measurements . . . 174
6.2.4 Empirical Potential Structure Refinement . . . 175
6.2.5 Molecular Dynamics . . . 177
6.2.6 ANGULA . . . 177
6.2.7 NMR Sample Preparation and Measurement . . . 178

6.3 Results . . . 179
6.3.1 Diffraction Data . . . 179
6.3.2 Clonidine Conformation . . . 181
6.3.3 Solvent Structure . . . 182
6.3.4 Clonidine Solvation . . . 185
6.3.4.1 Radial Distribution Functions . . . 185
6.3.4.2 Spatial Density Maps . . . 192
6.3.4.3 Whole Molecule Analysis . . . 196
6.3.5 NMR . . . 197
6.3.5.1 Assignment of Hydrochloride and Freebase . . . 197
6.3.5.2 The Effect of Temperature . . . 198
6.4 Conclusions . . . 199

7 Conclusions and Further Work **203**
7.1 Conclusions . . . 203
7.2 Further Work . . . 208

A Self and Distinct Scattering Terms **213**

B 2D NMR Spectra **217**
B.1 Indole . . . 217
B.2 Cocaine . . . 219
B.3 Clonidine . . . 221

References **223**

Abbreviations

2D	2 dimensional
3D	3 dimensional
ACF	Auto correlation function
BET	Bromodomain and extraterminal domain
CN	Coordination number
CNS	Central nervous system
DFT	Density functional theory
EPSR	Empirical Potential Structure Refinement
FT	Fourier transform
IMAP	Intermolecular Association Plot
HIV	Human immunodeficiency virus
IHB	Internal hydrogen bond
IR	Infrared
LJ	Lennard-Jones
MC	Monte Carlo
MD	Molecular Dynamics
NME	New molecular entity
NMR	Nuclear magnetic resonance
PSA	Polar surface area
RDF	Radial distribution function
RMC	Reverse Monte Carlo
SANDALS	Small Angle Neutron Diffractometer for Amorphous and Liquid Samples
SD	Spatial density
SDM	Spatial Density Map
SLD	Spatial length denisty
THF	Tetrahydrofuran
TOF	Time-of-flight
Trp	Tryptophan
WMA	Whole Molecule Analysis

Chapter 1

Introduction

1.1 Why Study the Structure of Pharmaceuticals in Solution?

Although many structural techniques, most notably crystallography, have provided invaluable insight into the solid-state structure of molecules,[51, 52] these solid-state structures do not necessarily correspond to the structures adopted by molecules in solution.[8, 52] Furthermore, crystallographic structures yield limited information regarding molecular interactions in solution and how these interactions are influenced by solvation. Understanding the structure of pharmaceuticals in solution is thus crucial because of the relationship between molecular structure and biological function and because of the importance of interactions between pharmaceuticals and other biomolecules *in vivo*, where in the majority of cases pharmaceuticals experience an aqueous environment, in determining the effectiveness of a drug.[134, 154, 207, 209, 228, 247]

As proteins bind to ligands in specific conformations, the topographies of the complementary surfaces of the ligand and protein determine both affinity and specificity.[207, 304] Therefore, the structure and conformation of pharmaceuticals in physiologically relevant environments (i.e in solution) are important considerations in drug design.[175, 304] Fruitful design modifications of a lead structure are facilitated by an understanding of the bioactive conformation of the ligand as it binds to its target.[89] Structure-based drug design has already led to the development of a number of commercially available

drugs,[169] most notably human immunodeficiency virus (HIV) drugs.[27] Moreover, one of the goals in biotechnology is to transform the process of developing a drug from a trial and error empirical operation into a rational, structure-based process.[7, 89]

Most approaches to drug design focus solely on direct interactions between a drug and its corresponding target, not taking into account the effect of the physical milieu in which the drug exists *in vivo*.[154, 298] Such an approach neglects the enigmatic role played by water in biomolecular complexes, as water is able to increase the promiscuity of an interaction yet can also increase affinity and specificity.[154] Furthermore, water molecules have been shown to have a direct and crucial function in molecular recognition and catalysis,[134] and the ability to engineer water binding sites into the interface between a drug and its target is useful in drug design.[154] For example, the design of inhibitors of HIV protease and HIV replication rely on mimicking the hydrogen bonding features of a key structural water molecule.[156] Hydration experienced by pharmaceuticals also ultimately plays a role in determining whether they reach their target, the mechanism that drives a molecule from water to the protein active site being related to the hydration energy of the drug.[209]

Understanding the structure adopted by pharmaceuticals in solution is also important, because current predictive solubility models, which are mainly based on crystalline structures, are inadequate in the case of most functional drugs. This means that ∼40% of new molecular entities (NMEs) are insufficiently soluble to be delivered to patients.[244] Consequently, much effort has gone into developing experimental and computational methods that enable the solubility of drugs to be more accurately predicted.[20, 55] Understanding the interactions that take place in solution between pharmaceuticals and solvent molecules will inform such prediction methods. The importance of increasing the proportion of NMEs that are medically viable is highlighted by Fig. 1.1, which shows that, despite ever greater expenditure on research and development (R&D), the number of NMEs available per year has not consistently increased.

Additionally, understanding the structure of pharmaceuticals in solution is also crucial in determining how certain drugs are able to cross the blood-brain barrier (BBB), an amphiphilic membrane which serves to protect the brain from noxious agents but also

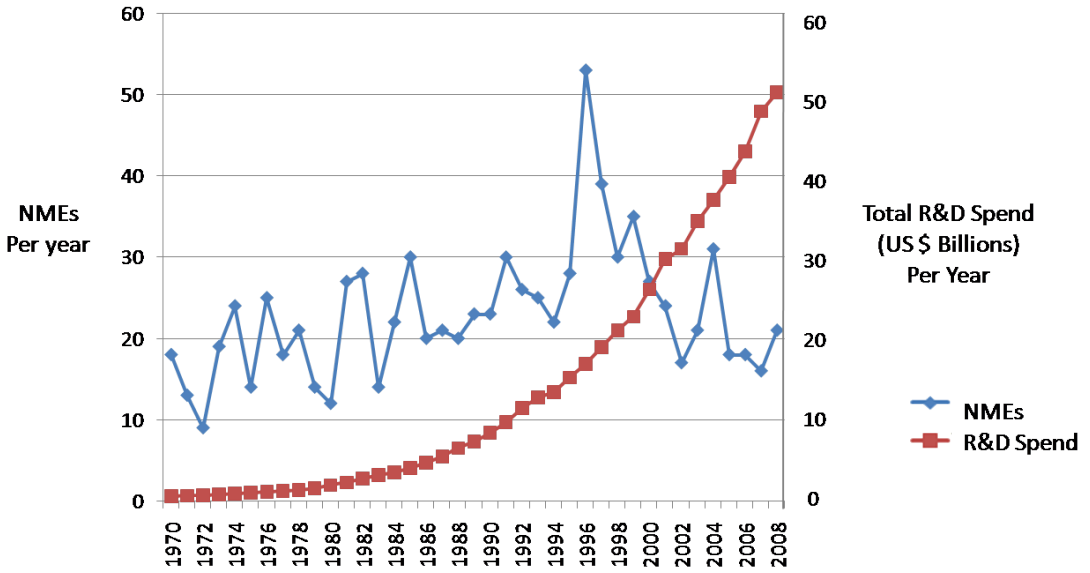


Figure 1.1: The number of new molecular entities (NMEs) available per year (blue) and research and development (R&D) expenditure per year (red) between 1970 and 2008. (Taken from [311].)

hinders delivery of therapeutics to the brain, as their ability to do so is also linked to structure and has ramifications for central nervous system (CNS) drug design. Factors linked to the ability of a molecule to cross the BBB include molecular size, the number of hydrogen bond donors and acceptors and lipophilicity. Examples of CNS disorders include Alzheimer's disease, Parkinson's disease and multiple sclerosis.[305] A detailed discussion of the BBB is offered in Section 1.3.

1.2 Studying the Atomic Structure of Solutions

1.2.1 Theory of Liquid Structure

Since the formal development of statistical mechanics, efforts have been made to provide a complete and consistent theory of liquids that may be validated by its ability to interpret and quantitatively represent their macroscopic properties.[66, 117] However, as the liquid state is transitory between the free atoms or molecules in random motion that constitute a gas and the solid state, where a high degree of spatial order may be assumed, it has traditionally proved difficult to apply classical concepts to define the physical properties of liquids.[29, 117] Moreover, theoretical study, even of the simplest

liquids, is difficult due to the complexity of the intermolecular potential, delaying the development of realistic and tractable potential functions.[124]

Early approaches adopted to describe the structure of liquids often employed simplified intermolecular forces, neglecting, for example, intermolecular attractive forces, approximated molecules as rigid spheres or postulated that too many hydrogen bonds connecting neighbouring molecules are broken.[124, 144, 219] Although successful for the very simplest systems, such as liquid argon, these approaches are too crude for more complex liquids.[124] More nuanced descriptions of the interactions between the molecules of liquids considered the different roles of attractive and repulsive forces in determining liquid structure.[302] Some models developed to describe the structure of liquids rely on the notion that a liquid has an underlying quasi-crystalline structure,[10, 144, 237] for example, a lattice structure blurred by thermal motion,[144] helping to surmount difficulties in dealing with strong molecular interactions.[10] However, the simplest variation of this approach, the so-called hole theory of liquids, which treats liquids as solids with certain atoms omitted, inadequately describes entropy.[23] Liquid structure has also been evaluated by treating the attractive potential as a perturbation on the rigid sphere potential.[13, 293] More complex theories necessary to describe the structure of mixtures of liquids rely on the principle that if two substances separately obey the law of their corresponding states, their mixture may also be expected to obey it as if it were a single-component fluid characterised by some suitable average parameters.[161]

Perhaps the most important tool for describing the structure of liquids, radial distribution functions (RDFs), relies on the assumption that at ordinary temperatures and pressures (i.e. not near the critical point) liquids are molecularly homogeneous.[22, 151] RDFs describe the distribution of one atomic species relative to another, technically the ratio of the average local density at a given distance from an arbitrary molecule to the average density of the fluid in bulk, and they can be used to understand the microscopic characteristics of a liquid.[143, 144, 145, 167, 192] For instance, the fluctuation theory of solutions relates the RDFs of a system to the quantities important in describing solvation thermodynamics of solutes in stable solutions.[142]

Computational advances have allowed implementation of some of the aforementioned

theoretical concepts in the form of computer models, which act as a bridge between theory and experimental results. The upshot of computer models is the capacity to simulate and analyse the behaviour of liquids and their structure.

1.2.2 Simulation Techniques

Simulation provides a powerful tool for understanding, not only liquid structure, but the structure of molecules, including pharmaceuticals, in solution. The two preeminent families of simulation techniques are molecular dynamics (MD) and Monte Carlo (MC), both of which were developed during the 1950s.[5, 103]

Classical MD consists of a numerical, step-by-step solution of Newton's equations of motion for a collection of particles. The interactions between the particles in the system are described by a forcefield (e.g. AMBER, CHARMM, GROMOS),[103, 183] which is an empirical approximation used to calculate interactions and evaluate the effective potential energy of the system as a function of particle coordinates.[164] A forcefield consists of both the set of equations used to calculate the potential energy and forces from particle coordinates, as well as a collection of parameters used in the equations.[164] Perhaps the principal advantage of MD is that the only input required by the model is a description of the interatomic/intermolecular interactions, with no assumptions made about the processes or mechanism to be investigated. In MD successive configurations of the simulation are connected in time, meaning that dynamical information can be measured. Nowadays, applications of MD extend to physics (e.g. properties of statistical ensembles, theory of liquids), chemistry (e.g. molecular structures, reactions), materials science (e.g. crystal defects, film growth) and biology (e.g. drug design, structure of membranes, protein folding, dynamics of large biomolecules).[103]

The main difference between MD and MC lies in the modes of sampling the configurational space available to the system. Unlike MD, where new configurations are generated by solving the classical equations of motion for all particles simultaneously over a small time step to determine the new atomic positions and velocities, MC generates new configurations by moving particles in accordance with the Metropolis algorithm along a Markov chain.[130] Particles are moved one at a time by a small displacement

in the simulation. This move is accepted or rejected based on an acceptance criterion, typically a decrease in the energy of the system, that guarantees that configurations are sampled in the simulation from a statistical mechanics ensemble distribution.[62] This method of sampling means that one drawback of MC is that the size of the molecule simulated can influence the configurations sampled. For instance, MC simulation of proteins in solution can prove difficult, because any move which significantly alters the internal coordinates of the protein without also moving the solvent particles will likely result in a large overlap of atoms and, thus, the rejection of the configuration.[62] Another disadvantage of MC compared with MD is that dynamic properties cannot be measured by MC, because MC is not time-dependent and each configuration depends only on its predecessor. However, MC simulations can be performed using fully flexible bonds without slowing down performance, unlike MD simulations, where hydrogen atoms often need to be constrained for the simulation to be computationally efficient.[287] A notable type of MC is reverse Monte Carlo (RMC), which, rather than rearranging an initial structure so that its energy is minimised, adjusts a structural model to minimise differences between simulated and experimental diffraction data.[135] The MC method has found applications across the physical sciences, as well as in finance and economics.[131, 179, 222]

Simulation techniques from each family are used in this thesis. Empirical Potential Structure Refinement (EPSR), a Monte Carlo-based method designed to model disordered materials, which is described in detail in Section 2.2, is used in conjunction with experimental data to undertake analysis of the systems studied at a microscopic level.[257, 259, 260] As an independent comparison to EPSR, for each system studied MD simulations were also run under the same conditions.

1.2.3 Experimental Techniques

The are a number of experimental techniques by which the structure of liquids and small molecules, such as pharmaceuticals, in solution has been studied. These techniques can be broadly categorised into two groups: those used to study macroscopic properties and those used to study the microscopic properties of liquids.[96] Typically, macroscopic properties, such as thermodynamic properties and transport coefficients,

are more easily determined experimentally than microscopic properties.[96] Techniques used to determine microscopic structure, the focus of this thesis, include various spectroscopic and scattering methods.

In Infrared (IR) spectroscopy, absorption of light in the IR region of the electromagnetic spectrum by molecules can be used to determine molecular structure,[67, 138, 188] as chemical bonds in different environments absorb light at varying frequencies. One important limitation of IR spectroscopy in the context of understanding the structure of molecules in physiologically relevant environments is that H_2O cannot be used as a solvent in IR spectroscopy because of its intense absorption in this frequency range, though use of D_2O may overcome this issue.[307] Raman spectroscopy measures the change in energy of photons scattered by a sample, the change in energy of the incident photon being related to the vibrational energy spacing of the molecule. Typically monochromatic light from a laser in the visible, near infrared (NIR) or near ultraviolet (NUV) range is used to determine the vibrational modes in the system from the energy shifts of the incident photons.[281, 312]

Nuclear Magnetic Resonance (NMR), also a spectroscopic technique, is commonly used to (indirectly) measure molecular structure and has become a standard technique across organic chemistry and biochemistry.[149, 285, 297] NMR exploits the existence of isotopes with non-zero nuclear spin (e.g. ^{1}H, ^{13}C, ^{31}P, ^{15}N, ^{19}F). A nucleus with spin I has 2I+1 associated magnetic moments aligned with an external magnetic field upon its application. For example, the magnetic moment of a nucleus with I=1/2, such as hydrogen, placed in an external magnetic field can be aligned either parallel with the field, the energetically preferred orientation, or antiparallel. The exact distribution of nuclei in a sample placed in an external magnetic field whose magnetic moments are either parallel or antiparallel with the field is governed by a Boltzmann distribution. Electromagnetic radiation, typically radiowaves, is used to excite nuclei from the lower energy state to the higher energy state, the difference in energy between the two orientations depending on the external magnetic field strength.[225] When this radiation ceases, nuclei relax back into the lower energy state in which their magnetic moments are parallel to the external magnetic field. The fluctuation of the magnetic field associated with this relaxation process is called resonance and can be detected and converted

into a spectrum *via* Fourier transformation. Peak positon, multiplicity, intensity and shape in the resulting spectra allow the chemical environments of atoms to be inferred, as these spectral features are sensitive to molecular structure.

In addition to the many spectroscopic techniques which can be used to study the atomic structure of molecules in solution, diffraction techniques can be used to measure interatomic structure. For liquids diffraction is one of the most important scattering techniques. Diffraction techniques enable direct structural measurement of molecules in solution by measuring the change in momentum of incident particles on a sample.

Neutron diffraction offers a number of advantages compared with more common techniques, most notably X-ray diffraction. Firstly, the wavelengths of neutrons generated to study matter are comparable to interatomic spacings, which, as neutrons interact with the atomic nucleus *via* the strong nuclear force, means that neutrons can be used to study atomic density.[223] In contrast, X-rays interact with the electron cloud, so that scattered photons in X-ray diffraction offer no information about atomic density but rather electronic density. The short range of the strong nuclear force with which neutrons interact ($\sim 10^{-15}$m) means that neutrons are highly penetrating.[252] The attenuation of a beam of low energy neutrons by aluminium, for example, is about 1% mm^{-1} compared with around 99% mm^{-1} for X-rays. The advantage of the highly penetrating nature of neutrons is that they interact weakly with matter, aiding interpretation of scattering data. Being highly penetrating also means that neutrons are non-destructive even to complex, delicate, biological materials.[223] The highly penetrating nature of neutrons is a boon in the study of liquids in particular, because neutrons are able to penetrate the container in which liquids must necessarily be measured. Unlike X-ray scattering, neutron scattering is isotopically sensitive. Therefore, isotopic substitution methods can be exploited as a powerful tool to yield structural and dynamical information in even greater detail. As neutrons have a magnetic moment, they can also be used to study the magnetic properties of materials.

The main disadvantage of neutron diffraction is the inherently low intensities of neutron sources. X-ray instruments at synchotron radiation facilities can provide fluxes of 10^{18} photons s^{-1} mm^{-2}, compared with 10^4 neutrons s^{-1} mm^{-2} in the same energy

bandwidth for powerful neutron scattering instruments.[223] Furthermore, although the highly penetrating nature of neutrons means that they are non-destructive and can be used to measure bulk properties, large samples are required to compensate for low scattering intensities.[223, 252] There are also kinematic restrictions, as not all energy and momentum transfers can be measured.

Principally neutron diffraction, complemented by NMR where appropriate, is used to study the atomic structure of pharmaceuticals in solution in this thesis. As mentioned in Section 1.2.2, by also harnessing the power of computer simulation techniques (EPSR and MD), atomic-level details can be determined.

1.3 Molecules Able to Cross the Blood-brain Barrier

A key theme of this thesis, as alluded to in Section 1.1, is the study of drugs which are able to cross the blood-brain barrier (BBB). Of interest are both the mechanisms by which drugs are able to cross the BBB and the common features that these drugs might share in terms of their structure in solution and how their interaction with solvent molecules may govern their interactions with the BBB.

The BBB (Fig. 1.2) is a highly specialised structure, formed of endothelial cells joined by tight junctions, which separates components of the circulating blood from neurons.[11, 318] It acts to protect the CNS from harmful chemicals and helps to maintain the chemical composition of the neuronal milieu.[318] Although successful in protecting the brain from noxious agents, the BBB also significantly hinders the delivery of therapeutics to the brain.[38] In fact, 98% of drugs fail at clinical trials due to inadequate BBB permeability.[85, 205] This low success rate has prompted the development of both *in vitro* and computational models to better assess the permeability of the BBB to drugs. Strategies for delivering drugs across the BBB encompass a plethora of both invasive and non-invasive techniques.[305] For example, functionalised nanoparticles have recently gained attention as an attractive carrier for brain targeting.[305] Understanding how substances are able to cross the BBB has important medical implications. A first step in this process is understanding the structure adopted by drugs in solution at the

BBB.

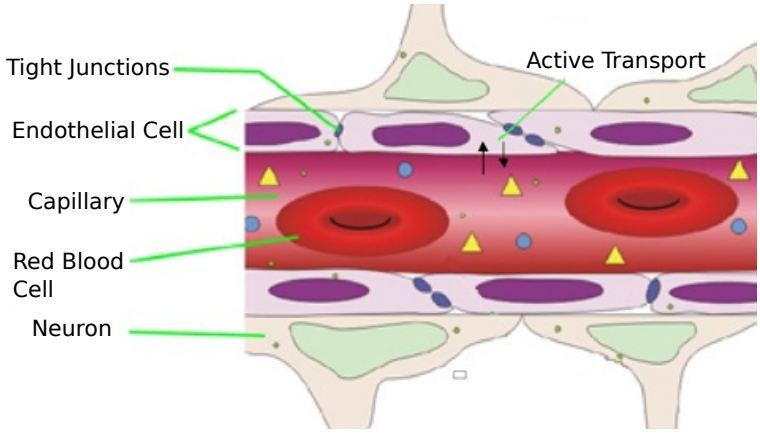


Figure 1.2: A simplified diagram of the human blood-brain barrier. The thickness of the endothelial cell is ∼200 nm. (Adapted from [310].)

Currently there is no definitive consensus regarding how substances cross the BBB. Previously, the BBB was considered a passive barrier.[72, 193] However, more recent studies consider the BBB a dynamic interface,[41] with substances transported by a number of mechanisms. For instance, small, hydrophobic molecules, such as O_2 and CO_2, cross the BBB by transcellular diffusion through the lipid bilayer. In contrast, nutrients, such as sugars, essential to brain function, cross the BBB by transporters which allow selective uptake.[136, 137]

Diffusion across the BBB favours drugs of low molecular weight (up to 400 Da)[93] and high lipid solublity.[12] However, on crossing the BBB, the drug must then partition into the aqueous environment of the interstitial fluid of the brain to exert a pharmacological effect.[12] Consequently, a substance that is highly lipid soluble can be sequestered by the capillary bed and not reach the cells beyond the BBB. Lipid solubility also favours uptake by the peripheral tissues, which can lower the concentration of the drug in blood. Thus, while lipid solubility can increase transport rate across the BBB, it can also lower the amount of the drug presented to the brain.[12] Being lipophilic and hydrophilic (amphiphilic) means that substances are able to cross the BBB and partition into the aqueous environment of the interstitial fluid of the brain.[74]

The most commonly used metric to quantify permeability across the BBB is $\log BB$, which is defined as the ratio between the concentration of a compound in brain, c_{brain},

and in the blood, c_{blood}:

$$\log \text{BB} = \log(\frac{c_{\text{brain}}}{c_{\text{blood}}}). \tag{1.1}$$

Unfortunately, measurement of $\log \text{BB}$ is both time consuming and expensive.[171, 187] Experimentally, $\log \text{BB}$ is usually determined by measurement of the concentration of a drug in the brain and blood of rats after its administration.[294] However, $\log \text{BB}$ is rather a crude measure, as it does not differentiate between intracellular and extracellular concentrations.[88] Molecular size, lipophilicity, the solvent accessible surface area or number of groups that can establish potential hydrogen bonds are all putative predictors of $\log \text{BB}$.[88, 201, 294] For most prescribed CNS drugs $\frac{c_{\text{brain}}}{c_{\text{blood}}}$ =0.3.[58]

Similarly, $\log \text{P}$, where P is the partition coefficient expressing the ratio of the concentration of a solute in octanol, c_{octanol}, to the concentration of a solute in water, c_{water}, is often used to quantify lipophilicity:

$$\log \text{P} = \log(\frac{c_{\text{octanol}}}{c_{\text{water}}}). \tag{1.2}$$

However, being a good hydrogen donor, octanol is probably not a typical apolar solvent.[187] In addition, besides the previously discussed issues relating to lipid solubility and the amount of a drug presented to the brain, $\log \text{P}$ is only defined for a compound in its neutral state.[1, 187]

Perhaps the main limitation of both $\log \text{BB}$ and $\log \text{P}$ is that they arise from multiple mechanisms and consequently do not offer medicinal chemists specific guidance regarding the structural modifications which must be made to a drug in order to improve its ability to cross the BBB. The multitude of variables thought to influence the metrics used to quantify BBB penetration and their limitations provide motivation for investigation of the mechanisms by which molecules cross the BBB and a microscopic assessment of the behaviour of such compounds in physiologically relevant environments.

1.4 Aims of Thesis

The main aim of this thesis aims to increase understanding of the atomic structure of four small molecules in solution, indole, cocaine hydrochloride, alprazolam and clonidine hydrochloride, and how their structure in solution relates to their biological function. Reasons why each of these molecules, in particular, is of interest are outlined in Sections 3.1, 4.1, 5.1 and 6.1. However, there are a some overarching aims related to the study of the structure of these molecules in solution.

This thesis aims to **compare how different hydrophobic and hydrophilic groups of these molecules interact with the solvent molecules in solution** on a microscopic level using a combination of experimental, neutron and X-ray diffraction complemented by NMR, and computational techniques. A comparison between the interaction of hydrophobic and hydrophilic groups of indole, cocaine hydrochloride, alprazolam and clonidine hydrochloride with solvent molecules includes making an assessment of which groups are most/least solvated and the nature of such interactions. Moreover, an aim of this thesis is to form a picture of how the different constituents of a molecule together contribute to its interactions with its environment.

Another common objective of this thesis is to **gain understanding of the mechanisms by which cocaine hydrochloride, alprazolam and clonidine hydrochloride cross the BBB**. Explicitly, does the conformation of molecules in solution and their solvation potentially aid explanation of their function and ability to cross the BBB? As referred to in Section 1.1, the relationship between the environment and the structure adopted by molecules in solution is a central theme of this research.

Additionally, it is hoped that the research presented in theis can **augment what is known about these molecules from crystallographic studies**. For example, are waters seen in the crystal structures of these molecules consistent with the solvation implied by the techniques used here?

An important ancillary aim of this thesis is to **develop computational methods of characterisation of the solvation of molecules in solution so as to address which parts of the molecules are preferentially solvated**. It should be noted that,

although different computational techniques are used in this thesis, it is not the aim of this thesis to evaluate the merits of different computational techniques in determining the atomic structure of molecules in solution. However, inevitably, the results obtained by the different computational techniques used are compared and contrasted throughout.

1.5 Structure and Scope of Thesis

Chapter 2 of this thesis (*Theoretical Background*) begins with a discussion of neutron diffraction (Section 2.1), the key experimental technique employed to determine the atomic structure of indole, cocaine, alprazolam and clonidine in solution. This Section includes an outline of the fundamentals of neutron diffraction theory (Section 2.1.1), elucidation of the principles and design considerations behind the Small Angle Neutron Diffractometer for Amorphous and Liquid Samples (Section 2.1.2), with which all neutron diffraction data in this thesis were obtained, and discussion of subsequent data correction (Section 2.1.3).

The proceeding section in Chapter 2 (Section 2.2) focuses on Empirical Potential Structure Refinement (EPSR), a computational method which fits simulated diffraction data to experimental diffraction data in order that comprehensive atomic level detail about the studied system may be extracted. In this thesis, for each system studied Molecular Dynamics (MD) has been used as an independent comparison to EPSR. Due to the ubiquity of MD in condensed matter physics[5, 103, 130, 183] and its secondary use to neutron diffraction in conjunction with EPSR in this thesis, a full discussion of the theory behind MD is not included.

Chapter 2 finishes by outlining the principles behind ANGULA (Section 2.3), a program designed for detailed analysis of small molecules and liquids, and the means by which ANGULA allowed analysis of both EPSR and MD simulations to be undertaken.

The remainder of this thesis is divided into four largely self contained chapters, comprising of introduction, methods, results and conclusion sections. These Chapters correspond to investigations of indole (Chapter 3), cocaine hydrochloride (Chapter 4), alprazolam (Chapter 5) and clonidine hydrochloride (Chapter 6) in solution.

The final chapter of this thesis (Chapter 7) offers a summary of the findings of the research presented and describes trends seen across the different systems studies. Speculative implications of the findings made in this thesis for drug design are discussed. Suggestions are also made for further work that could be undertaken to extend and enhance this thesis.

Chapter 2

Theoretical Background

2.1 Neutron Diffraction

2.1.1 Fundamentals

A particle is scattered when there is a change in its momentum caused by one or more localised non-uniformities in its path (e.g. other particles, density fluctuations and bubbles). Scattering is considered elastic when the kinetic energy of the incident particle is conserved, resulting in only a change in the direction of propagation of the particle. Conversely, inelastic scattering, such as Raman scattering, is when the incident particle either gains or loses kinetic energy. The phenomena of elastic and inelastic scattering can each be further classified as either coherent or incoherent scattering. Coherent scattering means that outgoing particles have a phase relationship with one another, whereas incoherently scattered particles have no phase relationship. Diffraction is elastic, coherent scattering. By measuring the diffraction of particles incident on a sample, structural information can be gleaned. In this Section, the general underpinnings of diffraction are adumbrated and discussed with regards to neutron diffraction in particular.[a] Special emphasis is given to amorphous materials (e.g. glasses and liquids). This Section outlines the quantities measured during a diffraction experiment and provides a

[a]The mathematical formalism used to describe diffraction in this Section is based on *Elementary Scattering Theory* by D.S. Sivia.[252] As with any mathematical treatment of a physical phenomenon, other approaches are possible.[273]

mathematical description of how such quantities are related to the structure of a sample.

The momentum transfer, $\mathbf{P}$, experienced by an incident particle upon collision with a localised non-uniformity, such as another particle, can be expressed as

$$\mathbf{P} = \hbar\mathbf{k}_\mathrm{i} - \hbar\mathbf{k}_\mathrm{f} = \hbar\mathbf{Q}, \tag{2.1}$$

where $\mathbf{k}_\mathrm{i}$ and $\mathbf{k}_\mathrm{f}$ are the initial and final wave vectors, respectively, and $\hbar$ is the reduced Planck constant. $\mathbf{Q}$ is often used to express the difference between the initial and final wave vectors:

$$\mathbf{Q} = \mathbf{k}_\mathrm{i} - \mathbf{k}_\mathrm{f}. \tag{2.2}$$

Similarly, energy transfer, E, may be written as

$$E = \hbar(\omega_\mathrm{i} - \omega_\mathrm{f}) = \hbar\omega, \tag{2.3}$$

where ω_i, ω_f and ω are the initial, final and change in radial frequency of the scattered particle, respectively. The momentum and energy transferred are encoded in a function, $F_\mathrm{tot}(\mathbf{Q}, \omega)$, termed the scattering law or total dynamic structure factor. For diffraction, where, by definition, scattered particles experience no change in energy

$$F_\mathrm{tot}(\mathbf{Q}, \omega = 0) = F_\mathrm{tot}(\mathbf{Q}). \tag{2.4}$$

$F_\mathrm{tot}(\mathbf{Q})$ is termed the total static structure factor and is related to the total dynamic structure factor by[63, 73]

$$F_\mathrm{tot}(\mathbf{Q}) = \int_{-\infty}^{+\infty} F_\mathrm{tot}(\mathbf{Q}, \omega)\, d(\hbar\omega). \tag{2.5}$$

Fig. 2.1 shows an incident particle being scattered through an angle of 2θ upon collision with another particle. The direction of momentum transfer (Eq. 2.2) is also shown as $\mathbf{Q}$. From this triangle a relationship can be written relating $\mathbf{k}_\mathrm{i}$ and $\mathbf{k}_\mathrm{f}$, whose magnitudes are equal for an elastically scattered particle, to the wavelength, λ, and θ:

$$|\mathbf{k}_\mathrm{i}| = |\mathbf{k}_\mathrm{f}| = \frac{|\mathbf{Q}|}{2\sin\theta} = \frac{2\pi}{\lambda}. \tag{2.6}$$

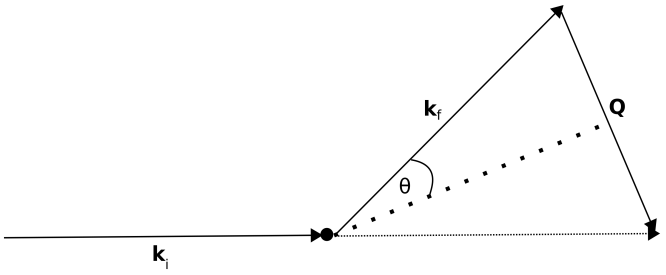


Figure 2.1: An incident particle, wave vector $\mathbf{k}_\mathrm{i}$, being elastically scattered through an angle of 2θ following a collision with another particle. The outgoing particle has a wave vector $\mathbf{k}_\mathrm{f}$.

Simple rearrangement of Eq. 2.6 yields

$$|\mathbf{Q}| = \frac{4\pi \sin\theta}{\lambda}, \tag{2.7}$$

a variation of Bragg's Law, which describes constructive scattering by a crystal structure.

Having set out some of the basic formalism for describing diffraction, it is necessary to consider the experimental quantities measured. Foremost among these is the differential cross section, $\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}$, which is defined as the ratio of the number of particles deflected by $(2\theta, \phi)$, where ϕ describes the rotation of $\mathbf{k}_\mathrm{f}$ about the plane of $\mathbf{k}_\mathrm{i}$ (azimuthal angle), per unit time, per unit solid angle, to the incident flux of particles, Φ. If the rate of arrival of particles at a detector, which subtends a solid angle of $\Delta\Omega$, scattered in the direction $(2\theta, \phi)$ by a sample is denoted as R,

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \frac{R}{\Delta\Omega\,\Phi N}. \tag{2.8}$$

The variable N arises from normalisation to the number of atoms in the sample. Eq. 2.8 corresponds to a measurement of the number of particles scattered into the solid angle $\Delta\Omega$ in the direction $(2\theta, \phi)$. If the same measurement were made as a function of the energy of the scattered particles, E_f, then this quantity is defined by the partial differential cross section, $\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega \mathrm{d}E_\mathrm{f}}$, which can be related to the differential cross section by

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} = \int_0^\infty \left(\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega \mathrm{d}E_\mathrm{f}}\right) \mathrm{d}E_\mathrm{f}. \tag{2.9}$$

The implication of Eq. 2.9 is that measurement over a range of energies is equivalent

to measurement of the differential cross section (Eq. 2.8), so that the number of particles scattered into $\Delta\Omega$ in the direction $(2\theta, \phi)$ can be found when the incident flux is unknown.

Both elastic and inelastic scattering are encompassed by Eq. 2.8. Diffraction measurements do not involve energy discrimination, but the total scattering can usually be approximated as elastic:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \approx \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{el}} = \frac{R_{el}}{\Delta\Omega\,\Phi N}. \tag{2.10}$$

Inelastic neutron scattering is often minimised through experimental design, such as by performing low temperature measurements or by positioning detectors at low angles (Section 2.1.2). However, attempts are usually made to correct for this discrepancy in the analysis of neutron diffraction measurements of liquids (Section 2.1.3.2.5), where the large number of low energy translational and rotational modes, which are excited by the neutron scattering process, exacerbate inelasticity effects. Furthermore, inelastic neutron scattering is particularly prominent for samples containing light atoms, where, due to mass equivalence, large amounts of energy can be exchanged in the scattering process.[263]

The cross section, σ, expresses the likelihood of an interaction between the incident particle and the sample in the form of an effective area. It can be related to the scattering rate through multiplication by the incident flux ($R = \sigma\Phi$). Cross sections are often quoted in barns (1 barn = $10^{-28}\mathrm{m}^2$). To understand the theoretical origin of the cross section, it is expedient to consider a beam of particles, ψ_{i}, traveling in the $\mathbf{z}$ direction, with wavelength λ, incident on a single fixed atom. Such a beam can be described by the complex plane wave:

$$\psi_{\mathrm{i}} = \psi_0 e^{\mathrm{i}kz}, \tag{2.11}$$

where k is the wavevector of the incident beam and i is the imaginary unit. ψ_0 subsumes temporal variation and phase offset. $\Phi = |\psi_0|^2$. If the particles in the beam are elastically scattered, $|\mathbf{k}_i| = |\mathbf{k}_f|$. After interacting with the atom, the particles move radially outwards. If the origin is centered on the atom, so that $\mathbf{k}_f$ is parallel to the displacement

vector of a scattered particle, $\mathbf{r}$, then the sinusoidal part of the scattered wave is

$$e^{\mathrm{i}\mathbf{k}_f \cdot \mathbf{r}} = e^{\mathrm{i}kr}, \tag{2.12}$$

where $r = |\mathbf{r}|$. The number of outgoing particles per unit area will fall with distance r, in accordance with the inverse square law. Consequently, the resulting scattered wave will take the form

$$\psi_f = \psi_0 f(\lambda, \theta) \frac{e^{\mathrm{i}kr}}{r} \tag{2.13}$$

for $r > 0$. In Eq. 2.13, $f(\lambda, \theta)$ is a probability weighting, indicating the likelihood of a particle of wavelength λ being deflected in a given direction. This function does not depend on ϕ because the isotropic nature of the scattering (*vide infra*) means that it is symmetric. ψ_0 ensures that the rate of scattering is proportional to the incident flux.

The characteristics of $f(\lambda, \theta)$ depend on the nature of the incoming particles. For neutrons $f(\lambda, \theta)$ is usually invariant with both λ and θ and equals the negative of what is termed the scattering length, b:

$$f(\lambda, \theta) = -\mathrm{b}. \tag{2.14}$$

The magnitude of b is related to the strength of the scattering of the neutrons by the nucleus, whereas the sign of b indicates whether the neutrons are scattered in phase or π out of phase.[b] The isotropic nature of nuclear scattering results from the short range of the strong nuclear force (10^{-15}m) compared with the wavelength of thermal neutrons used in scattering experiments (10^{-10}m). Eq. 2.14 describes neutron scattering only by the nucleus of the atom.

The neutron scattering length of an element is determined by the interaction of the incident neutron with the nuclei of its atoms, depending on the constituents of the nuclei and the orientation of their spins. The magnitude of the coherent neutron scattering length of an atom is, in fact, the weighted average of the scattering lengths of its nuclear spin states. The associated standard deviation is the incoherent neutron scattering length. Nuclei with a spin of 0 have a incoherent scattering length of 0 fm. This definition

[b] Although in reality complex, the imaginary part of the scattering length is normally negligible.[248]

Element/Nuclide	Z	I	Coherent (fm)	Incoherent (fm)
^{1}H	1	1/2(+)	-3.7406	25.2479
^{2}H	1	1(+)	6.6710	4.0430
^{12}C	6	0(+)	6.6511	0.0000
^{13}C	6	1/2(-)	6.1990	-0.5290
^{14}N	7	1(+)	9.3720	2.0200
^{16}O	8	0(+)	5.8030	0.0000

Table 2.1: Bound coherent and incoherent neutron scattering lengths for some common elements and nuclides.[248] Z and I denote the atomic number and spin of the nuclear ground state, respectively. (+) and (-) indicate triplet and singlet spin states, respectively.

also means that neutron scattering lengths are isotope dependent. Table 2.1 shows the values of the scattering lengths for some common elements and nuclides. The sign of the scattering length is notably designated negative for hydrogen and positive for deuterium. The difference in sign between scattering lengths of isotopes can be exploited by isotopic substitution methods in order to be able to discriminate between different interatomic interactions.[71, 268]

In the case of X-rays, $f(\lambda, \theta)$ is a real-valued function referred to as a form factor. Form factors have the same sign for all elements and the magnitudes of X-ray form factors are proportional to the atomic number of the element. The X-ray form factor decreases with increasing θ and λ and is not dependent on spin.

Returning to the theoretical origin of σ, the number of particles of wavelength λ deflected through a given scattering angle $(2\theta, \phi)$, per unit area, per unit time, is given by the modulus squared of Eq. 2.13. Hence, remembering that $\Phi = |\psi_0|^2$, $R = \sigma\ \Phi$ and recognising that $\mathrm{d}A = r^2 \sin 2\theta \mathrm{d}\phi \mathrm{d}2\theta$,

$$\sigma = 2\pi \int_{2\theta=0}^{\pi} |f(\lambda, \theta)|^2 \sin 2\theta \mathrm{d}2\theta. \tag{2.15}$$

Applying Eq. 2.14 means that Eq. 2.15 is reduced to

$$\sigma = 4\pi \mathrm{b}^2, \tag{2.16}$$

providing an expression for the neutron scattering cross section. The total neutron

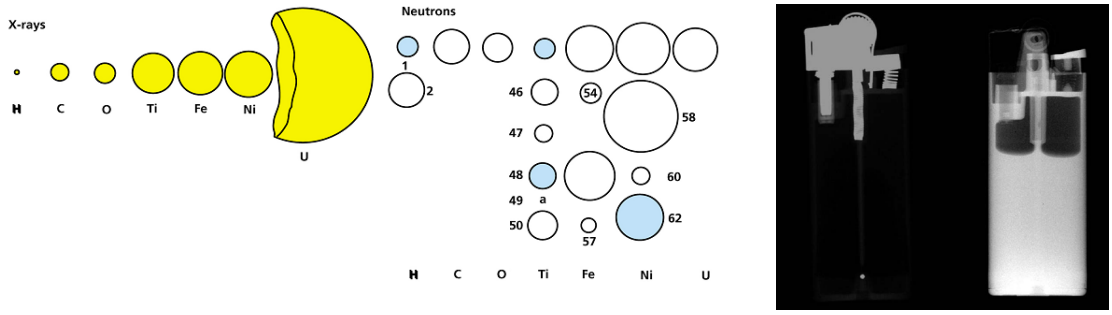


Figure 2.2: A pictorial representation of the scattering cross sections of different elements and nuclides for X-ray and neutron diffraction taken from [17]. For neutrons blue shading represents a negative scattering length. The numbers are the atomic masses of the elements/nuclides. A cigarette lighter measured with X-rays (left) and neutrons (right) is also shown. (Taken from [87].)

scattering cross section can be expressed as the sum of the coherent scattering cross section and the incoherent scattering cross section:

$$\sigma = \sigma_{\text{coherent}} + \sigma_{\text{incoherent}}. \tag{2.17}$$

Eq. 2.17 is possible because, as discussed earlier, the neutron scattering lengths, b, of isotopes whose nuclei have non-zero spin are specified by an average value $\langle \text{b} \rangle$ and a standard deviation (Δb):

$$\sigma_{\text{coherent}} = 4\pi \langle \text{b} \rangle^2 \qquad \sigma_{\text{incoherent}} = 4\pi (\Delta \text{b})^2. \tag{2.18}$$

Coherently scattered neutrons carry information on the correlations of one atom with respect to its neighbours in the sample.[120] As there is no constructive interference in incoherent scattering, incoherently scattered neutrons only carry information about individual atoms. Incoherent scattering can be considered a uniform background (Section 2.1.3.2.4). Fig. 2.2 illustrates how the neutron scattering cross section does not increase with atomic number as the X-ray scattering does. The difference in the sensitivity of the two techniques to different atoms is underlined by an image of a cigarette lighter, whose main body is composed of hydrocarbons, measured by X-ray and neutron diffraction (Fig. 2.2).

The idealised scenario leading to Eq. 2.16 neglects the contribution of absorption of incident neutrons, meaning that the total neutron cross section, σ_{total}, is actually the

sum of the neutron scattering cross section and the neutron absorption cross section, $\sigma_{\text{absorption}}$:

$$\sigma_{\text{total}} = \sigma + \sigma_{\text{absorption}}. \tag{2.19}$$

To form an expression for scattering from an assembly of atoms, however, Eq. 2.13 needs to be adapted. If the incident beam is represented by a complex plane wave with wavevector $\mathbf{k_i} = (0, 0, k)$, a particular atom, labelled by the index j, will make a small contribution to the scattered wave:

$$\delta[\psi_{\text{f}}]_{\text{j}} = \psi_0 f_{\text{j}}(\lambda, \theta) \frac{e^{\text{i}[\mathbf{r_j}(\mathbf{k_i} - \mathbf{k_f}) + \mathbf{r}\cdot\mathbf{k_f}]}}{|\mathbf{r} - \mathbf{r_j}|}, \tag{2.20}$$

where $f_{\text{j}}(\lambda, \theta)$ defines the relevant interaction characteristics of the j^{th} atom, located at $\mathbf{r_j}$, $\mathbf{k_f}$ is the wavevector of the outgoing particle and $\mathbf{r}$ is an arbitrary position from the origin (Fig. 2.3). This expression reduces to Eq. 2.13 when $\mathbf{r_j}$ is set to 0. Summing the contribution from all N atoms in the sample and using the substitution $\mathbf{Q} = \mathbf{k}_{\text{i}} - \mathbf{k}_{\text{f}}$ (Eq. 2.2),

$$\psi_{\text{f}} = \psi_0 e^{\text{i}\mathbf{k}_{\text{f}}\cdot\mathbf{r}} \sum_{\text{j}=1}^{N} f_{\text{j}}(\lambda, \theta) \frac{e^{\text{i}\mathbf{Q}\cdot r_{\text{j}}}}{|\mathbf{r} - \mathbf{r_j}|}. \tag{2.21}$$

This application of the superposition principle relies on the assumption that the scat-

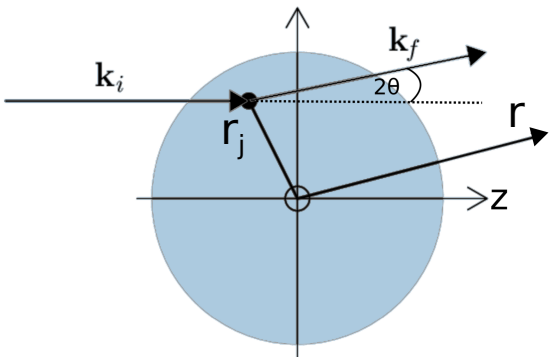


Figure 2.3: The contribution of the j^{th} atom, located at $\mathbf{r_j}$ relative to the origin, of a sample to the scattered wave.

tering process is weak. This simplification is known as the Born approximation, which states that the scattered waves have negligible effect on the incident beam. The interaction between incident neutrons and the nuclei of a sample is typically characterised

by the Fermi pseudo potential:[273]

$$V(\mathbf{r}) = \frac{2\pi\hbar^2}{\mathrm{m}} \langle \mathrm{b_j} \rangle \, \delta(\mathbf{r} - \mathbf{r_j}), \tag{2.22}$$

where m is the mass of a neutron, $\langle \mathrm{b_j} \rangle$ is the average scattering length of the $\mathrm{j^{th}}$ atom and $\delta(\mathbf{r})$ is the Dirac delta function. A quantum mechanical treatment of scattering uses the Fermi pseudo potential and Fermi's golden rule, the quantum equivalent of the Born approximation.[273] Multiple scattering effects have been ignored in the derivation of Eq. 2.21 because of the assumption of a weak scattering process. It is assumed that the length scale of the scattering nuclei is negligible with respect to the separation between the scattering sources in space. Given that the sample size is much smaller than the distance to the detector the approximation

$$|\mathbf{r} - \mathbf{r_j}| \sim |\mathbf{r}| = r \tag{2.23}$$

can also be made. Consequently, the modulus squared of Eq. 2.21 can be written as

$$|\psi_\mathrm{f}|^2 = \frac{\Phi}{r^2} \left| \sum_{\mathrm{j}=1}^{N} f_\mathrm{j}(\lambda, \theta) e^{\mathrm{i}\mathbf{Q}\cdot r_\mathrm{j}} \right|^2 . \tag{2.24}$$

Particles of wavelength λ are deflected elastically into a detector with a small area δA, in the direction $(2\theta, \phi)$, at a rate, R_el, given by

$$R_\mathrm{el}(2\theta, \phi) = |\psi_\mathrm{f}|^2 \delta A = \Phi \delta\Omega \left| \sum_{\mathrm{j}=1}^{N} f_\mathrm{j}(\lambda, \theta) e^{\mathrm{i}\mathbf{Q}\cdot r_\mathrm{j}} \right|^2 , \tag{2.25}$$

where $\delta\Omega = \delta A / r^2$ is the solid angle subtended by the detector at the sample. Revisiting the definition of the differential cross section given by Eq. 2.10, an expression is obtained which relates the differential cross section, the quantity measured in a scattering experiment, of an assembly of N atoms to $f_\mathrm{j}(\lambda, \theta)$:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega} \right)_\mathrm{el} \propto \left| \sum_{\mathrm{j}=1}^{N} f_\mathrm{j}(\lambda, \theta) e^{\mathrm{i}\mathbf{Q}\cdot r_\mathrm{j}} \right|^2 . \tag{2.26}$$

This relationship is expressed as a proportionality, because it is unnormalised.

The scattering characteristics of an object can be described by a scattering length density (SLD) function $\beta(x, y, z)$ [area^{-1}] of the scattering material and an infinitesimally small volume, dV, which encapsulates a point source of isotropic scattering, such as another particle, at a given location, $\mathbf{r}$:

$$f(\lambda, \theta) = \beta(\mathbf{r})\mathrm{d}V. \tag{2.27}$$

Substituting this elemental form of $f(\lambda, \theta)$ into Eq. 2.26 and summing over the x, y and z arguments of $\beta(\mathbf{r})$ leads to the continuum generalisation of Eq. 2.26:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{el}} \propto \left| \iiint_V \beta(\mathbf{r}) e^{\mathrm{i}\mathbf{Q}\cdot r} \mathrm{d}^3\mathbf{r} \right|^2 \propto F_{\mathrm{tot}}(\mathbf{Q}), \tag{2.28}$$

where V is the total volume within which scattering occurs. Therefore, the elastic differential cross section is related to the structure of the sample through a Fourier transformation (FT) of its SLD function. The connection between the SLD distribution of a given atom centered at the origin, $\beta_{\mathrm{j}}(\mathbf{r})$, and the corresponding function $f_{\mathrm{j}}(\lambda, \theta)$ can be seen when $\beta(\mathbf{r})$ is expressed as the sum of N discrete contributions:

$$\beta(\mathbf{r}) = \sum_{\mathrm{j}=1}^{N} \beta_{\mathrm{j}}(\mathbf{r} - \mathbf{r}_{\mathrm{j}}), \tag{2.29}$$

where $\mathbf{r_j}$ is the location of the j$^{\mathrm{th}}$ atom. Substitution of Eq. 2.29 into Eq. 2.28 implies that

$$\iiint \beta_j(\mathbf{r} - \mathbf{r_j}) e^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r} = e^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r_j}} \iiint \beta_j(\mathbf{r}) e^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r}. \tag{2.30}$$

Thus, comparing Eq. 2.30 with Eq. 2.26 and Eq. 2.28 leads to an expression for the relationship between the form factor and the SLD function of an individual atom:

$$f_j(\lambda, \theta) = \iiint \beta_j(\mathbf{r}) e^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r}. \tag{2.31}$$

By taking the auto correlation function (ACF) and FT of Eq. 2.29 and by exploiting Eq.s 2.28, 2.30 and 2.31 (Appendix A), the differential cross section can be expressed

as the sum of two terms:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{el}} \propto \overbrace{\sum_{\mathrm{j}=1}^{N} |f_{\mathrm{j}}(Q)|^2}^{\text{Self Scattering}} + \underbrace{F(Q)}_{\text{Distinct Scattering}} , \tag{2.32}$$

where

$$F(Q) = 2\sum_{\mathrm{j>k}} Re\{f_{\mathrm{j}}(\mathbf{Q}) f_{\mathrm{k}}(\mathbf{Q})^* e^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{r}_{\mathrm{j}}-\mathbf{r}_{\mathrm{k}})}\}. \tag{2.33}$$

The first term in Eq. 2.32 depends only on the scattering properties of individual atoms, not their position, and is referred to as the self scattering term. $F(Q)$, which represents the correlation between the j^{th} and k^{th} atoms, contains the basic structural information on atomic positions, the information sought in a diffraction experiment.[269] $F(Q)$ is usually referred to as the distinct scattering term. Both terms can be expressed as a function of Q=$|\mathbf{Q}|$, because, as molecules in amorphous materials are found in all orientations, they are assumed spherically symmetric. In reality, the differential cross section contains a third term that corresponds to inelastic scattering:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right) \propto \overbrace{\sum_{\mathrm{j}=1}^{N} |f_{\mathrm{j}}(Q)|^2}^{\text{Self Scattering}} + \underbrace{F(Q)}_{\text{Distinct Scattering}} + \overbrace{P(Q,\omega)}^{\text{Inelastic Scattering}} . \tag{2.34}$$

As discussed earlier, this term is the subject of corrections in the analysis of neutron diffraction data (Section 2.1.3.2.5). Being due to inelastic scattering (i.e. energy transfer), $P(Q,\omega)$, is dependent on energy as well as Q.

Usually, Eq. 2.32 is normalised by the number of atoms in the sample. By also incorporating normalisation factors corresponding to the average value of the SLD, $\langle\beta\rangle$, and the illuminated volume, V, and by expressing the Fourier integral in its one dimensional form,[c] Eq. 2.32 can be expressed as

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{el}} = \frac{1}{N}\sum_{\mathrm{j}=1}^{N} |f_{\mathrm{j}}(Q)|^2 + \frac{4\pi \ \langle\beta\rangle^2 V}{N} \int_0^R r^2 G(r) \frac{\sin(Qr)}{Qr} \mathrm{d}r, \tag{2.35}$$

[c] By using spherical polar coordinates, it can be shown that $\mathscr{F}[f(\mathbf{r})] = \iiint f(\mathbf{r}) e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r} = 4\pi \int_0^R r^2 f(r) \frac{\sin(kr)}{kr} \mathrm{d}r$

where $G(r)$, a pair distribution function, is the normalised real space counterpart of $F(Q)$ (Appendix A). The first term on the right hand side is the the average value of the modulus squared of the form factor, ($\left\langle |f(Q)|^2 \right\rangle$). The second term on the right hand side can be more readily assessed by calculating the the FT of $G(r) - 1$ instead of $G(r)$:

$$\int_0^R r^2[G(r) - 1]\frac{\sin(Qr)}{Qr}\mathrm{d}r = \int_0^R r^2 G(r)\frac{\sin(Qr)}{Qr}\mathrm{d}r + \Delta(Q). \tag{2.36}$$

The discrepancy, $\Delta(Q)$, the FT of a constant, is peaked sharply around the origin. Since the region around Q = 0 is not accessible experimentally, as it coincides with particles that have not been scattered by the sample, and because the upper limit of the integral on the right hand side can be replaced with infinity as $G(r) - 1 \to 0$ for large r,

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{el}} = \left\langle |f(Q)|^2 \right\rangle F_{\mathrm{tot}}(Q). \tag{2.37}$$

The total static structure factor, $F_{\mathrm{tot}}(Q)$, is given by

$$F_{\mathrm{tot}}(Q) = 1 + \frac{4\pi\ \alpha(Q)}{Q}\int_0^\infty r[G(r) - 1]\sin(Qr)\mathrm{d}r \tag{2.38}$$

for $Q \neq 0$, where

$$\alpha(Q) = \frac{\langle\beta\rangle^2 V}{\left\langle |f(Q)|^2 \right\rangle N}. \tag{2.39}$$

The expression for α is independent of Q for neutrons, as $f_{\mathrm{j}}(Q)$ equals the coherent scattering length of the j^{th} nucleus. With $\beta = \rho\,\langle \mathrm{b}\rangle$, where $\rho = N/V$ is the atomic number density,

$$\alpha(Q) = \rho. \tag{2.40}$$

Eq. 2.38 can be rearranged to express the pair distribution function, which represents the correlation between the j^{th} and k^{th} atoms in real space, in terms of $F_{\mathrm{tot}}(Q)$:[75]

$$G(r) = 1 + \frac{1}{2\pi^2 r}\int_0^\infty Q\frac{[F_{\mathrm{tot}}(Q) - 1]}{\alpha(Q)}\sin(Qr)\mathrm{d}Q. \tag{2.41}$$

The pair distribution function is straightforward to interpret for monatomic samples, because distances between the peaks correspond to atoms of the same type. For a sample

consisting of M atomic species, $G(r)$ can be decomposed into a linear combination of $M(M+1)/2$ radial distribution functions (RDFs), g_{JK}:

$$G(r) = \sum_{\mathrm{J=1}}^{\mathrm{M}} \sum_{\mathrm{K=1}}^{\mathrm{M}} \gamma_{\mathrm{JK}} g_{\mathrm{JK}}(r). \tag{2.42}$$

For neutrons

$$\gamma_{\mathrm{JK}} = (2 - \delta_{\mathrm{JK}}) c_{\mathrm{J}} c_{\mathrm{K}} \left\langle \mathrm{b_J} \right\rangle \left\langle \mathrm{b_K} \right\rangle, \tag{2.43}$$

where the Kronecker delta function, δ_{JK}, equals 1 if J=K and 0 otherwise, and c_{J} and c_{K} are the concentrations of atomic species J and K, respectively, in the sample.[152] RDFs are frequently used in this thesis to study the relative distributions of pairs of atoms. The peaks in RDFs indicate the position of coordination shells.[223] Integration of RDFs also allows calculation of coordination numbers (CNs),[180] which give the average number of atoms of species K within a given radius of atoms of species J:[271]

$$n_{\mathrm{J}}^{\mathrm{K}} = 4\pi\rho c_{\mathrm{K}} \int_{r_1}^{r_2} r^2 g_{\mathrm{JK}} \mathrm{d}r. \tag{2.44}$$

The decomposition of Eq. 2.42 is mirrored in Q space:[283]

$$F_{\mathrm{tot}}(Q) = \sum_{\mathrm{J=1}}^{\mathrm{M}} \sum_{\mathrm{K=1}}^{\mathrm{M}} \gamma_{\mathrm{JK}} S_{\mathrm{JK}}(Q), \tag{2.45}$$

where $S_{\mathrm{JK}}(Q)$ is the partial structure factor. Partial structure factors, the reciprocal space equivalents of RDFs, by analogy with Eq. 2.38, are defined as

$$S_{\mathrm{JK}}(Q) = 1 + \frac{4\pi\ \alpha(Q)}{Q} \int_0^{\infty} r[g_{\mathrm{JK}}(r) - 1] \sin(Qr) \mathrm{d}r. \tag{2.46}$$

A simple example of the application of Eq. 2.45 to a system of two molecular species (i.e. water) is shown in Eq. 2.47:

$$F_{\mathrm{tot}}(Q) = c_{\mathrm{H}}^2 \left\langle \mathrm{b_H} \right\rangle^2 S(Q)_{\mathrm{HH}} + c_{\mathrm{O}}^2 \left\langle \mathrm{b_O} \right\rangle^2 S(Q)_{OO} + 2c_{\mathrm{H}} c_{\mathrm{O}} \left\langle \mathrm{b_H} \right\rangle \left\langle \mathrm{b_O} \right\rangle S(Q)_{HO}, \tag{2.47}$$

where the subscripts H and O denote hydrogen and oxygen atoms, respectively.

Using Eq. 2.43, Eq. 2.45 and Eq. 2.46 an expression for the distinct scattering term,

$F(Q)$, often referred to as the structure factor, in terms of atomic concentrations, neutron scattering lengths, the number density and RDFs is realised:

$$\overbrace{F(Q)}^{\text{Structure Factor}} = \sum_{\mathrm{J,K}\geq\mathrm{J}} (2-\delta_{\mathrm{JK}}) \overbrace{c_{\mathrm{J}}\, c_{\mathrm{K}}}^{\text{Concentration}} \underbrace{\langle \mathrm{b_J}\rangle\, \langle \mathrm{b_K}\rangle}_{\substack{\text{Scattering}\\ \text{Lengths}}} 4\pi \overbrace{\rho}^{\substack{\text{Number}\\ \text{Density}}} \int_0^\infty r^2[\, \underbrace{g_{\mathrm{JK}}(r)}_{\substack{\text{Radial}\\ \text{Distribution}\\ \text{Function}}} -1] \overbrace{\frac{\sin(Qr)}{Qr}\mathrm{d}r}^{\text{Fourier Transform}} \quad . \tag{2.48}$$

Typically, $F(Q)$ is quoted with units of barns atom^{-1} sr^{-1}. For X-rays, replacing the neutron scattering lengths in Eq. 2.48 with the X-ray form factors yields an expression for the static structure factor. Eq. 2.48 shows the relationship between important known physical quantities of a sample (i.e. composition and density) and the structure and, as shown by Eq. 2.34, is directly related to the quantity measured in a scattering experiment.

2.1.2 The SANDALS Instrument

All neutron diffraction data presented in this thesis were obtained using the Small Angle Neutron Diffractometer for Amorphous and Liquid Samples (SANDALS), which is located at the ISIS facility (STFC, UK). The SANDALS instrument is a diffractometer, which saw its first neutrons in 1989, purposely built for the study of amorphous materials and liquids.[16]

Neutron production at the ISIS facility begins with an ion source that produces negative hydrogen ions (H^-) from hydrogen gas. H^- ions are accelerated and separated into bunches by a Radio Frequency Quadrupole accelerator. The bunches of H^- ions are further accelerated by a linear accelerator to speeds equivalent to 37% of the speed of light (37%c). Subsequently, the ions enter a synchrotron, a 163 m circumference ring of magnets is used to focus the beam. As the ions enter the synchrotron, a 0.3 μm thick aluminium oxide foil strips away electrons from the H^- ions. The resulting bunches of protons are accelerated in the synchroton using radio frequency electric fields to energies of 800 MeV (84%c). These protons collide with a solid tungsten target, producing 15-20 neutrons for every incident proton. This process takes place under constant cooling to dissipate the heat from the 160 kW proton beam.

This method of neutron production is termed spallation. The other principal class of neutron source, nuclear reactors (e.g. the Institut Laue-Langevin, Grenoble), produces neutrons through a nuclear induced fission process. The grouping of protons, hence neutrons, into bunches means that spallation sources are pulsed sources, whereas reactor sources produce a continuous flux of neutrons. The spallation source used by the SANDALS instrument produces 4×10^{14} neutrons at a frequency of 50 Hz.

Neutrons produced by both spallation and reactor sources have energies that are too high for investigating condensed matter.[223] Consequently, neutrons used by the SANDALS instrument are moderated by liquid methane at 110 K. Hydrogenated materials, such as methane and water, are typically used as moderator materials because of the large absorption cross section of hydrogen essential for the energy exchange process. Without moderation the high energy of the neutrons may mean too many neutrons pass straight through a sample for sample measurement to be undertaken. Moderation reduces neutrons to an average speed of $\sim$2200 m s^{-1}. Such neutrons are termed thermal neutrons. The energy distribution of thermal neutrons is described by a Maxwellian distribution. Higher energy neutrons may be under-moderated and are known as epithermal neutrons. The energy range of the moderated neutrons can be controlled by changing the moderator temperature. The spectrum of the moderated neutrons used by the SANDALS instrument is shown as a function of wavelength, which is related to the speed of the neutrons *via* the de Broglie equation, in Fig. 2.4.

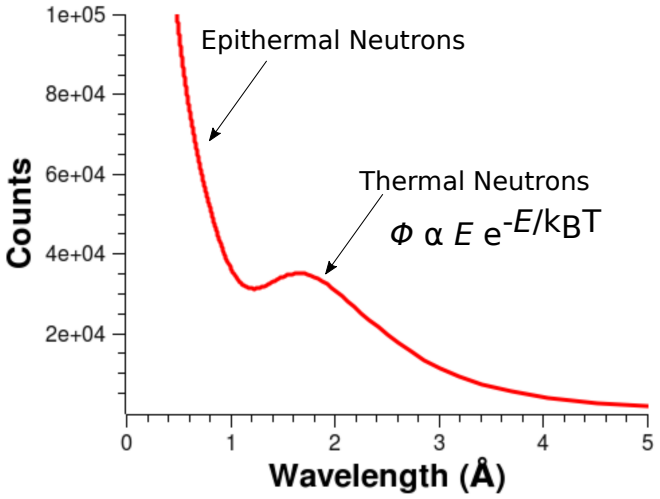


Figure 2.4: The spectrum of the moderated neutrons used by the SANDALS instrument. The flux, Φ, of thermal neutrons is related to the energy of the incident neutrons, E, and the moderator temperature, T. k_B is the Boltzmann constant. (Adapted from [17].)

Pulses of neutrons, which contain neutrons of wavelengths in the range 0.05 - 4.95 Å, travel down the 11 m, B_4C collimator from the moderator. As the neutrons travel down the instrument, the faster, shorter wavelength neutrons arrive at the sample first. Neutrons can thus be sorted by the time they arrive at a detetctor and the diffraction measurement performed as a function of time.[223] This technique is known as time-of-flight (TOF). The advantage of TOF is that it permits calculation of neutron wavelengths, hence Q (Eq. 2.7), from the de Broglie equation, so that monochromation of the neutron beam is unnecessary.[d] Monochromation reduces neutron flux, meaning samples have to be measured for longer to realise the same signal. The reason that the moderator thickness is not increased in order to reduce the flux of epithermal neutrons is because doing so would affect the TOF resolution. TOF calculation is made possible by the pulsed nature of the neutron source.

Using a white beam, in conjunction with detectors at a range of angular positions (*vide infra*), also facilitates measurement over a wide Q range (0.1 - 50 $Å^{-1}$). The wide Q range of the SANDALS instrument means that intermolecular and intramolecular distances can be determined. Furthermore, as the real space resolution of a measured correlation function is determined by the maximum Q value, measuring data over a wide Q range increases the resolution of the RDFs extracted.[95]

The detector array of the SANDALS instrument consists of 633 ZnS detectors divided 18 into groups over a range of 3.8-41°. (2θ is as defined in Fig. 2.1 in Section 2.1.1.) Having a large number of detectors at low angles, the SANDALS instrument is particularly well designed for measuring structure factors of samples containing light atoms such as hydrogen and deuterium.[16, 128] Detectors at low angles (<45°), forward facing detectors, are suitable for measuring samples consisting of light atoms, because at angles >45° correcting for inelastic scattering is particularly difficult. The distance from sample to detector ranges between 0.75 m and 4.0 m. A diagram showing the vacuum tank in which samples are placed and the detector array is shown in Fig. 2.5. Samples are measured under vacuum (1 - 10^{-2} mbar), so that scattering by air molecules is minimised and the beam is not attenuated.

[d] $\lambda = \frac{h}{m_n v} \implies \lambda = \frac{ht}{m_n L}$, where t is the time, measured from the target impact, when a diffracted neutron is deflected at a given flight path distance, L, from the moderator. h and m_n are the Planck constant and neutron mass, respectively.[252]

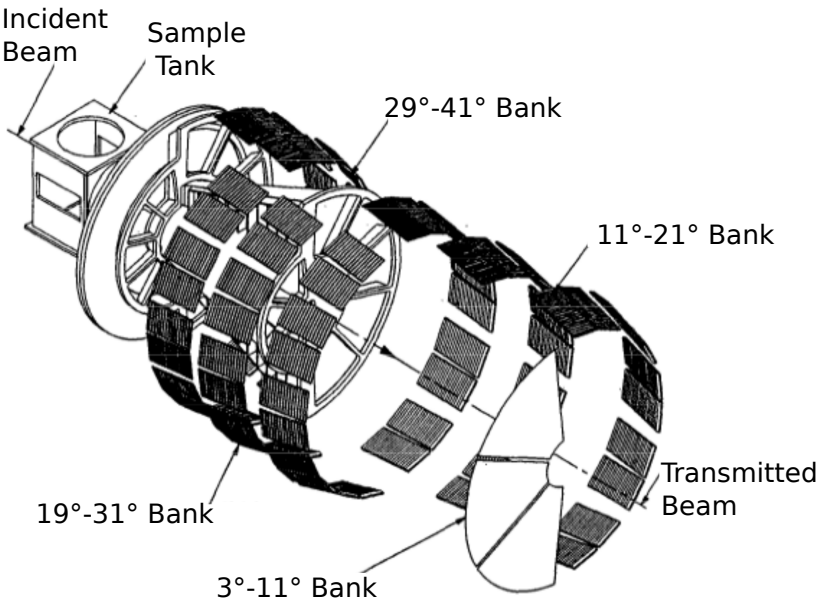


Figure 2.5: A diagram showing the vacuum tank and detector array of the SANDALS instrument. (Adapted from [256].)

Although the resolution of the instrument could be improved by increasing the distance between the sample and the detectors, increasing the Q range by increasing 2θ, the greater the distance of the detectors from the sample, the lower the measured signal. In fact, only ∼1% of the neutrons incident on a sample measured using the SANDALS instrument are scattered. This issue is particularly pertinent for amorphous materials and liquids, because they lack the repeating structure of crystals.

A set of B_4C jaws can be used to adjust the normally circular beam (diameter 32 mm).[16] Like the collimator, the jaws are made of B^{10} because B^{10} is highly absorbing. Consequently, a less divergent beam is incident on the sample. Having a beam that can be covered by the sample (i.e. the area of the beam is smaller than the area of the sample) reduces background.

Samples may be mounted on a candlestick, essentially a metal rod to which a single sample at a time may be attached, for measurement under ambient conditions or to a sample changer. The sample changer is attached to a water bath (-5°C-110°C), circulating 8 litres of 50% ethylene glycol-water mixture, for temperature control

2.1.3 Measurements, Corrections and Processing

The aim of measuring a sample with the SANDALS instrument or, indeed, any diffractometer is to obtain the differential cross section of the sample (Eq. 2.8 and Eq. 2.9), such that the structure of the sample may be determined (Eq. 2.48). Accurately obtaining the differential cross section is thus crucial for successful analysis of diffraction data. This Section delineates the measurements necessary to obtain the differential cross section of a sample (Section 2.1.3.1) and addresses the corrections and procedures required to ensure that the differential cross section is measured accurately and that structural information may be reliably determined from its measurement (Section 2.1.3.2). Steps taken to correct and process the diffraction data in this thesis were carried out *via* GudrunN, a program that analyses diffraction data detector by detector before finally merging the data into a single spectrum.[264]

2.1.3.1 Measurements

There are four measurements necessary for the differential cross section of a sample on an absolute scale to be obtained: the sample background, the empty container in which the sample was measured, the sample in its container and a standard for calibration.

2.1.3.1.1 Sample Background The first measurement necessary to obtain the differential cross section of a sample is the sample background. Measurement of sample background is made with the neutron beam on using the sample environment in which samples will be measured (Section 2.1.2). In this way, any background radiation may be eliminated from the final analysis.

2.1.3.1.2 Empty Container Measurement must also be made of each of the containers in which samples are to be measured, so that their contribution to the measured differential cross section may also be removed from the final analysis. Two types of containers were used in this thesis.

The first type of containers are flat Ti/Zr canisters (Fig. 2.6), which have walls 1 mm

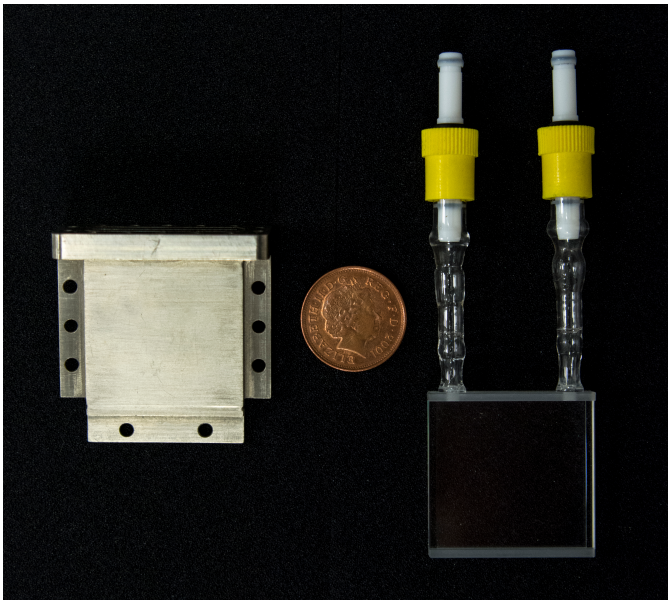


Figure 2.6: A Ti/Zr canister (left) and SiO_2 cell (right) used to contain samples measured by neutron diffraction. A 2 pence piece is shown (centre) for scale. Photograph courtesy of Richard Gillams.

thick, a cavity for the sample 1 mm thick and a density of 0.0542 atoms $Å^{-3}$. A thin sample cavity is important to limit multiple scattering effects and attenuation. Ti/Zr alloy canisters are used as Ti and Zr have coherent scattering lengths of -3.438 fm and 7.160 fm,[248] respectively, and can be combined in such a way (68% Ti and 32% Zr) to produce negligible coherent scattering from the sample containers. The canisters are sealed using Teflon O-rings held in place by Ti/Zr lids screwed to the main body of the canisters.

The second type of containers are specially designed SiO_2 cells (Fig. 2.6) of wall thickness 2 mm, sample cavity thickness 1 mm and density 0.0663 atoms $Å^{-3}$. Their design includes Young's taps, which are vacuum tight. Measurement of samples contained in these SiO_2 cells was necessary because of the possibility of some of the samples corroding the Ti/Zr canisters, resulting in contamination of the samples.

2.1.3.1.3 Sample in Container Samples were measured in the containers described above.

2.1.3.1.4 Vanadium Standard Normalisation to a standard measured under the same conditions (e.g. beam size and detector positions) as the samples in their containers allows diffraction data to be put on an absolute scale by eliminating calibration constants, such as detector efficiency.[178, 264] An almost ideal material for performing this normalisation procedure is vanadium, predominantly because it can be approximated as a purely incoherent scatterer whose scattering does not vary with Q.[264] Scattering almost completely incoherently means that vanadium has only very small Bragg peaks, which can be removed from the spectrum in GudrunN. Furthermore, the mass of vanadium atom renders inelastic scattering small. (See Section 2.1.3.2.5 for a discussion of the means by which the total differential cross section is corrected for inelastic scattering.) The differential cross section of vanadium as a function of Q is shown in Fig. 2.7 (a), where the small range of values of the differential cross section, small Bragg peaks and small inelasticity effects are apparent. As vanadium is a solid material with a known density ($0.07921\,\text{atoms}\,\text{Å}^{-3}$), does not need to be measured in a container and may be formed into cylinders or flat plates as required, its use as a standard has a number of practical advantages. The lack of a container means that diffraction data from a vanadium standard only need to have the background subtracted, although monitor normalisation (Section 2.1.3.2.2) should still be applied. The vanadium plate used in the work presented here had a thickness of 3.48 mm.

Combining the measurements described in Sections 2.1.3.1.1, 2.1.3.1.2 and 2.1.3.1.3 with the differential cross section of vanadium to form an expression for the differential cross section of the sample yields:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Sample}} = \frac{\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Sample in Container}} - \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Empty Container}}}{\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Vanadium Standard}} - \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Sample Background}}}. \tag{2.49}$$

In Eq. 2.49 the scattering from the empty container has been subtracted from the scattering of the sample measured in the same container, and this difference has subsequently been normalised to a vanadium standard whose scattering has been corrected for background radiation.

2.1.3.2 Corrections and Processing

Calculation of the normalised differential cross section of the sample in Eq. 2.49 neglects errors that may arise from the detectors, fluctuation of the incident flux, attenuation and multiple scattering corrections. The quantity defined by Eq. 2.49 also ecompasses both inelastic scattering and incoherent scattering. As discussed in Section 2.1.1 (Eq. 2.34), diffraction is concerned with elastic scattering, and incoherent scattering offers no information about the positions of atoms in the sample. Thus, an important aspect of processing the diffraction data is eliminating as much as possible inelastic scattering and incoherent scattering from the differential cross section.

2.1.3.2.1 Detectors The time after a neutron event during which another measurement by the same detector is not possible, typically 2-10 μs for the SANDALS instrument, is termed the deadtime.[269] If τ is the deadtime of a given detector, R_{m} the measured count rate in the time channel and Δ the width of the time channel, then the deadtime of the detector, D, is given by

$$D = \Delta R_{\mathrm{m}} \tau. \tag{2.50}$$

Therefore, the count rate, R, which would be measured for a detector with zero deadtime is greater than R_{m} in proportion to the time the detector is silent:

$$R = \frac{\Delta}{\Delta - D} R_{\mathrm{m}} = \frac{R_{\mathrm{m}}}{1 - R_{\mathrm{m}}\tau}. \tag{2.51}$$

However, Eq. 2.51 applies only when the time channel is broad compared with the deadtime (i.e. $\Delta > D$).[269] If the time channel is narrower than the deadtime, previous time channels may contribute to the deadtime in a given channel. For example, if channels n to m contribute to the deadtime in channel m, the length of time channel m is dead is given by[269]

$$D_{\mathrm{m}} = \tau \sum_{j=n}^{m} \Delta_{\mathrm{j}} R_{\mathrm{j}}, \tag{2.52}$$

where Δ_{j} and R_{j} are the channel width and count rate in channel j, respectively. The limits of j are determined by inspection. Substitution of Eq. 2.52 into Eq 2.51, exchanging

D for D_{m}, thus gives the correction for when the channel are narrower than the deadtime.

In reality, the deadtime correction is more complicated because, as it is not practical to have a separate input for each detector into the data acquisition electronics, an encoder is used to create a binary address describing the detector fired. Consequently, if the deadtime of the encoder is longer than the deadtime of the detector, it is the deadtime of the encoder which determines the deadtime of the detector. Furthermore, since the encoder can only process one event at a time, all the detectors that feed into that encoder are effectively dead when any one detectors fires. In this situation, Eq. 2.52 must also be summed over all channels which feed into an encoder:

$$D_{\mathrm{m}} = \tau \sum_{j=n}^{m} \sum_{k} \Delta_{\mathrm{j}} R_{\mathrm{j,k}}. \tag{2.53}$$

In Eq. 2.53 $R_{\mathrm{j,k}}$ is the count rate in time channel j and encoder channel k and Δ_{j} is the encoder deadtime. Thus, even though the deadtime for an individual detector may be small, the grouping of multiple detectors into an encoder results in an enhancement of the count rate as far as the deadtime is concerned. So the deadtime correction could be much larger than might be apparent from the count rate of an individual detector.[269]

The large number of detectors used by the SANDALS instrument means that during measurement one or more of these detectors may not function optimally. Therefore, GudrunN includes a mechanism of removing such detectors from the data analysis. This mechanism involves calculating the ratio of the counts recorded for each sample measured to the counts recorded for the vanadium standard (Section 2.1.3.1.4). For each group of detectors the spread of these ratios and their associated standard deviations is assessed so that detetctors in a given group whose measured counts is far from the average can be eliminated.

2.1.3.2.2 Incident Beam Small fluctuations in the neutron flux due to moderator temperature mean that all measurements made in Section 2.1.3.1 must be normalised to the number of neutrons detected per second by a detector placed in the incident beam before the sample, the monitor detector, because the number of neutrons that impinge on any samples is directly proprotional to this quantity. Similarly, a transmission monitor

placed after the sample can be used to calculate the fraction of neutrons transmitted by the sample by dividing the count rate of the transmission monitor by the count rate measured by the monitor detector. The transmission monitor can be used to confirm the identity of a sample by computing the number of neutrons scattered relative to the expected number given the composition (DCS Level). These normalisation procedures should also include coefficients to account for detector efficiency and be corrected for deadtime.

2.1.3.2.3 Attenuation and Multiple Scattering Attenuation and multiple scattering corrections performed by GudrunN are based on previously published works whose details are beyond the scope of this section.[255] The procedure involves successive solutions of a one dimensional integral equation, each iteration corresponding to a neutron collision.[255] These calculations require the beam size, sample geometry and sample composition. Correcting for attenuation and multiple scattering requires measurement of both the sample in its container and the empty container, each of these quantities being affected by attenuation and multiple scattering.[269] The calculation is performed using the the average scattering cross section of the sample and the container for a particular wavelength, assuming the atoms in that component scatter isotropically, although multiple scattering itself is not isotropic.[264]

2.1.3.2.4 Self Scattering The term of interest, the distinct scattering term ($F(Q)$), oscillates about a baseline set by the self scattering (Eq.2.34).[264] Inputting the composition of the sample into GudrunN means that the self scattering term can be calculated from the concentration of the different atomic species and their corresponding coherent and incoherent scattering lengths so that the self scattering term can be removed from the final analysis.

As the incoherent scattering length of hydrogen is ∼6 times greater than the incoherent scattering length of deuterium (see Table 2.1), samples which contain a higher proportion of hydrogen compared with deuterium experience a Q independent increase in their differential cross sections. This is illustrated by Fig. 2.7, which shows the measured differential cross section of seven isotopically distinct samples of indole dissolved

in methanol/water solutions. As the proportion of hydrogen increases relative to the proportion of deuterium, the differential cross section becomes greater. Consequently, the fully protonated sample has the greatest total differential cross section and the sample consisting of indole dissolved in a mixture of CD_3OD and D_2O has the smallest differential cross section.

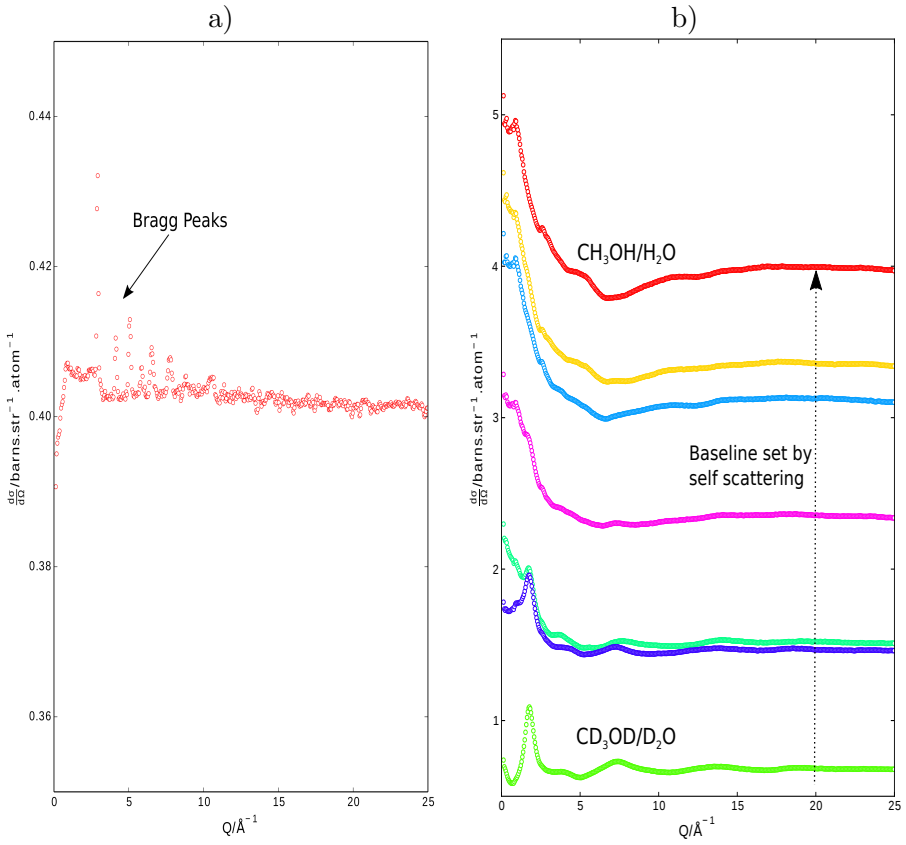


Figure 2.7: The differential cross section, $\frac{d\sigma}{d\Omega}$, as a function of Q for the vanadium standard (left). $\frac{d\sigma}{d\Omega}$ as a function of Q for indole dissolved methanol/water solutions. These methanol/water solutions vary in their deuteration from a fully protonated sample (CH_3OH/H_2O) to indole dissolved in CD_3OD/D_2O.

2.1.3.2.5 Inelasticity Corrections Although in Section 2.1.1 the approximation of solely elastic scattering of the incident neutrons by the sample is promptly made, it is important consider inelastic scattering term, $P(Q,\omega)$, included in Eq. 2.34. Inelastic scattering, which predominantly occurs at low Q, is primarily experienced by light atoms, such as hydrogen, due to having a similar mass as neutrons.

To determine $P(Q,\omega)$, strictly speaking, it is necessary to know the total dynamic structure factor, $F_{\text{tot}}(Q,\omega)$ (Eq. 2.4). However, knowing $F_{\text{tot}}(Q,\omega)$ preempts the need

for a diffraction experiment, as the total static structure factor, $F_{\mathrm{tot}}(Q)$, is obtainable by direct integration of $F_{\mathrm{tot}}(Q,\omega)$. Such an experiment is unfortunately impractical because of the time required for measurement of the complete dynamic structure factor.[269] Consequently, efforts have been made to estimate $P(Q,\omega)$. The best known method of calculating the inelastic scattering is the Placzek correction.[e][217]

Inelastic scattering contributes a Q dependent self scattering background to the total differential cross section, but does not have a major impact on the distinct scattering. The essence of the Placzek correction lies in recognising that inelasticity affects primarily the self scattering. In fact, analysis of the the self scattering and distinct scattering contributions to $F_{\mathrm{tot}}(Q,\omega)$ has shown that, whereas for self scattering the energy transfer is finite and increases with Q, the distinct scattering energy transfer always averages to zero, regardless of the value of Q.[264] The Placzek correction involves a Taylor expansion of $F_{\mathrm{tot}}(Q,\omega)$ about its elastic value. However, this expansion does not work for light atoms, as the Placzek correction depends on the mass of the nuclei of the sample being more massive than the mass of the incident neutrons.

As the Placzek correction is not valid for light atoms, the diffraction data from the samples measured in this thesis were not corrected for inelastic scattering by means of the Placzek correction. The diffraction data measured for the vanadium standard described in Section 2.1.3.1.4 were corrected for inelastic scattering using the Placzek correction in GudrunN, because the mass of a vanadium atom is much greater than that of a neutron.

When the Placzek expansion method does not work, typically an *ad hoc* method is employed.[73, 263] The way this is done within GudrunN is through a series of exponentials in Q space:

$$f(Q) = Ae^{\frac{-Q}{b}}. \tag{2.54}$$

A and b are amplitude and decay constants, respectively. However, care must be taken when using this method, as such a background subtraction risks introducing systematic effects unrelated to the structure of the material or removing small angle neutron

[e]In the nomenclature of this thesis, the Placzek correction refers specifically to the methodology of correcting for inelastic scattering developed by Placzek[48, 217] and is not used as a generic term for inelasticity corrections as is sometimes the case.

scattering.[264] The amplitude of any function subtracted from the samples is typically smaller for more deuterated samples, as the inelastic scattering from hydrogen is larger than the inelastic scattering from deuterium. Inelastic scattering of X-rays, Compton scattering, can be calculated for each atom within the independent atom approximation.[111]

2.2 Empirical Potential Structure Refinement

The different neutron scattering lengths of isotopes can be used to determine partial structure factors (Eq. 2.46). For example, the opposite signs of the neutron scattering lengths of hydrogen and deuterium (Table 2.1) can be exploited to find $S_{\mathrm{HH}}(Q)$, $S_{\mathrm{OO}}(Q)$ and $S_{\mathrm{HO}}(Q)$ from measurement of the total static structure factor, $F_{\mathrm{tot}}(Q)$ (Eq. 2.45), of H_2O (Eq. 2.47), D_2O and HDO by means of a set of simultaneous equations, subsequently allowing RDFs be computed from the experimental data alone. However, use of neutron diffraction alone, even with isotopic substitution methods, is insufficient to extract all interatomic information when considering systems containing a large number of different atoms. By using modelling techniques that produce results consistent with experimental data, otherwise inaccessible information about a system can be gained. One such modelling technique is Empirical Potential Structure Refinement (EPSR).

EPSR is a computational modelling technique specifically designed for building atomistic models of disordered systems.[46, 157, 257, 258, 265, 267] Using neutron and/or X-ray diffraction data allied to physical and chemical information (i.e. density and molecular structure) as constraints, EPSR produces a 3D structural model consistent with experimental data. As a Monte Carlo-based technique (Section 1.2.2), EPSR is based on the Boltzman-like distributions of statistical mechanics. As such, EPSR should produce the most disordered, and therefore most probable, solutions to a problem which are compatible with the specified constraints.[262] Unlike reverse Monte Carlo (RMC), from which EPSR evolved, EPSR does not employ hard sphere constraints to model atomic interactions.[179, 262] Similar to all modelling processes, the structure produced by EPSR is not unique: EPSR does not produce the only possible 3D structure which could be consistent with the experimental data.

Starting with a simulation box of atoms/molecules, a store of coordinates characterised atomic density (ρ), energy (U) and pressure (P), the EPSR process works by moving each atom, molecule and functional group until the associated simulated diffraction data matches the experimental diffraction data. There are four ways in which the atoms, molecules and functional groups within the model are moved: whole molecule translations, whole molecule rotations, rotation of specific molecular sidechains and individual atomic moves within molecules. Once good agreement is realised between the experimental diffraction data and the simulated diffraction data (i.e. there is a good fit), a set of analytical tools can be used to interrogate the 3D structure produced by the EPSR process. Tools used to analyse atomic/molecular interactions and the 3D structure include RDFs (Eq. 2.42), CNs (Eq. 2.44) and torsional angles. Analysis was also conducted using ANGULA (Section 2.3).

Forces between atoms in the system involve only interatomic forces and, in the case of ions, electric charges. Consequently, the force acting between an atom α (σ_α, ϵ_α) and an atom β (σ_β, ϵ_β) can be described by a Lennard-Jones potential (Fig. 2.8), which is composed of repulsive and attractive terms:

$$U^{\mathrm{LJ}}(r_{\alpha\beta}) = 4\epsilon_{\alpha\beta}\left[\overbrace{\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}}\right)^{12}}^{\text{Repulsive}} - \underbrace{\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha\beta}}\right)^{6}}_{\text{Attractive}}\right], \tag{2.55}$$

where $r_{\alpha\beta}$ is the distance between the atoms, which is calculated from their coordinates using Pythagoras' theorem. The Lorentz-Berthelot mixing rules[28] are invoked for calculating the values of σ, the distance at which the Lennard-Jones potential is zero, and ϵ, the depth of the potential well, for different atom types:

$$\sigma_{\alpha\beta} = \frac{1}{2}(\sigma_\alpha + \sigma_\beta) \qquad \epsilon_{\alpha\beta} = (\epsilon_\alpha \epsilon_\beta)^{\frac{1}{2}}. \tag{2.56}$$

Intramolecular interactions are described by an intramolecular harmonic potential, defining molecular structure in EPSR, which allows for the flexibility of molecules in a disordered system. The total intramolecular potential of the system summed over all i

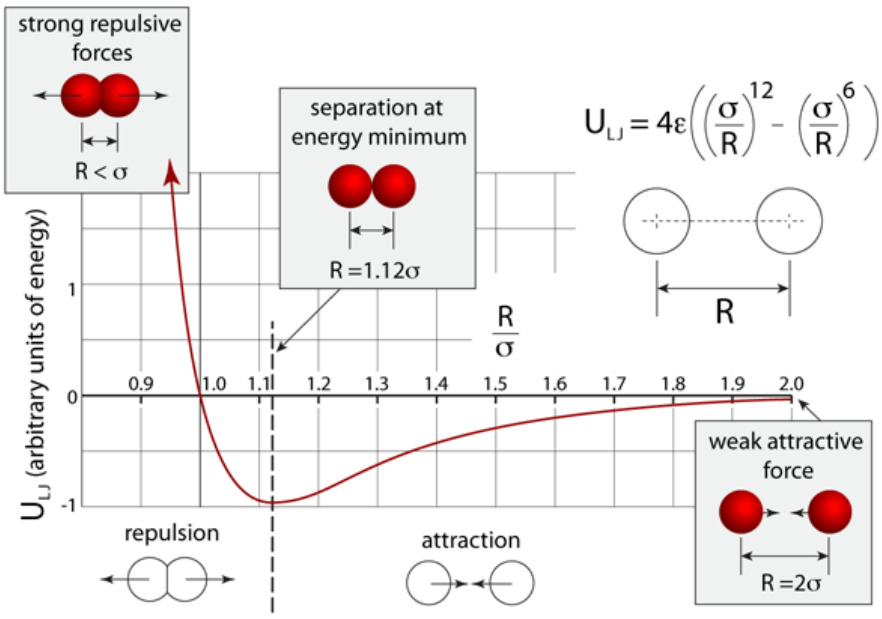


Figure 2.8: The Lennard-Jones potential used to characterise interatomic interactions in EPSR. (Taken from [309].)

molecules is

$$U^{\text{Intra}}(r_{\alpha_i\beta_i}) = C\sum_{i}\sum_{\alpha\beta\neq\alpha}\frac{(r_{\alpha_i\beta_i} - d_{\alpha\beta})^2}{2w_{\alpha\beta}^2}, \tag{2.57}$$

where C [$\text{Å}^{-1}\,u^{-1}$] is a constant[f], $r_{\alpha_i\beta_i}$ is the separation of atom α and atom β in molecule i, $d_{\alpha\beta}$ is the average distance between the atoms α and β of molecule i and $w = d_{\alpha\beta}/\mu_{\alpha\beta}^{\frac{1}{2}}$ is a broadening function (μ is the reduced mass[g]) used to avoid having to refine an independent Debye-Waller factor [h] for each intramolecular distance.

Intermolecular interactions are more complicated. Not only are interatomic forces present, in the sense of intramolecular interactions, but the sum of the forces of the atoms of a molecule on other molecules are also important. Intermolecular interactions are described by the sum of a Lennard-Jones potential (Eq. 2.55) plus effective Coulomb charges. In EPSR this is termed the intermolecular reference potential, U^{Ref}. Summing over all pair of atoms on distinct molecules, the total intermolecular energy of the system

[f] C is determined by comparing the simulated structure factors with the measurements at large Q.[262]

[g] $\mu = \text{M}_\alpha\text{M}_\beta/(\text{M}_\alpha + \text{M}_\beta)$, where M_α is the mass of atom α and M_β is the mass of atom β.

[h] The Debye-Waller factor, also known as the B factor or temperature factor, is used to describe the attenuation of neutron or X-ray scattering by the thermal motion.

is

$$U^{\mathrm{Ref}}(r_{\alpha_{\mathrm{i}}\beta_{\mathrm{i}}}) = \frac{1}{2}\sum_{\mathrm{ij}\neq\mathrm{i}}\sum_{\alpha\beta}\left(\overbrace{4\epsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha_{\mathrm{i}}\beta_{\mathrm{j}}}}\right)^{12}-\left(\frac{\sigma_{\alpha\beta}}{r_{\alpha_{\mathrm{i}}\beta_{\mathrm{j}}}}\right)^{6}\right]}^{\text{Lennard-Jones Potential}}+\underbrace{\frac{q_\alpha q_\beta}{4\pi\epsilon_0 r_{\alpha_{\mathrm{i}}\beta_{\mathrm{j}}}}}_{\text{Coulomb Potential}}\right), \quad (2.58)$$

where the factor 1/2 arises from the number of distinct interactions, q_α and q_β are the charges of atom α and atom β, respectively, and ϵ_0 is the permittivity of free space. U^{Ref} is used throughout the EPSR process but used on its own at the outset to ensure physically realistic geometries and atomic overlap. Values of ϵ, σ and q are input at the start of the EPSR process.

To save computation time, it is necessary to truncate the derivative of U^{Ref} at a reasonable radius value from the atomic site.[257, 258] Truncation of the first term of Eq. 2.58 is performed using a smoothly decaying function, $T(r)$:

$$T(r) = 1 \qquad r < r_{\mathrm{minpt}}, \quad (2.59)$$

$$T(r) = 0.5\left(1+\cos\pi\left(\frac{r-r_{\mathrm{minpt}}}{r_{\mathrm{maxpt}}-r_{\mathrm{minpt}}}\right)\right) \qquad r_{\mathrm{minpt}} < r < r_{\mathrm{maxpt}}, \quad (2.60)$$

$$T(r) = 0 \qquad r > r_{\mathrm{maxpt}}. \quad (2.61)$$

r_{minpt} and r_{maxpt} are the points at which $T(r)$ drops below 1 and 0, respectively. Typically values of $r_{\mathrm{minpt}} = 9\,\text{Å}$ and $r_{\mathrm{maxpt}} = 12\,\text{Å}$ are used.[262] The Coulombic term of Eq. 2.58 is truncated using a Heaviside step function.

In the Monte Carlo method employed by EPSR, where an atom or molecule may be moved is determined by a random number generator that generates numbers, R, between -1 and 1. For instance, an atom at postion (x_1, y_1, z_1) would move to a new position (x_2, y_2, z_2) defined by

$$x_2 = x_1 + (R \times \delta s), \qquad y_2 = y_1 + (R \times \delta s), \qquad z_2 = z_1 + (R \times \delta s), \quad (2.62)$$

where δs is a predefined step size. However, whether the move is actually accepted is dependent on the Metropolis condition: if the move results in a decrease ($\Delta U < 0$) in

the total energy of the simulation box ($U = U^{\mathrm{Ref}} + \mathrm{k_B}T\ U^{\mathrm{Intra}}$)[i], the move is accepted; if the proposed move increases U ($\Delta U > 0$), the move is accepted with a probability:

$$P_{\mathrm{m}} = e^{\frac{-\Delta U}{k_B T}}. \tag{2.63}$$

The simulation box in which these moves take place is a cubic simulation box defined by periodic boundary conditions (i.e. if a molecule leaves the simulation box on one side it re-enters on the opposite side), meaning that the original simulation box is effectively infintely repeated in all directions. As a result, a more manageable number of particles can be used to compute accurately bulk properties, such density. However, this repeating lattice of boxes results in Bragg peaks in the simulated diffraction pattern. To surmount this problem the minimum image convention is used, which means that only the interactions within a radius from a particle equivalent to half the length of the simulation box are calculated.

The core of the EPSR process is the empirical potential, U^{Ep}, which acts as a perturbation to U^{Ref} (Eq. 2.58).[258] In essence, U^{Ep} indicates how U^{Ref} should be modified so that experimental and simulated data agree more closely. Generally U^{Ep} cannot be used in computer simulations, because they already contain information arising from many-body cooperative effects (i.e. the packing of molecules in the material).[262] The potential U^{Tot} runs the refinement:

$$U^{\mathrm{Tot}}(r) = U^{\mathrm{Ref}}(r) + U^{\mathrm{Ep}}(r). \tag{2.64}$$

The most successful way of expressing U^{Ep} so that it represents the true difference between experimental and simulated data and contains minimal artefacts associated with statistical noise, systematic errors and truncation effects has been found to be in the form of a series of Poisson functions:[260]

$$U^{\mathrm{Ep}}(r) = \mathrm{k_B}T \sum_{\mathrm{k}} \mathrm{C_k} p_{\mathrm{n_k}}(r, \sigma_{\mathrm{r}}), \tag{2.65}$$

[i]Only the change in U^{Intra} determines the acceptance of movements of individual atoms.[259]

where

$$p_{\mathrm{n}}(r,\sigma_{\mathrm{r}}) = \frac{1}{4\pi\rho\sigma_{\mathrm{r}}^{3}(\mathrm{n}+2)!}\left(\frac{r}{\sigma_r}\right)^{\mathrm{n}} e^{\frac{-r}{\sigma_{\mathrm{r}}}}. \tag{2.66}$$

In Eq. 2.65 $\mathrm{C_k}$ is real and can be positive or negative, $\mathrm{k_B}$ is the Boltzmann constant and T is temperature. σ_r is a width function. Eq. 2.66 peaks near r≈nα for large n, with a width that gradually increases with n. Thus, Eq. 2.66 is a natural function with which to represent an intermolecular potential that varies rapidly at short distances and becomes more slowly varying at longer distances.[262] Approaching a Gaussian centred on $r = n\sigma$ as n becomes very large, the values of n are associated with the likely range of the empirical potential as chosen by the user. Eq. 2.66 can be expressed in reciprocal space *via*

$$p_{\mathrm{n}}(Q,\sigma) = 4\pi\rho \int p_{\mathrm{n}}(r,\sigma)e^{i\mathbf{Q}\cdot\mathbf{r}}\mathrm{d}\mathbf{r} \tag{2.67}$$

$$= \frac{1}{(\mathrm{n}+2)(\sqrt{1+Q^2\sigma^2})^{(\mathrm{n}+4)}}\left[2\cos(\mathrm{n}\alpha) + \frac{(1-Q^2\sigma^2)}{Q\sigma}\sin(\mathrm{n}\alpha)\right], \tag{2.68}$$

where α=arctan($Q\alpha$) and Q is as defined in Eq. 2.7. The reciprocal space equivalent of Eq. 2.65 is given by

$$U^{\mathrm{Ep}}(Q) = \sum_{\mathrm{k}} \mathrm{C_k} p_{\mathrm{n_k}}(r,\sigma_Q). \tag{2.69}$$

Having established a reasonable U^{Ref} and after a state of equilibrium being reach by the the Monte Carlo simulation, U^{Ep} needs to be linked to the experimental data. Following Eq. 2.48, the structure factor[j] of the $\mathrm{i^{th}}$ experimental data set can set be written as

$$F_{\mathrm{i}}(Q) = \sum_{\mathrm{j}=1}^{\mathrm{N}} \gamma_{\mathrm{ij}} S_{\mathrm{j}}(Q), \tag{2.70}$$

where j refers to an atomic pair $\alpha\beta$ and $\gamma_{\mathrm{ij}} = (2-\delta_{\alpha\beta})c_\alpha c_\beta b_\alpha b_\beta$. To generate a perturbation to each of the site-site potential functions an inversion of the matrix γ_{ij} is required. If sufficient isotopic samples are measured such that all the partial structure factors can be separated, in principle inverting γ_{ij} would not present a problem. However, as this is rarely the case, and, even if it were, the small contribution of some partial structure factors to the overall scattering signal raises questions about their reliablity, a feedback factor, f, is assigned to the experimental data.[262] A modified matrix, γ'_{ij},

[j]It is important to note that at this stage the neutron diffraction data in EPSR is the distinct scattering, $F(Q)$, as the self scattering has been removed from the data (Section 2.1.3.2.4).

is then formed, where input from the experimental data is accepted with a confidence f and input from the model is accepted with a confidence (1-f):

$$\gamma'_{\mathrm{ij}} = f\gamma_{\mathrm{ij}} \qquad 1 \leq \mathrm{i} \leq M \tag{2.71}$$

$$\gamma'_{\mathrm{ij}} = (1-f)\delta_{(\mathrm{i}-M)\mathrm{j}} \qquad M \leq \mathrm{i} \leq (M+N), \tag{2.72}$$

where M is the number of measured diffraction datasets, N is the number of distinct atomic interactions and δ is the Dirac delta function. This leads to an overdetermined (M+N)×N matrix in each case provided $0 < f < 1$. The inverse of this matrix, $\gamma_{\mathrm{ij}}^{-1}$, is then sought such that

$$P_{\mathrm{ii}'} = \sum_{\mathrm{j}=1}^{N} \gamma'_{\mathrm{ij}}\gamma_{\mathrm{ji}'}^{-1} - \delta_{\mathrm{ii}'} = 1 \tag{2.73}$$

has a minimum norm, giving the least squares solution for $\gamma_{\mathrm{ij}}^{-1}$ in the case of an overdetermined set of linear equations as here.[260] In EPSR a Monte Carlo loop is used to achieve inversion of the matrix.[262]

At the m^{th} iteration, the empirical potential for the j^{th} atom pair is

$$U_{\mathrm{m}}^{\mathrm{Ep\,(j)}}(r) = \mathrm{k_B}T \sum_{\mathrm{k}} \mathrm{C}_{\mathrm{k,m}}^{(\mathrm{j})} p_{\mathrm{n_k}}(r, \sigma_{\mathrm{r}}). \tag{2.74}$$

Initially, when m=1, $\mathrm{C}_{\mathrm{k,m}}^{(\mathrm{j})} = 0$. Following the m^{th} iteration, the difference between the experimental partial structure factors, $S_{\alpha\beta}^{\mathrm{exp}}(Q)$, and the simulated partial structure factors, $S_{\alpha\beta}^{\mathrm{sim}}(Q)$, is calculated, and represented by a sum of Poisson functions in the form of Eq. 2.69. In this way, difference coefficients, $\mathrm{C}_{\mathrm{k}}^{(\mathrm{i})}$, are calculated for each data set after each iteration. Changing as the EPSR process proceeds, these coefficients should go to zero when the simulation approaches the data. The difference coefficients are then accumulated in the potential coefficients:

$$\mathrm{C}_{\mathrm{k,m+1}}^{(\mathrm{j})} = \mathrm{C}_{\mathrm{k,m}}^{(\mathrm{j})} + \sum_{\mathrm{i}=1}^{\mathrm{M}} \gamma_{\mathrm{ji}}^{-1} \mathrm{C}_{\mathrm{k}}^{(\mathrm{i})}. \tag{2.75}$$

Revised values of $\mathrm{C}_{\mathrm{k,m+1}}^{(\mathrm{j})}$ are used in Eq. 2.65 to update the empirical potential. This process is repeated until the difference coefficients become so small that U^{Ep} no longer. The EPSR process is then allowed to continue in order to accumulate ensemble infor-

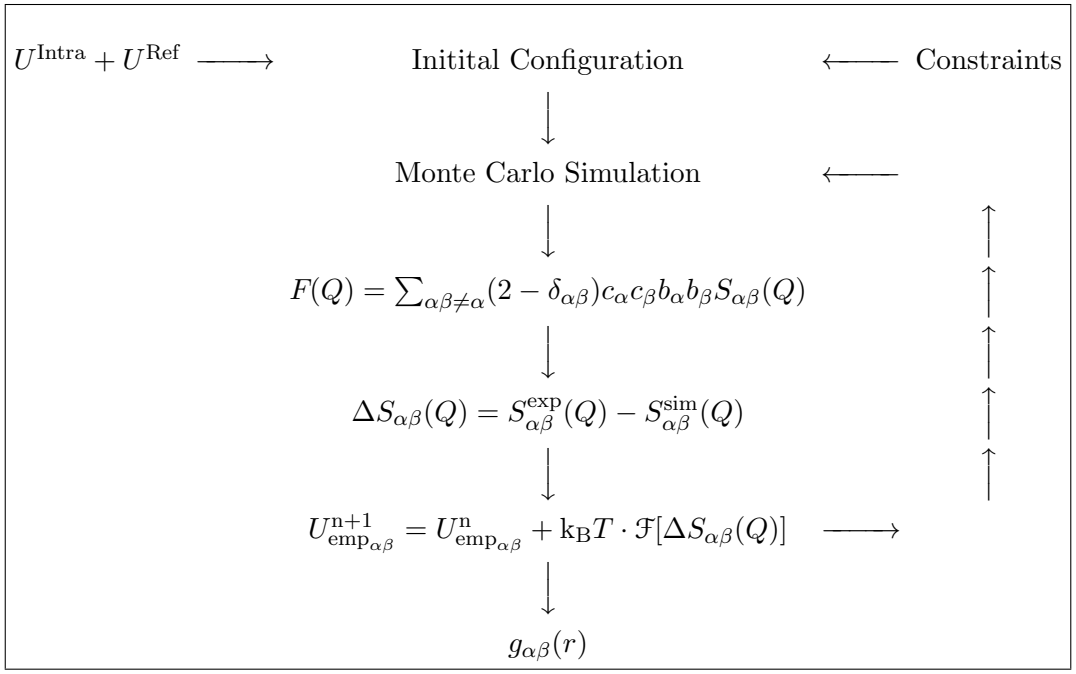


Figure 2.9: A summary of the EPSR process, where the difference between the experimental and the simulated partial structure factors is used to update the empirical potential in an iterative loop to drive the 3D structure in the EPSR simulation box towards consistency with experimental data.

mation. A summary of the EPSR process is shown in Fig. 2.9. An estimate of the partial structure factors can finally be made from those of the experimental data and the simulation based on confidence in each:

$$S_j(Q) = \sum_{i=1}^{M} \gamma_{\mathrm{ji}}^{-1} f S_{\mathrm{i}}^{\mathrm{exp}}(Q) + \sum_{i=M+1}^{M+N} \gamma_{\mathrm{ji}}^{-1}(1-f) S_{\mathrm{i-M}}^{\mathrm{sim}}(Q) \tag{2.76}$$

Eq. 2.75 illustrates how the empirical potential is cumulative, developing an amplitude and structure as the simulation proceeds.[262]

2.3 ANGULA

ANGULA is a program written to enable detailed study of small molecules and liquids.[k] Ultimately, it allows the relative position and relative orientation of molecules to be calculated.[34, 86, 125, 126, 202, 203, 204, 236, 235, 277] Able to read atomic coordinate

[k] ANGULA was developed by our collaborator Luis Carlos Pardo of the Universitat Politècnica de Catalunya, who has been generous enough to tailor aspects of the program for my work.[84]

files generated by a number of computational methods, such as EPSR or MD, the key information provided by ANGULA stems from the assignment of an orthogonal Cartesian coordinate system to molecules or fragments of molecules in a system. ANGULA also includes more prosaic features, such as the ability to plot a frequency distribution for a given property of the molecules in a system or the ability to select molecules by a chosen criterion (e.g. bond length, torsional angle, intermolecular distance, molecular species). The chief means by which information was extracted and processed from the systems studied in this thesis by employing ANGULA is outlined in this section.

2.3.1 Assigning the Coordinate System

As the information provided by ANGULA is rooted in the assignment of an orthogonal coordinate system(s) to each of the molecules chosen for analysis, the process of assigning the coordinate system is described in this section using the example of a benzene ring. A labelled benzene ring with a coordinate system assigned is shown in Fig. 2.10.

The origin of the coordinate system is placed by weighting the atoms of the molecule to which the coordinate system is being assigned. For example, if it were decided that it is most convenient to place the origin of the coordinate system in Fig. 2.10 at atom C1, C1 would be given a weighting of 1 and all other carbon and hydrogen atoms of benzene a weighting of 0. Alternatively, if the origin were required at the midpoint of the bond joining C1 and C2, both C1 and C2 would be given a weighting of 1 and all other atoms a weighting of 0. In Fig. 2.10 the origin has been placed at the geometrical centre of the benzene ring, meaning that each of the carbons have been given a weighting of 1.

Typically, the **z** axis is defined next. By placing what is termed a pseudo atom at a chosen location, using the same weighting procedure described in the above paragraph to do so, a vector from the origin to the pseudo atom, which corresponds to the **z** axis, can be drawn. In Fig. 2.10 the **z** axis begins at the geometrical centre of the benzene ring and passes through C3 (the chosen position of the pseudo atom).

In defining the **x** and **y** axes, the vector cross product is invoked. For the **y** axis in Fig. 2.10, two vectors, the first being the **z** axis itself and the second being from the

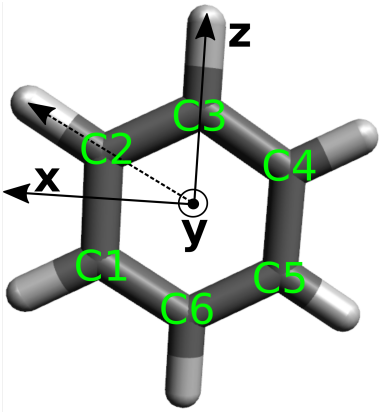


Figure 2.10: A coordinate system assigned to a benzene ring for analysis using ANGULA, where the **y** axis is coming out of the plane of the page. The dashed line is the vector used to define the **y** axis by the cross product of the **z** axis and this dashed line.

origin to C2 (dashed line), are used to map out a plane which then, by definition of the cross product, results in a vector perpendicular to the plane mapped by the two vectors. If the plane mapped out corresponds to the plane of the **z** axis, the vector produced by the cross product will be perpendicular to the **z** axis. The direction of vector produced by the cross product is given by the right-hand rule. The anticommutativity of the cross product means that the direction of the resulting vector can be reversed by changing the order of the cross product: $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$. The cross product described above results in a **y** axis coming out of the plane of the page in Fig. 2.10. Similarly, the **x** axis is defined as the cross product of the **y** and **z** vectors ($\mathbf{x} = \mathbf{y} \times \mathbf{z}$). By definition of the cross product, **x** is then normal to both **y** and **z**.

2.3.2 Spatial Distribution

The origins of the sets of orthogonal coordinate systems asssigned to molecules in ANGULA, as described in Section 2.3.1, can be used to determine the position of the molecules relative to one another. A central molecular species is chosen and the origin of the coordinate system assigned to a second molecular species can be plotted relative to the origin of the coordinate system of the first (central) molecular species. Typically, the distribution of the origin of the coordinate system assigned to solvent molecules is computed relative to the origin of the coordinate system assigned to solute molecules or a fragment of the solute molecules (e.g. a benzene ring, carbonyl group, amine group,

etc). It is in these terms (i.e. the distribution of a solvent around a solute) that ANGULA is discussed in the remainder of this section. In principle, any given molecular species can be plotted relative to any other respective molecular species. This procedure can be performed for a nearest neighbour analysis, where in each frame of the simulation being analysed the nearest neighbour solvent molecule relative to each solute molecule, as defined by the distance between the origins of their coordinate systems, is selected for analysis, or for a distance range, where all solvent molecules whose origin is within a chosen distance of the origin of the coordinate system of each solute molecule in each frame of the simulation are analysed.

In the work presented here, information regarding the relative position/distribution of molecules is usually displayed in the form of a 3D spatial density map (SDM),[33, 34, 86, 126, 125, 277] an example of which is shown in Fig. 2.11. For a nearest neighbour analysis, each frame of a simulation contributes a number points to the SDM equal to the number of solute molecules, corresponding to the position of the origin of the coordinate system assigned to the nearest neighbour solvent molecules around each central solute molecule. For a distance range analysis, the number of points contributed to the SDM equals the total number of solvent molecules within the chosen distance range of every solute molecule in the system, summed over all frames of the simulation. For both methods of analysis, as more frames are analysed the number of points around the central molecule in the SDM, representative of all such solute molecules in the system, accumulate such that an aggregate distribution is plotted. The density of these points around the central molecule is indicated by the scale bar on the right-hand side of the SDM displayed in Fig. 2.11, which shows the density of points per $Å^3$ (i.e. density of solvent molecules per $Å^3$ - an absolute scale). The level of the density shown in the SDMs is normalised both to the number of frames and the number of solute molecules, enabling comparison when these two factors are varied. Clearly, direct comparison between SDMs is only valid for systems of the same concentrations of molecular species.

Although in Fig. 2.11 the whole indole molecule has been plotted at the centre of the SDM, for other molecules studied only the fragment of the molecule to which the coordinate system has been assigned is shown in the SDM for clarity. When considering the SDMs it is also prudent to recognise that the density of solvation may be affected by

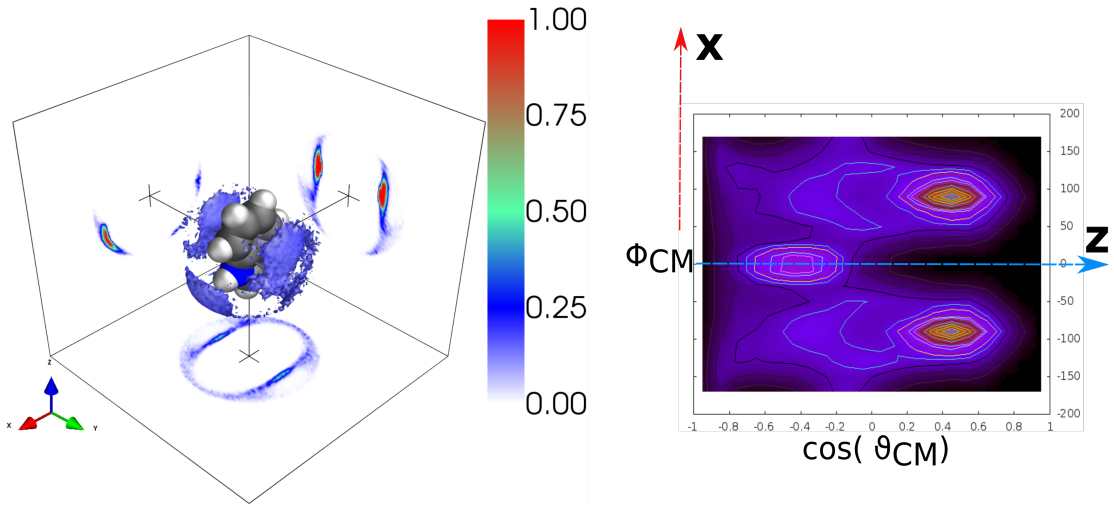


Figure 2.11: A spatial density map (SDM) showing the distribution of water molecules relative to indole (left). Water molecules within 5 Å of indole have been selected for analysis. The density of the isocontour surface is indicated by the scale bar, where the isocontour surface has been cut to include the top 50% of water molecules. Cut-throughs have been displaced by 6 Å from the origin for clarity. A 2D spatial density contour map showing the distribution of water molecules within 5 Å of indole in spherical polar coordinates (right). For each of these plots the origin of the coordinate system assigned to the indole molecule was at the midpoint of the bond between the benzene and pyrrole rings of indole. The origin of the water molecule was the oxygen atom.

the location of the coordinate system assigned to the molecules. For instance, whether the origin of the coordinate system is at the carbon or oxygen atom of methanol can impact on the solvation density displayed in the SDM.

Cut-throughs through the origin of the solute molecule, which is placed at the origin of the SDMs, are included in the plot. These cut-throughs are displaced from the origin for clarity in the each of the **x**, **y** and **z** directions. The isocontour surface (or isopycnic surface), which encloses points of equal density, contains a given percent of molecules. Unless stated otherwise, this is set to the top 50% of surrounding molecules. As the isocontour surface is increased to contain a larger proportion of solvent molecules, the density of the surface typically decreases. As the cut-throughs are defined through the origin, they are at a constant density regardless of the proportion of molecules included in the isocontour surface.

Spatial density (SD) contour maps can also be used to show the position of solvent molecules relative to the solute in 2D. Again, this can be performed for a nearest neighbour or a distance range analysis. These contour maps plot the position of the solvent

molecules in terms of the spherical polar coordinates of the origin of the coordinate system assigned to them (θ_{CM}, ϕ_{CM}) relative to the origin of the solute. The position of the solvent molecules in polar coordinates can be standardly converted to Cartesian coordinates: x=$r$sin(θ_{CM})cos(ϕ_{CM}), y=$r$sin(θ_{CM})sin(ϕ_{CM}) and z=$r$cos(θ_{CM}), where r is the distance between the origins of the coordinate systems. The SDM and SD contour map in Fig. 2.11 can be compared by noting that θ_{CM} is the rotation about the **z** axis of the indole molecule, where positive values signify an anticlockwise rotation. As with SDMs, SD contour maps are coloured to illustrate the spatial density of the solvent molecules. Contour maps display the cosine of the angle on the **x** axis to account for the symmetry of the system. (i.e. The information contained in the region between θ_{CM}=0° and θ_{CM}=180° is the same as between θ_{CM}=0° and θ_{CM}=-180°. As cos(θ_{CM})=cos(-θ_{CM}), taking the cosine of the angle accounts for this symmetry. This procedure is only necessary for one dimension at a time.) An example of an SD contour map is included in Fig. 2.11, where a similar distribution of solvent molecules either side of the indole molecule (0.4, ±100) and in front of the the NH group (-0.4, 0) is observed as in the SDM. Unlike SDMs, SD contour maps are in no way normalised. SDMs are usually the preferred method of displaying the spatial distribution of molecules in the this thesis.

2.3.3 Whole Molecule Analysis

Whole molecule analysis (WMA) is based on the same principle as the SDMs, but rather than plotting the distribution of the origin of the coordinate system of the solvent molecules around the origin of the coordinate system assigned to the solute, or fragment of the solute, WMA allows the distribution of the origin of the solvent to be plotted within a specified distance range of any atom of the solute.[125, 277] The density of solvent molecules for these conditions can then be plotted relative to the solute with a chosen isocontour level, which may be chosen to illustrate, for instance, areas of the solute molecule that are preferentially solvated. These plots are typically not normalised and do not comprise of cut-throughs or include a scale bar. Frequently, a number of solute molecules from the system have been overlaid in the same graphic when WMA data is presented in order to highlight the range conformations adopted by the solute molecules in solution.

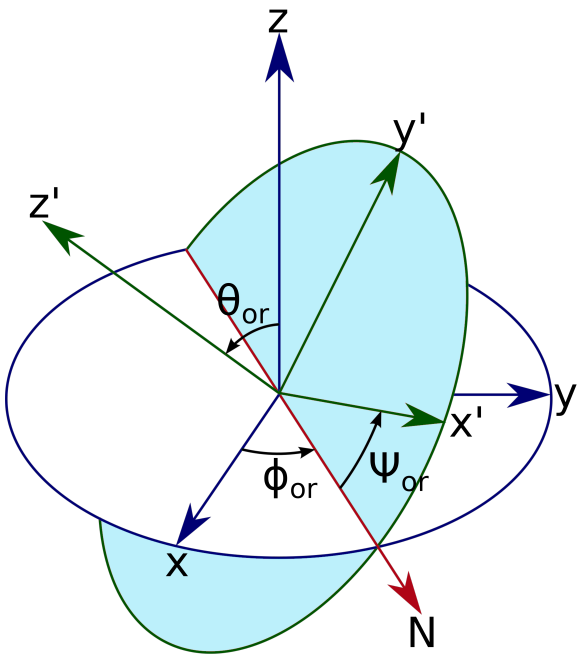


Figure 2.12: A diagram defining Euler angles, ϕ_{or}, θ_{or} and ψ_{or}, as used to determine the relative orientation of molecules in ANGULA. **N** is the line of nodes. **xyz** represents the original/fixed reference frame (dark blue) assigned to the solute molecules. $\mathbf{x'y'z'}$ represents the local coordinate system (green) assigned to the solvent molecules. Here the original/fixed reference frame and local coordinate system are overlaid, but in reality there may be any distance between their origins.

2.3.4 Orientation

In addition to the relative position of molecules, the relative orientation of molecules can also be plotted for any subset of solvent molecules (e.g. the relative orientation of the nearest neighbour waters of the solute) by means of Euler angles.[126] Essentially, the difference in the rotation of the coordinate system assigned to the solvent atoms, often termed the local coordinate system ($\mathbf{x'y'z'}$), can be computed relative to the orientation of the coordinate system assigned to the solute (**xyz**), the original/fixed reference frame, to find the relative orientation of the two molecular species. [1]

The definition of these Euler angles, denoted as ϕ_{or}, θ_{or} and ψ_{or}, is embodied by Fig. 2.12. ϕ_{or}, θ_{or} and ψ_{or} specifically represent the rotations about the **z**, **N** and $\mathbf{z'}$ axes, respectively. **N**, termed the line of nodes, is defined as the intersection between

[1]Geometrically derived Euler angle relationships: $\phi_{or} = \cos^{-1}(\frac{z'_y}{\sqrt{1-z'^2_z}}), \theta_{or} = \cos^{-1}(z'_z), \psi_{or} = \cos^{-1}(\frac{y'_z}{\sqrt{1-z'^2_z}})$, where subscripts x, y and z denote the projection of the local coordinate system ($\mathbf{x'y'z'}$) onto the **x**, **y** and **z** axes of the original reference frame.

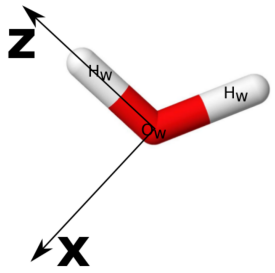


Figure 2.13: A coordinate system assigned to a water molecule for analysis using ANGULA, where the **y** axis is coming out of the plane of the page.

the **xy** and $\mathbf{x'y'}$ planes, meaning that it is perpendicular to the $\mathbf{zz'}$ plane.

There are 6 possible sequences of 3 elemental rotations required to transform from **xyz** to $\mathbf{x'y'z'}$ starting from the 2 coordinate systems being aligned. These rotations can either be intrinsic or extrinsic. Intrinsic rotations occur about the axes of the local coordinate system, whereas extrinsic rotations occur about the axes of the coordinate system of the fixed reference frame. It should be noted that Euler angles between two frames are only defined if the coordinate systems have the same handedness.

Fig. 2.14 demonstrates how the relative orientation of two molecules can be described through Euler angles. In this example, the relative orientation of a water molecule to a benzene molecule is considered. The coordinate system attached to the benzene molecule is as defined in Fig. 2.10. The coordinate system attached to the water molecule is defined in Fig. 2.13. Maintaining the convention adopted in this section, the coordinate systems attached to the benzene and water molecules may be considered the original/fixed reference frame and local coordinate system, respectively.

On the left-hand side of Fig. 2.14 a water molecule at an aribitary distance from the benzene ring (Euler angles do not provide information about distances between molecules) and a randomly chosen orientation is considered. Intially, each of ϕ_{or}, θ_{or} and ψ_{or} are 40°. These 3 Euler angles are sequentially reduced to 0°, such that the 2 coordinate systems finally become aligned. By calculating the rotations 2 coordinate sytems must undergo in terms of Euler angles to become aligned, ANGULA is able to compute the relative orientation of molecules.

Information extracted from a system regarding ϕ_{or}, θ_{or} and ψ_{or} can be plotted as a

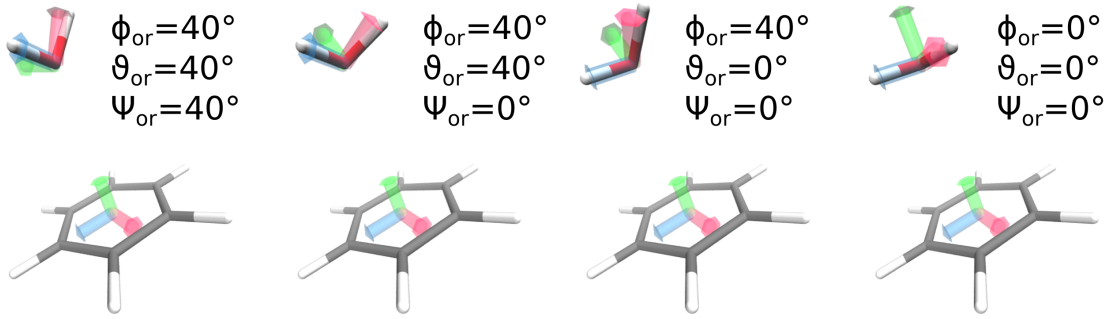


Figure 2.14: How the relative orientation of one molecule to another can be described using Euler angles. This example is for the orientation of a water molecule, whose local coordinate system is defined in Fig. 2.13, relative to a benzene molecule, whose original/fixed reference frame is is defined in Fig. 2.10. The **x**, **y** and **z** axes are red, green and blue, respectively.

frequency distribution, enabling the most probable relative orientation of molecules to be found. This data can also be represented in the form of a 2D map, in the same fashion as for the 2D contour maps showing spatial distribution (see orientational contour maps in Fig. 2.15 (B)).

2.3.5 Most Probable Position and Orientation

By combining positional and orientational information, the most probable nearest neighbour solvent molecule of a solute can be plotted.[33, 34, 125] The flow chart shown in Fig. 2.15 illustrates how this was achieved for the example of the most probable nearest neighbour water of the amine hydrogen of cocaine.

As shown in Fig. 2.15 (A), an SD contour map of the nearest neighbour water molecules of the amine hydrogen was first plotted. This SD contour map, as explained in Section 2.3.2, indicates the relative positions of the origins of the coordinate systems assigned to this fragment of the cocaine molecule and the water molecules in polar coordinates. From the SD contour map of the nearest neighbour water molecules around this hydrogen atom, the area of the highest density of nearest neighbour water molecules was then selected. Orientational contour maps corresponding to the nearest neighbour water molecules in this region were then plotted showing the Euler angles of the coordinate systems assigned to cocaine and water, as shown in Fig. 2.15 (B). Consequently, the most probable angles between the sets of axes, and hence the relative orientation of

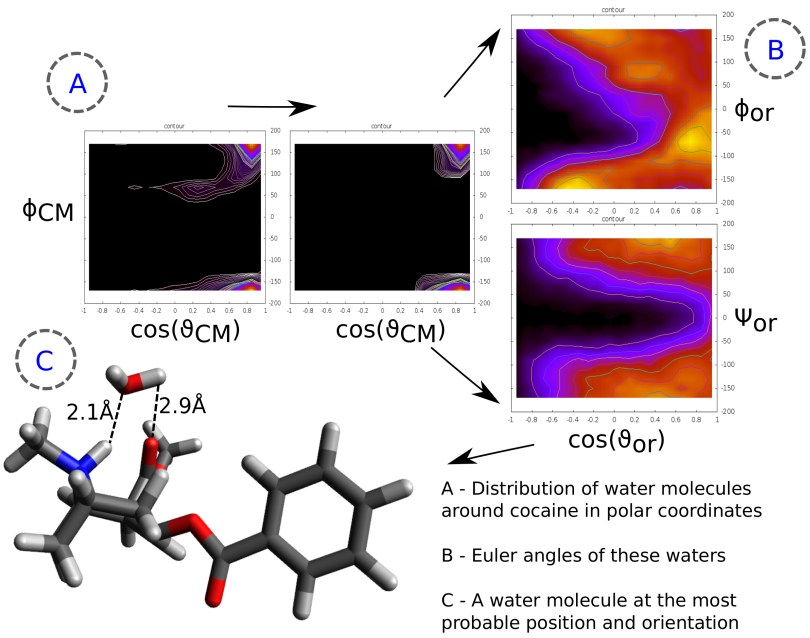


Figure 2.15: Extracting the most probable position and orientation.[125]

water molecules, was extracted. Using these values, along with the position of the peak of the RDF for the selected area of density of nearest neighbour water molecules, which provides the distance between the origins, allows Fig. 2.15 (C) to be plotted.

2.3.6 Intermolecular Association Plot

To assess the interactions between molecules of the same molecular species, particularly with view to determining how solute molecules are aggregating, an intermolecular association plot (IMAP) can be utilised. IMAPs do not tell you whether a molecule is aggregating, but which pairs of atoms are most likely to be involved in any aggregation that does occur.

Essentially, IMAPs show the likelihood any atom of the molecular species in question associating with any other atom of that same molecular species relative to all the other possible interatomic interactions for the chosen molecule. This information is presented in the form of a matrix with the number along each of the axes corresponding to an atom of the molecule. The relatively greater any association between any two atoms, the higher the probability density of the corresponding element in the matrix. The probability density is indicated by the scale bar which has arbitary units.

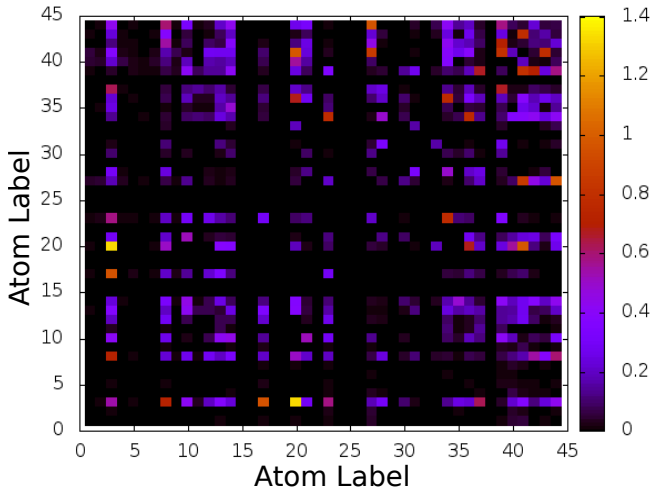


Figure 2.16: An intermolecular association plot showing the interactions between atoms of cocaine in aqueous solution.

The molecules used to produce IMAPs are selected from the simulation box by plotting an RDF that shows the distance between any atom of a solute molecule and any atom of any other solute molecule of the same molecular species. The matrix is then calculated over a distance range corresponding to the full width at half maximum of the peak in the described RDF. This avoids performing the matrix calculation for pairs of molecules at distances outside the range where any association occurs.

Chapter 3

Indole in Methanol/water Solutions

3.1 Introduction

Indole is the precursor of a number of alkaloids, a group of naturally occurring chemical compounds that includes caffeine, nicotine and cocaine, as well as the precursor of many pharmaceuticals, where its ring structure is seen in a variety of synthetic drug molecules.[24, 284] Indole is also the functional group of the amino acid tryptophan (Trp). Trp has been found to cluster preferentially at the membrane bilayer interface,[211, 295] where it is thought to anchor membrane proteins.[140, 197, 229, 245] Proteins that span or are anchored in membranes play critical roles in cell regulation, molecular transport and energy transduction.[245]

Exactly how amphiphilic proteins are anchored into cellular membranes by Trp depends on the complex environment of the membrane bilayer. It has been said that Trp favours the interfacial region over the aqueous solvent more strongly than other natural amino acid,[119, 303, 313] being mostly bound between the aqueous solvent and the headgroups of the lipid bilayer.[54, 213] Despite its role in protein structure and thereby function, details of the interactions of Trp with lipids and the aqueous solution at the membrane-water interface remain unknown.

The indole side chain governs interactions of Trp with the membrane-water interface. Indole has an quadrupole moment resulting from its aromatic, delocalized π electrons, creating regions of partial negative charge above and below its benzene ring.[61, 245]. Charged and aromatic residues, in general, have been suggested to act as interfacial anchoring residues for membrane proteins.[290] The predominantly hydrophobic structure of indole might mean that it is expected to have an inherent affinity for the lipid bilayer.[290] Although it is known that Trp locates preferentially at the membrane-water interface, the mechanism by which this binding occurs is less clear. Previous studies indicate that indole occupies a location partly exposed to a polar environment in bilayers but have not specified its exact location.[132]

Mixed hypotheses exist regarding the dominant factors that dictate indole, and hence tryptophan, binding in membranes. Some investigations suggest that cation$-\pi$ interactions govern Trp-lipid specificity,[25, 241, 306] or that the aromatic ring acts as a hydrogen bond acceptor either with water or hydroxyl groups present at the lipid-water interface.[278] Other investigations suggest that indole NH hydrogen bond donation to the surrounding lipid head groups governs Trp specificity.[53, 54] In contrast, some theories suggest that Trp residues are spatially orientated with the benzene ring sequestered into the hydrophobic domain with the indole NH group aligned so as to hydrogen bond to the membrane surface.[65, 109, 110]. Yet other investigations claim that hydrophobic interactions with the lipid tails also govern indole orientation, suggesting that hydrophobic rings are immersed in the lipid part of the bilayer while the NH moeites are free to bind with the lipid head groups.[245]

This multitude of theories means that further detailed investigation at the atomic scale is necessary to discern the physical basis underlying the preference of Trp for the interface.[148] Methanol/water solutions are used in this investigation, as, being amphiphilic, these solutions share some important characteristics of amphiphilic biological interfaces and allow an assessment of the interactions of indole with both hydrophilic and hydrophobic groups. Furthermore, as indole has a low solubility in water (1 indole: 3400 water),[163] use of methanol/water solutions means that indole can be dissolved at the high concentrations necessary for measurement by neutron diffraction. Besides the importance of understanding the interactions of indole with methanol and water

molecules in relation to gaining insight into the mechanism by which Trp anchors proteins to the interfacial region of the membrane bilayer, this investigation acts a starting point for studying drug hydration in solution and in amphiphilic environments.

3.2 Methods

3.2.1 Neutron Diffraction Sample Preparation

Indole (Fig. 3.1) was purchased from Sigma Aldrich (CAS 120-72-9). In order to remove any impurities, indole was recrystallized by dissolution in ethanol for 24 hrs in a round bottom Schlenk flask before the ethanol was removed by dynamic vacuum. 4 g of recrystallized indole was deuterated, meaning that the amine hydrogen (H_{in}) was exchanged for a deuterium, by repeating the process of dissolving indole in D_2O (Sigma Aldrich, 99.8%D) in a Schlenk flask and removing the water by dynamic vacuum three times. The water was prevented from freezing while being removed by vacccum by submerging the Schlenk flask containing the indole/water solution in a water bath heated to 303 K. Although water may be removed more easily by vacuum at higher temperatures, it was important that the temperature of the indole/water solution was not so high that the indole did not fully precipitate out of solution during the evaporation process. Deuteration of H_{in} was verified by measuring indole dissolved in $CDCl_3$ (Sigma Aldrich, 99.9%D), a solvent with no exchangeable hydrogens, using ^1H NMR to see if the NH peak was still visible in the spectrum, its absence signifying the deuteration of this group. (See Section 3.2.6 for details of NMR sample preparation.) Integration of the peaks in the spectrum indicated that H_{in} was ∼81% deuterated.

All isotopomers of methanol - CH_3OH (Sigma Aldrich), CH_3OD (Sigma Aldrich, 99.5%D), CD_3OH (Sigma Aldrich, 99.9%D) and CD_3OD (VWR, 99.8%D) - were dried over Mg, which was activated with I_2. The activation process was carried out by subliming iodine onto the side of an ampoule using a heat gun and stirring Mg turnings in the presence of I_2 for 24 hrs. Each methanol isotopomer was then refluxed over the activated Mg at 330 K for 48 hrs, having been degassed on a vacuum line prior to heating. A mineral oil heat bath was used to ensure uniform and efficient heating of each

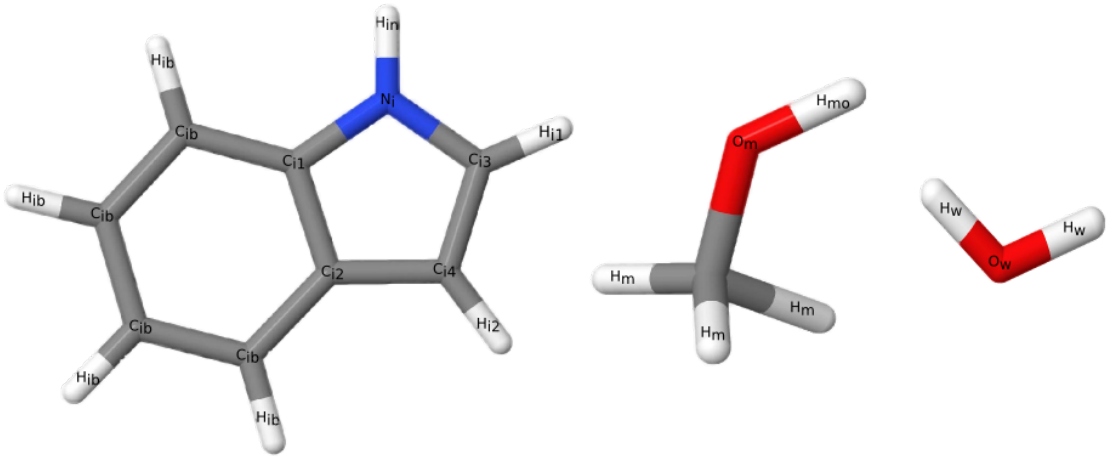


Figure 3.1: The molecular structures and labelling schemes of indole, methanol and water.

isotopomer. The pressure in the ampoule containing each isotopomer was monitored by a Pirani gauge (MKS, PDR900-1), which was accurate to 2 d.p. and had a minimum pressure reading of 1×10^{-5} Torr. Pirani gauges calculate pressure based on the principle that the heat loss of a hot object to the environment depends on the pressure of the surrounding gas.[249] After reflux, the methanol isotopomers were cryogenically distilled onto 3 Å molecular sieves, which had been pre-dried under vacuum, yielding anhydrous methanol. Sieves were stored in an oven at 493 K to ensure they remained dry. The purity of each of the isotopomers of methanol was verified by ^{1}H NMR (Section 3.2.6). The process of refluxing the methanol isotopomers and cryogenically distilling them over pre-dried molecular sieves was repeated until no contaminants were visible in the NMR spectra. H_2O was degassed using N_2 for a 1 hr period on a Schlenk line. D_2O was purchased from Sigma Aldrich for neutron diffraction measurements (99.994%D) and NMR measurements (99.98%D) and was also degassed.

Samples for neutron diffraction were prepared by mass using scales accurate to 4 d.p. under an N_2 atmosphere (<0.1 ppm O_2, <0.1 ppm H_2O). The 1.5 ml glass vials (Fisher, 1153-5884) in which the samples were prepared were sealed by crimping their aluminium lids. Gastight syringes were used to add solvents (methanol and water) to the solute through the resealable silicon/Teflon septa of the vials. Solvents were drawn from their containers using the gastight syringes under N_2 flow. The lids of the vials were wrapped with parafilm and stored in a fridge (253 K) to reduce the chance of any evaporation of the solvents. The mass of the vials before the addition of each component

Sample Name	Indole	Methanol	Water
h-indole/CH_3OH/H_2O	h-indole	CH_3OH	H_2O
h-indole/CHD_3OH/H_2O	h-indole	$CH_{1.5}D_{1.5}OH$	H_2O
d-indole/CD_3OD/D_2O	d-indole	CD_3OD	D_2O
d-indole/CHD_3OD/D_2O	d-indole	$CH_{1.5}D_{1.5}OD$	D_2O
hd-indole/CH_3OHD/HDO	$h_{0.5}d_{0.5}$-indole	$CH_3OH_{0.5}D_{0.5}$	HDO
hd-indole/CD_3OHD/HDO	$h_{0.5}d_{0.5}$-indole	$CD_3OH_{0.5}D_{0.5}$	HDO
hd-indole/CHD_3OHD/HDO	$h_{0.5}d_{0.5}$-indole	$CH_{1.5}D_{1.5}OH_{0.5}D_{0.5}$	HDO

Table 3.1: The components of the seven samples studied. The molar ratio used was 1 indole: 29 methanol: 30 water. Subscripts indicate isotopic proportions.

was recorded to be able to monitor the total mass of the vial and its contents in case of evaporation at any stage.

Seven different isotopomers of indole/methanol/water solutions for neutron diffraction were prepared at a relative molar ratio of 1 indole: 29 methanol: 30 water (∼1 M), with approximately 1.4 ml of each sample being prepared. The samples are listed in Table 3.1, where h-indole and d-indole refer to the indole molecule with an H atom and a D atom attached to N_i, shown in Fig. 3.1, respectively.

3.2.2 Neutron Diffraction Measurements

Diffraction data were collected using the SANDALS instrument (Section 2.1.2) at 298 K. The samples were transferred from the vials to the Ti/Zr canisters described in Section 2.1.3.1.2 using gastight syringes. Only the flat face of the main body of each canister was placed in the neutron beam, an area of 30 mm × 30 mm being illuminated. Both the canisters and the syringes were cleaned before use with distilled water and acetone (C_3H_6O), before being dried in a vacuum oven, to avoid contamination of the samples. Acetone is frequently used as a drying agent because of its ability to dissolve water and because of its low boiling point (∼330 K), which means that it evaporates rapidly. Diffraction data were also collected from the empty canisters, empty instrument and for a vanadium standard for reasons outlined in Section 2.1.3.1.

The μA hrs each sample was measured is shown in Table 3.2. More data were collected for samples that contained a higher proportion of hydrogen, due to the the

Sample Name	Measurement Time (μA hrs)	DCS Level
h-indole/CH_3OH/H_2O	1510	91%
h-indole/CHD_3OH/H_2O	1350	99%
d-indole/CD_3OD/D_2O	1414	104%
d-indole/CHD_3OD/D_2O	1010	101%
hd-indole/CH_3OHD/HDO	1510	93%
hd-indole/CD_3OHD/HDO	1510	94%
hd-indole/CHD_3OHD/HDO	1643	96%

Table 3.2: The measurement time and DCS level (i.e. the scattering of the neutrons by the sample as a percentage of the expected scattering.) for each sample. The average current was $\sim$170 μA.

greater coherent scattering length of deuterium and the reduction in the signal that can arise from the cross term in $F(Q)$ when samples containing atoms with scattering lengths of opposite signs are measured. Each of the empty Ti/Zr canisters was measured for 500 μA hrs. The vanadium standard was measured for 840 μA hrs. The empty instrument was measured for 250 μA hrs.

Data were corrected for absorption, multiple scattering and inelasticity effects and subsequently converted to $F(Q)$, as described in Section 2.1.3.2. The scattering from each of the samples in Table 3.2 is within $\sim$ 10% of the theoretical value. Moreover, the mean DCS levels of the samples was $\sim$ 95% of the theoretical value. The vanadium standard and each of the empty Ti/Zr canisters had DCS levels of 99%-100% and 94%-100% of their theoretical values, respectively.

3.2.3 Empirical Potential Structure Refinement

A 42.5 Å cubic EPSR simulation box was constructed which contained 25 indole molecules, 725 methanol molecules and 750 water molecules at the measured solution density (ρ =0.0912 atoms Å^{-3}) at 298 K.

The SPC/E water model[18] was used, with added σ and ϵ values for the water hydrogen taken from the OPLS all-atom (OPLS-AA) force field.[133] The SPC/E water model was chosen, as it is considered to have the best agreement with experimental bulk water dynamics (e.g. self-diffusion coefficient) and atomic structure.[174, 317] The molecular structure of methanol was generated and energetically optimised using Avogadro.[98]

Methanol seed potentials were taken from the OPLS-AA force field.[133] The methanol and water labelling schemes and potentials are shown in Fig. 3.1 and Table 3.3, respectively.

Atom Label	ϵ **(KJ mol^{-1})**	σ**(Å)**	**Atomic Mass (u)**‡	$\mathbf{q_{EPSR}(e)}$	$\mathbf{q_{MD}(e)}$
C_{i1}	0.35980	3.3997	12.011	0.0805	0.2231
C_{i2}	0.35980	3.3997	12.011	-0.1055	0.2240
C_{i3}	0.35980	3.3997	12.011	0.0805	-0.0621
C_{i4}	0.35980	3.3997	12.011	-0.1055	-0.4350
C_{ib}	0.35980	3.3997	12.011	-0.2245	-0.3099
C_{ib}	0.35980	3.3997	12.011	-0.2245	-0.1263
C_{ib}	0.35980	3.3997	12.011	-0.2245	-0.1940
C_{ib}	0.35980	3.3997	12.011	-0.2245	-0.2679
H_{i1}	0.06276	2.5106	2.000	0.1814	0.1814
H_{i2}	0.06276	2.5996	2.000	0.2091	0.2091
H_{ib}	0.06276	2.5996	2.000	0.1587	0.1713
H_{ib}	0.06276	2.5996	2.000	0.1587	0.1425
H_{ib}	0.06276	2.5996	2.000	0.1587	0.1477
H_{ib}	0.06276	2.5996	2.000	0.1587	0.1734
H_{in}	0.65690	1.0691	2.000	0.3701	0.3701
N_i	0.71130	3.2500	14.007	-0.4474	-0.4474
C_m	0.27614	3.5000	12.011	0.1450	0.1450
H_m	0.12552	2.5000	2.000	0.0400	0.0400
O_m	0.71128	3.1200	16.000	-0.6830	-0.6830
H_{mo}	0.00000	0.0000	2.000	0.4180	0.4180
O_w	0.65000	3.1660	16.000	-0.8476	-0.8476
H_w †	0.19246	0.4000	2.000	0.4238	0.4238

Table 3.3: The parameters of the indole, methanol and water atoms used in EPSR1 and EPSR2/MD. Fig. 3.2 shows how these parameters relate to the molecular structure of indole. † σ and ϵ values were adjusted for H_w in EPSR1, but remain 0 in EPSR2/MD. ‡ The atomic mass of hydrogen used in EPSR2/MD was 1.000 u.

The molecular structure of indole was generated and energetically optimised using Avogadro. Indole parameters, listed in Table 3.3, were generated with ANTECHAMBER. [300, 299] (See Section 3.2.4 for more details.) Atoms with similar charges were grouped together and the mean of their parameters used. For instance, the atoms on the outside of the benzene ring (C_{ib}), whose charges have a standard deviation of 30%, were grouped together and their mean charge taken. Atoms either side of the nitrogen were also grouped together (C_{i1} and C_{i3}) but were still labelled separately to enable analysis of the distribution of solvent atoms around indole. Similarly, C_{i2} and C_{i4} were grouped together. The charges for these two groupings of atoms were informed by previous density functional theory (DFT) calculations,[194] which attribute negative charges to the atoms directly adjacent to the amine nitrogen, reflecting the electron withdrawing character of nitrogen atoms in most compounds. Other combinations (e.g. C_{i1} and C_{i2}) result in a different polarity to that suggested by DFT calculations. How the charges of the atoms of indole relate to its molecular structure is shown in Fig. 3.2.

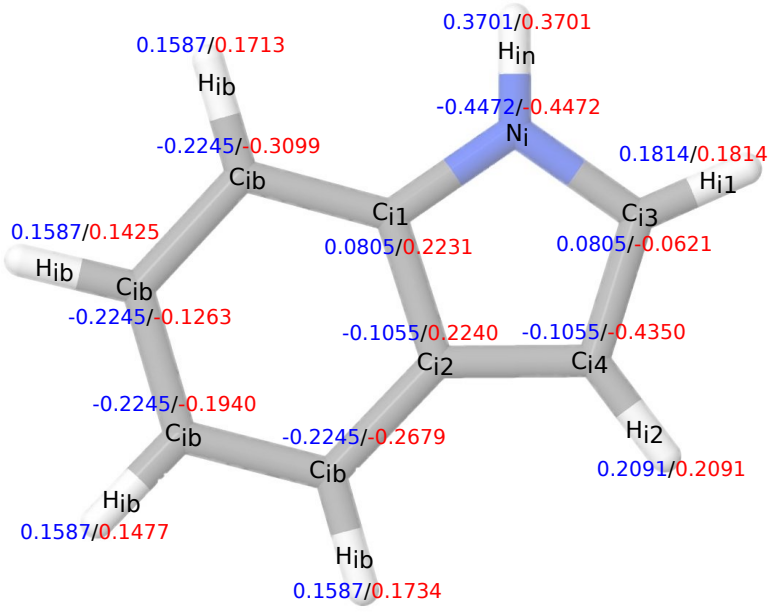


Figure 3.2: The charges of the indole atoms used in EPSR1 (blue; first number) and EPSR2/MD (red; second number) to 4 d.p.

The indole molecules were kept planar by the use of dummy atoms. ϵ, σ and q were 0 for these atoms, whereas their mass was 120.110 u. The atoms were placed above and below the the geometrical centres of each of the rings of indole, as shown in Fig. 3.3. Tests

were conducted to ensure that the solvation of indole was unaffected by the addition of these atoms, although a more physically realistic model was effected by their inclusion (Section 3.3.4.1).

Standard values were used for the atomic masses of the atoms of indole, methanol and water, save hydrogen atoms. An atomic mass of 2.000 u was used for H atoms in EPSR, as it has been shown that using an atomic mass of 1.00 u results in hydrogen atoms that are typically too mobile in EPSR.

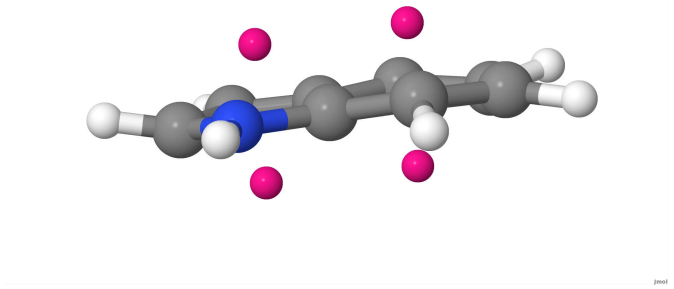


Figure 3.3: A diagram showing the location of the dummy atoms (pink) used to maintain the planarity of the indole molecules in the EPSR simulation.

This simulation, named EPSR1, was accumulated for 56690 steps. Additional EPSR simulations were performed to assess the effect of the fitting process on the results and also the effect of charge (e.g. uncharged indole molecules in methanol/water solutions).

3.2.4 Molecular Dynamics

MD simulations were run[a] using Gromacs 4.5.5 [220] with the OPLS-AA force field.[133] The SPC/E water model was used,[18] and methanol parameters were taken from the OPLS-AA force field. The topology and parameters of the indole molecule were generated in AMBER format with ANTECHAMBER[299, 300] and converted to Gromacs format using the amb2gmx.pl script.[186] Indole was optimized with Gaussian 09 at the Hartree-Fock level with the 6-31G basis set.[77] Restrained electrostatic potential (RESP) charges[15] were assigned as the atomic partial charges by optimizing the charge fitting to QM electrostatic potential (ESP) maps.

[a]These simulations were performed by Rosie Yapei Zhang, an exchange student, under the guidance of Sarah Rouse.

The initial system was generated by inserting 100 indole, 3000 water and 2900 methanol molecules into a simulation box and was energy minimised using the steepest descents integrator. A 5 ns simulation to allow water to relax with other components positionally restrained was performed using the NVT ensemble. Production simulations of 10 ns used the NPT ensemble. A timestep of 2 fs, the timescale for bond vibration, was used and bond lengths were constrained using the LINCS algorithm.[104] Van der Waals interactions were cut off at 10 Å. Periodic boundary conditions were used with Particle Mesh Ewald electrostatics cut off at 10 Å.[50] Simulations were performed at 298 K coupled to a Berendsen bath. For production simulations isotropic pressure coupling to a Berendsen barostat at 1 bar was used.[19]

A snapshot from the MD trajectory file was converted to an EPSR simulation box (EPSR2), and the starting potentials were subsequently refined by the EPSR process by fitting the same set of diffraction data as EPSR1 for 60000 steps. In this way, an EPSR simulation was run whose atomic parameters exactly matched those of the MD simulation such that comparisons could be made both between the simulation techniques and the effect of the charge distribution amongst indole atoms. Values of σ and ϵ were conserved across each of the 3 simulations for indole. The individual charges for each atom of indole as used in MD and EPSR2 are depicted in Fig. 3.2 and also listed in Table 3.3.

Additional MD simulations were run,[b] changing only the concentration of the methanol and water molecules, with indole at an infinite dilution (1 indole molecule: 1000 solvent molecules). These simulations were perfomed to assess whether the concentration of the solvent molecules affected any of the behaviour seen in the system.

3.2.5 ANGULA

ANGULA was used to study the 3D arrangements of solvent molecules around indole in each of EPSR1, EPSR2 and MD. The relative positions and orientations of solvent molecules with respect to indole were extracted for nearest neighbour and distance range analyses, which are described in Section 2.3, using ∼5000 snapshots of the simulation

[b]These simulations were run by Andres Henao Aristizabal.

box in each case.

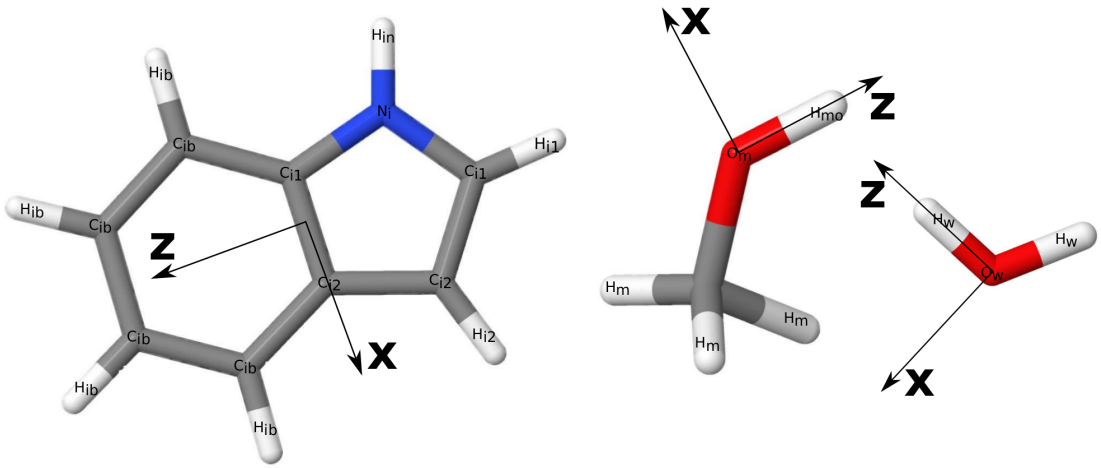


Figure 3.4: The coordinate systems assigned to indole, methanol and water using ANGULA. The **y** axis is coming out of the plane of the page.

The orthogonal coordinate systems assigned to each of the molecular species is shown in Fig. 3.4, where the origin of the indole molecule is defined as the mid-point of the carbon atoms joining the benzene and pyrrole rings. In Fig. 3.4, the origin of the solvent molecules is defined as the oxygen atoms (O_m and O_w) for both methanol and water. Analysis was also undertaken with the carbon atom (C_m) as the origin of the methanol molecules. The coordinate system with C_m at the origin is not defined here, as only SDMs were produced using this coordinate system for which only the definition of the origin is relevant. The majority of the analysis was conducted with the oxygen atoms as the origin of the solvent molecules for ease of comparison between the two molecular species.

3.2.6 NMR Sample Preparation and Measurement

NMR samples were prepared in an N_2 atmosphere when adding solute to NMR tubes. Chloroform was dried following the same procedure used to dry the isotopomers of methanol (Section 3.2.1) but was refluxed over CaH_2, rather than activated Mg, and cryogenically distilled onto 4 Å molecular sieves.

Solvents, with the exception of water, were added to the NMR tubes by vacuum transfer, using liquid nitrogen to reduce the partial pressure in the NMR tube. The

NMR tubes used were fitted with Young's taps, enabling their use on a vacccum line and limiting the possibility of exposure to the atmosphere, which could result in contamination. Water was added to the NMR tubes under N_2 flow using gastight syringes, as it is not possible to effectively transfer water using a vacuum line.

All NMR samples were prepared by mass, using a set of scales accurate to 4 d.p. An NMR tube holder was used to maintain the centre of mass of the NMR tube when measuring its mass. NMR spectra were collected on a 500 MHz spectrometer. Typically, ^{1}H spectra were recorded with 8 k points for 128 scans, the signal to noise ratio decreasing with the square root of the number of scans. A recycle delay of 1 s was used.

An NMR spectrum of indole dissolved in $CDCl_3$ was assigned using a combination of ^{1}H, ^{1}H-^{1}H COSY and ^{13}C-HSQC spectra. Indole was measured at 3 different concentrations in methanol/water solutions, including the concentration at which neutron measurements were performed, to observe if there were any concentration related effects on the spectra. Measurements of indole dissolved in methanol and indole dissolved in water were also made, enabling comparison between an aqueous and organic environment. The samples listed in Table 3.1 were measured by NMR following their measurement by neutron diffraction to confirm that the samples had not been contaminated in the Ti/Zr canisters, being transferred from the canisters to the NMR tubes with a gastight syringe under N_2 flow.

3.3 Results

3.3.1 Diffraction Data

Measured neutron diffraction data, EPSR (EPSR1 and EPSR2) fits to these data and $F(Q)$s/$G(r)$s generated from the MD trajectory are shown in reciprocal space and real space in Fig. 3.5 and Fig. 3.6, respectively. Curves showing the difference between the experimental and the simulated data are included in Fig. 3.5.

In general, there is no significant difference between the measured neutron diffraction data and the EPSR fits to these data in either reciprocal or real space. The difference

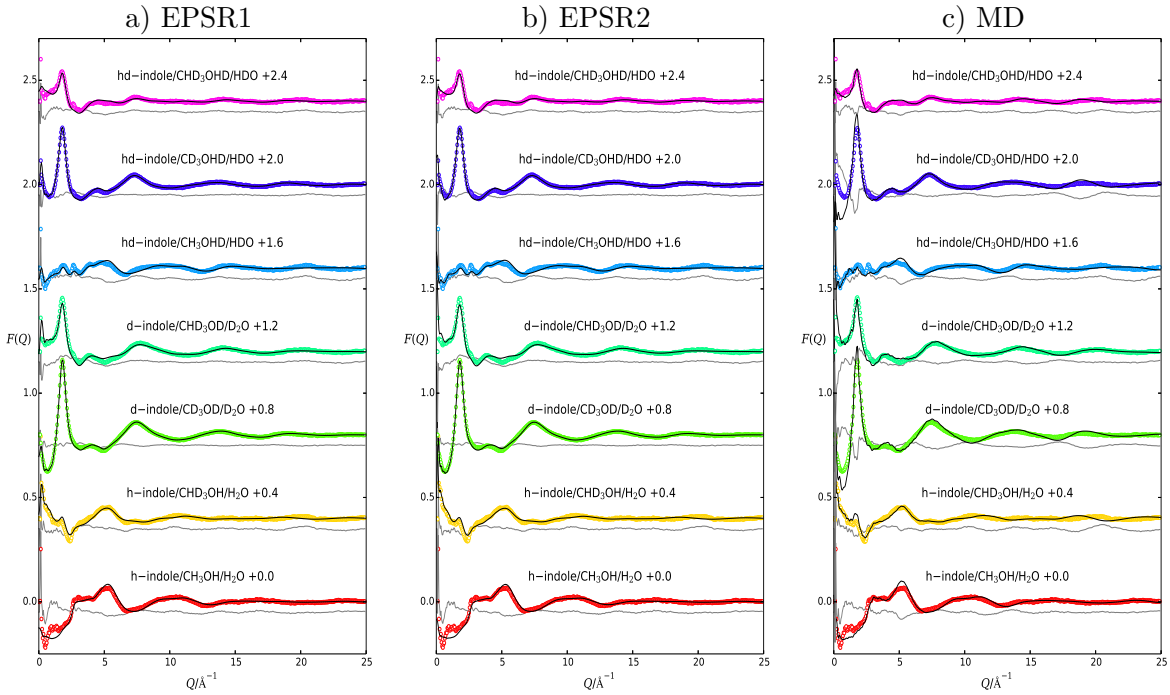


Figure 3.5: Measured neutron diffraction data (coloured lines) and EPSR fits to the data (black lines) for EPSR1 (a) and EPSR2 (b) in reciprocal space. Also shown (c) is a comparison between the data (coloured lines) and the $F(Q)$s generated from the MD simulation (black lines). The difference curves (grey lines) have been shifted by 0.05 for clarity.

curves are largely flat for $Q > 5\,\text{Å}^{-1}$ in Fig. 3.5 (a) and Fig. 3.5 (b) for each sample. Moreover, there appears to be no difference in the agreement between the measured diffraction data and the EPSR fits that is associated with the change in charge distribution (i.e. the EPSR1 and EPSR2 difference curves are almost identical), it being unclear which of EPSR1 and EPSR2 constitutes a better fit. This is unsurprising given that the majority of the scattering signal is from the solvent for which the charge distribution was unchanged between EPSR1 and EPSR2.

Both EPSR1 (Fig. 3.5 (a) and Fig. 3.6 (a)) and EPSR2 (Fig. 3.5 (b) and Fig. 3.6 (b)) show an improvement in their agreement with the measured neutron diffraction data, however, compared with the MD $F(Q)$s and $G(r)$s (Fig. 3.5 (c) and Fig. 3.6 (c)). The MD difference curves are not as flat as the difference curves in Fig. 3.5 (a) and Fig. 3.5 (b), and some of the peaks in real space, such as for samples hd-indole/CH_3OHD/HDO and hd-indole/CD_3OHD/HDO, for example, do not match the diffraction data as closely as EPSR1 and EPSR2 do in Fig. 3.6 (a) and Fig. 3.6 (b), respectively. The agreement of the simulated MD diffraction data with the experimental data is less favourable due,

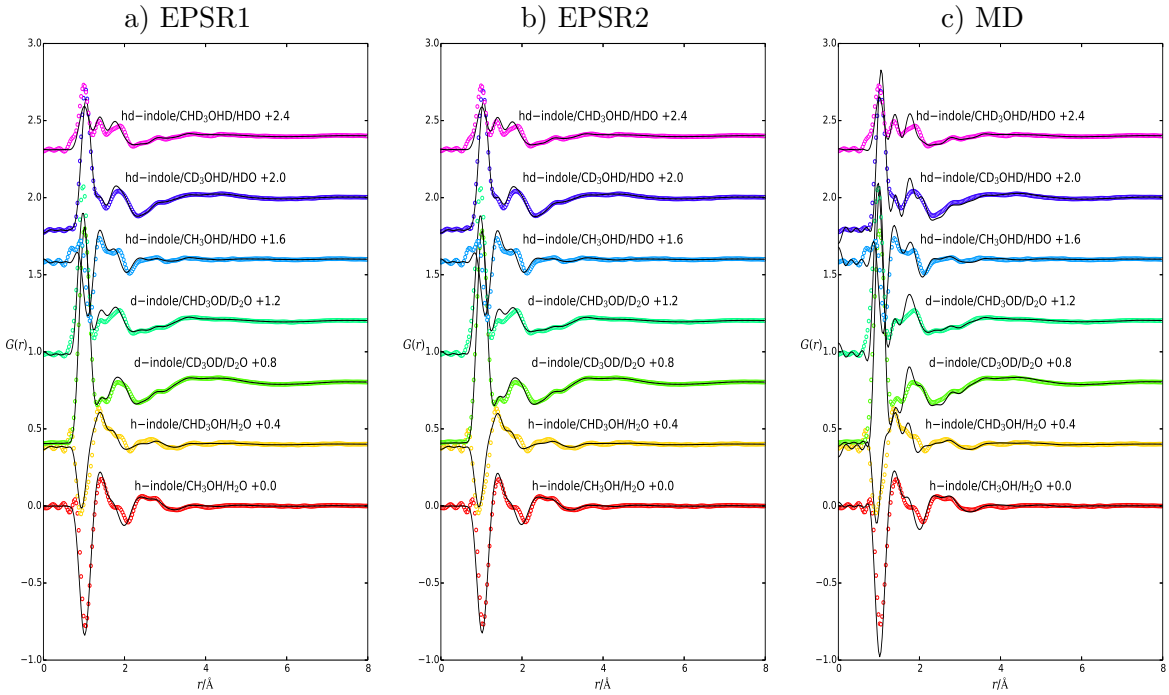


Figure 3.6: Measured neutron diffraction data (coloured lines) and EPSR fits to the data (black lines) for EPSR1 (a) and EPSR 2 (b) in real space. Also shown (c) is a comparison between the data (coloured lines) and the $G(r)$s generated from the MD simulation (black lines).

in part, to the larger distribution of intramolecular distances allowed in EPSR, which is more consistent with the measured neutron diffraction data. Nevertheless, the MD simulation produces a structure which results in simulated diffraction data similar to experimental diffraction data. For each simulation there are some differences between the simulated and the experimental data at low Q because of difficulty correcting for inelastic scattering in this region.

3.3.2 Solvent Structure

Fig. 3.7 shows the water-water and methanol-methanol RDFs extracted from the EPSR and MD simulations, with corresponding CNs being shown in Table 3.4. The CNs in this Section are computed at the peak minima (right hand column CN tables) of the corresponding RDFs. RDFs extracted from an EPSR simulation of pure water[266] and pure methanol[56] are included in Fig. 3.7 for comparison.

$g_{\mathrm{O_wO_w}}(r)$ has a first peak distance of $\sim$2.8 Å and a second peak at $\sim$4.5 Å in EPSR1,

EPSR2 and MD. The second peak position in $g_{\mathrm{O_wO_w}}(r)$ indicates that second water-water coordination shell is slightly compressed in each of the simulations performed for indole/methanol/water solutions compared with pure water, as has been previously reported for methanol/water solutions.[59] The peak positions in $g_{\mathrm{O_wH_w}}(r)$ (∼1.8 Å and ∼3.3 Å) and $g_{\mathrm{H_wH_w}}(r)$ (∼2.4 Å and ∼3.8 Å) match those of pure water for each simulation. These findings agree with previous neutron diffraction studies which revealed that the underlying water structure is not greatly affected by the presence of methanol molecules.[268] Furthermore, other neutron diffraction studies have found no evidence of structural enhancement in going from pure water to water in alcohol/water mixtures.[286]

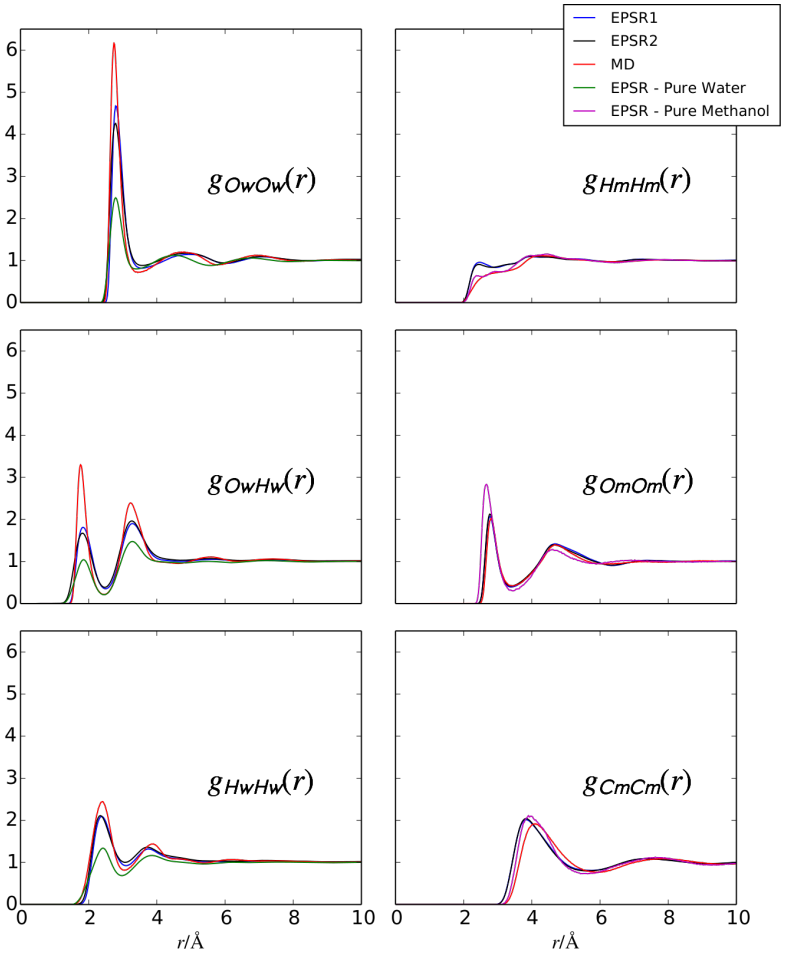


Figure 3.7: RDFs for water-water and methanol-methanol interactions extracted from EPSR1 (blue line), EPSR2 (black line) and MD (red line) simulations of indole/methanol/water. RDFs extracted from EPSR simulations of pure water[266] (green line) and pure methanol[56] (magenta line) are also included.

The methanol-methanol RDFs show peaks at ∼2.8 Å and ∼4.7 Å in $g_{\mathrm{O_mO_m}}(r)$ for each simulation. A peak at ∼3.9 Å is seen in $g_{\mathrm{C_mC_m}}(r)$ for EPSR1 and EPSR2. This

n_{α}^{β}	**EPSR1**	**EPSR2**	**MD**	r/Å
$n_{O_w}^{O_w}$	2.30	2.39	2.46	3.51
$n_{O_w}^{H_w}$	0.97	1.00	1.08	2.49
$n_{H_w}^{H_w}$	2.50	2.63	2.79	3.06
$n_{H_m}^{H_m}$	1.66	1.63	1.12	2.94
$n_{O_m}^{O_m}$	0.93	0.96	0.90	3.39
$n_{C_m}^{C_m}$	7.00	7.03	6.90	5.61

Table 3.4: CNs of the water-water and methanol-methanol RDFs shown in Fig. 3.7.

peak is at slightly greater distance of ∼4.1 Å for MD. $g_{H_mH_m}(r)$ shows a peak at ∼2.5 Å for each simulation, although this peak is sharper for EPSR1 and EPSR2 compared with MD. The positions of these peaks in the RDFS are in agreement with the peak positions previously reported for pure methanol and methanol/water solutions as determined by neutron diffraction.[6, 56] The methanol-methanol RDFs from EPSR1, EPSR2 and MD also agree with *ab initio* MD studies of pure methanol.[94]

The good agreement between the RDFs extracted from EPSR and MD simulations is in line with other studies which cite qualitatively similar results between EPSR and MD simulations of methanol/water solutions over a range of concentrations, temperatures and pressures.[60] Similar CNs are observed for each simulation in Table 3.4, although there is a slight trend for the water-water CNs to increase from EPSR1 to EPSR2 to MD. Changes in CNs are most likely associated with the difference in the propensity of water to interact with indole in each of the simulations, which is discussed in Section 3.3.3. Previous investigations using X-ray emission spectroscopy have shown that most methanol and water molecules in methanol/water solutions persist in structures similar to the ones found in the pure liquids (i.e. methanol and water are microsegregated).[91] Results from IR spectroscopy measurements and MD simulations of acetone/water and acetonitrile/water solutions indicate similar behaviour of solvent molecules in these mixtures as is found here for methanol and water molecules in methanol/water solutions.[292] Both components in water/DMSO mixtures have also been found to tend to preserve their structural order upon mixing.[270, 288]

Fig. 3.8 shows the RDFs for the solvent-solvent hydrogen bonding interactions between methanol and water molecules in the present solution. The first peak positions in

all of the RDFs in this figure show that, on average, H_{mo} and H_w form bonds at ∼1.8 Å to O_m and O_w. Coupled with the CNs shown in Table 3.5, this indicates that both the solvents have a similar propensity to interact with each other as with themselves.

The full hydrogen bonding potential in this system should theoretically equal one in the absence of any solute for both H_w and H_{mo} atoms. However, in the case of EPSR1, for water $n_{H_w}^{O_w}+n_{H_w}^{O_m} = 0.49 + 0.38 = 0.87$, indicating that the system is slightly undersaturated with respect to H_w hydrogen bonding. The equivalent figure for EPSR2 is 0.88. To a lesser degree, again in the case of EPSR1, the system is also undersaturated with respect to methanol H_{mo} hydrogen bonding: $n_{H_{mo}}^{O_w}+n_{H_{mo}}^{O_m} = 0.49 + 0.44 = 0.93$. The equivalent figure for EPSR2 is 0.94. The same analysis using the CNs extracted from MD yields $n_{H_w}^{O_w}+n_{H_w}^{O_m} = 0.54 + 0.42 = 0.96$ and $n_{H_{mo}}^{O_w}+n_{H_{mo}}^{O_m} = 0.5 + 0.42 = 0.92$. These numbers show that in MD, unlike in either EPSR1 or EPSR2, the hydrogen bonding potential of the water molecules is almost fulfilled *via* solvent-solvent interactions. EPSR1 and EPSR2 are ∼10% less saturated than MD. Although the EPSR2 system is marginally more saturated than EPSR1 with respect to H_w hydrogen bonding potential, the differences in saturation as a result of the charge distribution on the indole molecule appear to have less of an effect than the simulation technique employed. The saturation of the system with respect to methanol remains practically unchanged between EPSR and MD. Despite the aforementioned similar propensity for interaction, Table 3.5 underlines a slight overall preference for bonding to water over methanol.

n_α^β	**EPSR1**	**EPSR2**	**MD**	r**/Å**
$n_{H_w}^{O_w}$	0.49	0.50	0.54	2.49
$n_{H_w}^{O_m}$	0.38	0.39	0.42	2.55
$n_{H_{mo}}^{O_w}$	0.49	0.49	0.50	2.55
$n_{H_{mo}}^{O_m}$	0.44	0.45	0.42	2.61

Table 3.5: CNs of the hydrogen bonding RDFs shown in Fig. 3.8.

Comparing the RDFs extracted from simulations of indole in methanol/water solutions with those extracted from pure water and pure methanol indicates that the nature of the hydrogen bonding between atoms of the same solvent is maintained from the pure liquid systems. That MD simulation results in sharper peaks in the extracted RDFs shown in Fig. 3.7 and Fig. 3.8 compared with the RDFs extracted from EPSR1 and

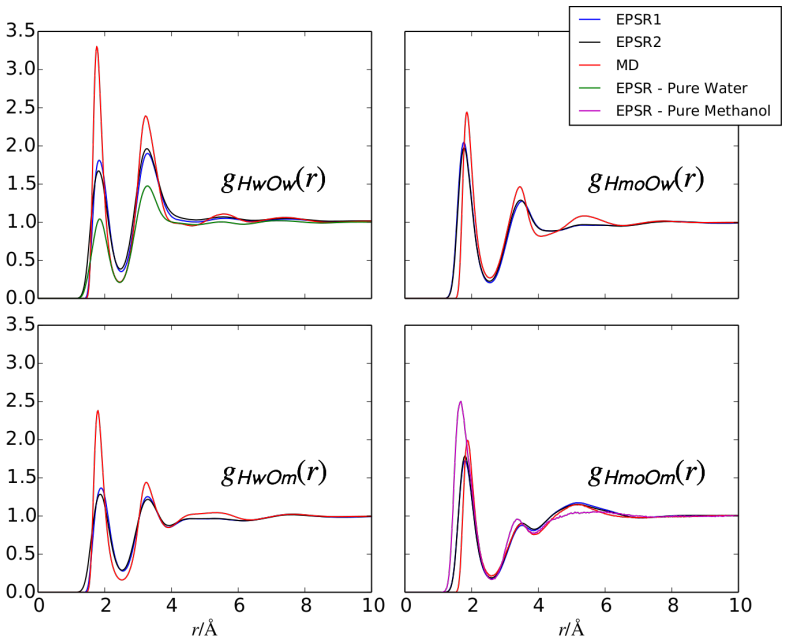


Figure 3.8: RDFs for hydrogen bonding of water (O_w and H_w) and methanol (O_m and H_{mo}) atoms extracted from EPSR1 (blue line), EPSR2 (black line) and MD (red line). Corresponding RDFs extracted from EPSR simulations of pure water[266] (green line) and pure methanol[56] (magenta line) are also included.

EPSR2 is likely due to the initial restrictions placed on the molecular configuration in MD, which results in them having a more well defined configuration.

3.3.3 Indole Solvation

3.3.3.1 Radial Distribution Functions

The hydrogen bonding interactions between the indole amine group and the solvent oxygens are shown in Fig. 3.9. Similar to Fig. 3.8 and Fig 3.7, the first peak positions in the RDFs are similar between the two solvents across each of the simulations. Both $g_{H_{in}O_w}(r)$ and $g_{H_{in}O_m}(r)$ show the presence of a hydrogen bond at ∼2.1 Å from the amine hydrogen to both water and methanol. Hydrogen bonds between 1.5 Å and 2.2 Å, such as those indicated by the peaks in $g_{H_{in}O_w}(r)$ and $g_{H_{in}O_m}(r)$ in Fig. 3.9, have been described as mostly electrostatic.[274] $g_{N_iO_w}(r)$ and $g_{N_iO_m}(r)$ both show peaks at ∼3.0 Å, which is consistent with a linear NH bond length of ∼0.9 Å. Comparing these numbers with previous investigations, the RDF peak corresponding to hydrogen bonding between the amine hydrogen of dopamine hydrochloride and water oxygens

was found at 1.75 Å,[36] and investigation of the hydration of the tripeptide glycyl-L-prolyl-glycine-NH_2 has shown that the average distance of the amine nitrogen to water oxygens is 2.9 Å.[32] CNs for $g_{\mathrm{H_{in}O_w}}(r)$ and $g_{\mathrm{H_{in}O_m}}(r)$ (Table 3.6) indicate that ∼50% of H_{in} donate hydrogen bonds to water while the other half of indole hydrogen bonding occurs between H_{in} and methanol, with no preference of binding to either solvent.

Despite these similarities, in $g_{\mathrm{H_{in}O_w}}(r)$ some slight differences start to emerge between the simulations. The second peak at 4.6 Å in $g_{\mathrm{H_{in}O_w}}(r)$ is more prominent for EPSR1, followed by EPSR2 and MD. This peak could be interpreted as the second coordination shell of water molecules around indole. However, it is also possible that this peak is a result of other interactions, unassociated with the NH group, that manifest themselves in this RDF (*vide infra*). Differences in the hydration of the amine group are also witnessed in $g_{\mathrm{N_iO_w}}(r)$, where the second peak in the RDF at 4.2 Å is again most prominent for EPSR1. Accordingly, $\mathrm{n}_{\mathrm{N_i}}^{\mathrm{O_w}}$ decreases by ∼18% from EPSR1 to MD. In contrast, CNs indicating methanol solvation of the amine group change by no more than ∼7% from EPSR1 to MD. The hydration of indole amine group in methanol/water solutions is comparable with the degree of hydrogen bonding experienced by the amine groups of other small molecules in solution. For example, ∼0.6 hydrogen bonds were found to form between water oxygens and proline amine hydrogens in aqueous solution,[182] compared with ∼0.5 in the present solutions.

Interestingly, indole, in the absence of methanol, is sparingly soluble in water (1 indole: 3400 water), while it is highly soluble in pure methanol. Even though indole-water hydrogen bonding is somewhat larger than intuitively expected for indole amine-water hydrogen bonds, it is unlikely that this bonding is what drives indole solubility in the present solutions.

$\mathrm{n}_{\alpha}^{\beta}$	**EPSR1**	**EPSR2**	**MD**	r**/Å**
$\mathrm{n}_{\mathrm{H_{in}}}^{\mathrm{O_w}}$	0.50	0.49	0.51	2.73
$\mathrm{n}_{\mathrm{H_{in}}}^{\mathrm{O_m}}$	0.48	0.52	0.48	2.73
$\mathrm{n}_{\mathrm{N_i}}^{\mathrm{O_w}}$	0.74	0.69	0.61	3.42
$\mathrm{n}_{\mathrm{N_i}}^{\mathrm{O_m}}$	0.65	0.70	0.66	3.51

Table 3.6: CNs of the indole/methanol/water RDFs shown in Fig. 3.9.

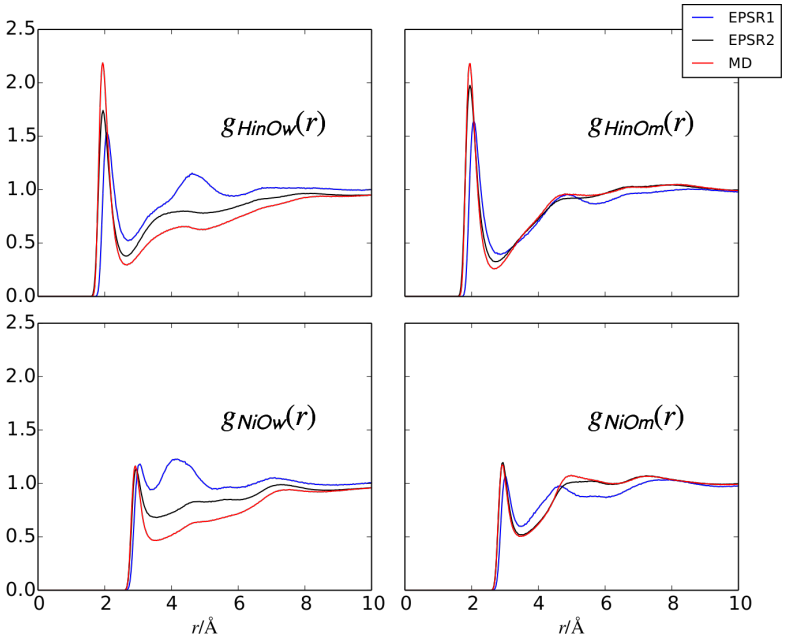


Figure 3.9: RDFs for interactions of the indole amine group (H_{in} and N_i) with solvent oxygens (O_w and O_m) extracted from EPSR1 (blue line), EPSR2 (black line) and MD (red line).

RDFs in Fig. 3.10, which are interactions between carbon atoms on the exterior of the benzene ring of indole and solvent atoms, show clearer differences between the three simulations. For $g_{C_{ib}O_w}(r)$, the first peak position varies across the simulations from ∼3.3 Å in EPSR1, ∼3.7 Å in EPSR2 and ∼3.8 Å in MD. The relative size and sharpness of this peak, especially in EPSR1, suggests that these waters are interacting directly with the benzene ring. The largest peak height is seen in EPSR1, with the density of atoms present in this peak increasing from MD to EPSR2 to EPSR1. This is reflected in the CNs (Table 3.7) for this function, (1.29, 1.84 and 2.94, for MD, EPSR2 and EPSR1, respectively) which indicate an increase of 129% in the average number of O_w atoms around C_{ib} atoms in EPSR1 compared with MD. The same trend of decreasing CNs moving from EPSR1 to EPSR2 to MD is also seen for $g_{C_{ib}H_w}(r)$. However, the first peak in $g_{C_{ib}H_w}(r)$ is less defined for either MD or EPSR2 than their first peaks in $g_{C_{ib}O_w}(r)$. CNs for EPSR1 show that that each benzene ring, on average, forms approximately 1 hydrogen bond with water at ∼2.1 Å, while EPSR2 and MD show that benzene forms hydrogen bonds with water around 20-30% as frequently.

Differences between EPSR1 and EPSR2 are most likely linked to differences in charges for the seed potentials, as, for instance, a positive charge on C_{i2} in EPSR2 would

result in a less favourable interaction between the benzene ring and water molecules. However, as EPSR2 and MD use identical charges, the greater hydration of the benzene ring in EPSR2 compared with MD is clearly a function of the EPSR fitting process which adapts the input seed potentials to be consistent with a set of diffraction data. As discussed in Section 3.2.3, the charge distribution used for EPSR1 is consistent with the electron withdrawing character of N and previous DFT studies.[194]

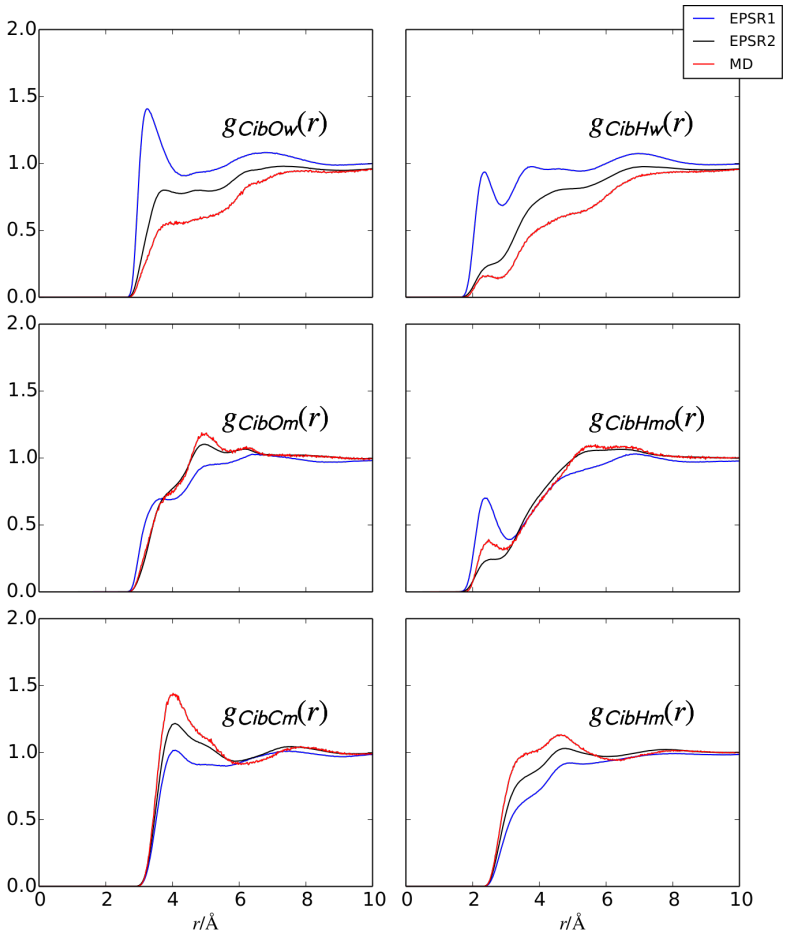


Figure 3.10: RDFs for interactions of the carbon atoms on the exterior of the benzene ring (C_{ib}) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR1 (blue line), EPSR2 (black line) and MD (red line).

RDFs in Fig. 3.10 for interactions between C_{ib} atoms and methanol atoms show that the position of the peaks in each function and the form of the functions is largely preserved between the three simulations. However, in this case, rather than solvation being at a maximum in EPSR1 and a minimum in MD, here the MD simulation shows the greatest number of interactions between methanol atoms and the benzene ring followed by EPSR2 and EPSR1. In fact, the average number of O_m atoms around C_{ib} atoms is 12% greater in MD than in EPSR1. The exception to this trend is $g_{C_{ib}H_{mo}}(r)$, where in

EPSR1 there are around twice as many H_{mo} atoms around the benzene ring as in EPSR2 and MD. The peak in $g_{C_{ib}H_{mo}}(r)$, located at ∼2.4 Å, is also more well defined than the other C_{ib}-methanol RDFs, certainly when compared to $g_{C_{ib}H_m}(r)$ (∼4.8 Å) and $g_{C_{ib}O_m}(r)$ (∼3.7 Å). Comparing the peak positions in $g_{C_{ib}O_m}(r)$ (∼3.7 Å) and $g_{C_{ib}C_m}(r)$ (∼4.1 Å) shows that, although there are some differences in the amount of solvent molecules at these peak positions across the three simulations (see CNs in Table 3.7), the methanol -OH group is consistently closer to the benzene ring of indole than the -CH_3 group, as is the water oxygen (3.3 Å-3.8Å).

n_α^β	**EPSR1**	**EPSR2**	**MD**	***r*/Å**
$n_{C_{ib}}^{O_w}$	2.94	1.84	1.29	4.47
$n_{C_{ib}}^{H_w}$	1.15	0.30	0.20	2.88
$n_{C_{ib}}^{O_m}$	5.40	6.01	6.04	5.73
$n_{C_{ib}}^{H_{mo}}$	0.41	0.19	0.26	3.03
$n_{C_{ib}}^{C_m}$	6.64	7.66	8.06	6.06
$n_{C_{ib}}^{H_m}$	23.30	25.57	26.76	6.24

Table 3.7: CNs of the indole/methanol/water RDFs shown in Fig. 3.10.

To ensure that the behaviour of the solvent molecules deduced from Fig. 3.10 is indeed related to interactions with the benzene ring, RDFs were also plotted showing the interaction between all six atoms which comprise the benzene ring of indole and the solvent molecules. These RDFs are displayed in Fig. 3.11, using the notation of C_B to indicate interactions with all six atoms of the benzene ring. Also shown in are RDFs for the interactions of all the atoms of the pyrrole ring (C_P) with solvent oxygens for EPSR1, EPSR2 and MD, so that the interactions between the solvent molecules and both of the rings may be compared.

Mirroring what was seen in Fig. 3.10, EPSR1, EPSR2 and MD each show peaks at a distance less than 4 Å in $g_{C_BO_w}(r)$ (∼3.3 Å , ∼3.9 Å and ∼3.9 Å for EPSR1, EPSR2 and MD, repectively). The presence of water oxygens at this distance indicates that waters are indeed preferentially located around the benzene ring of indole. Conversely, for the methanol oxygen, while $g_{C_BO_m}(r)$ shows the presence of methanol oxygens around indole at slightly less than 4 Å , the more prominent interactions occur over a larger distance range, evidenced by the peak at around 5 Å. Additionally, $g_{C_BC_m}(r)$, which

effectively is a measure of the hydrophobic-hydrophobic distances between methanol and the benzene ring of indole, displays a peak at at the same position as in Fig. 3.10 (∼4.1 Å). This indicates that these non-polar, hydrophobic interactions take place over a larger length scale than the interactions between the benzene ring and water molecules, as discussed.

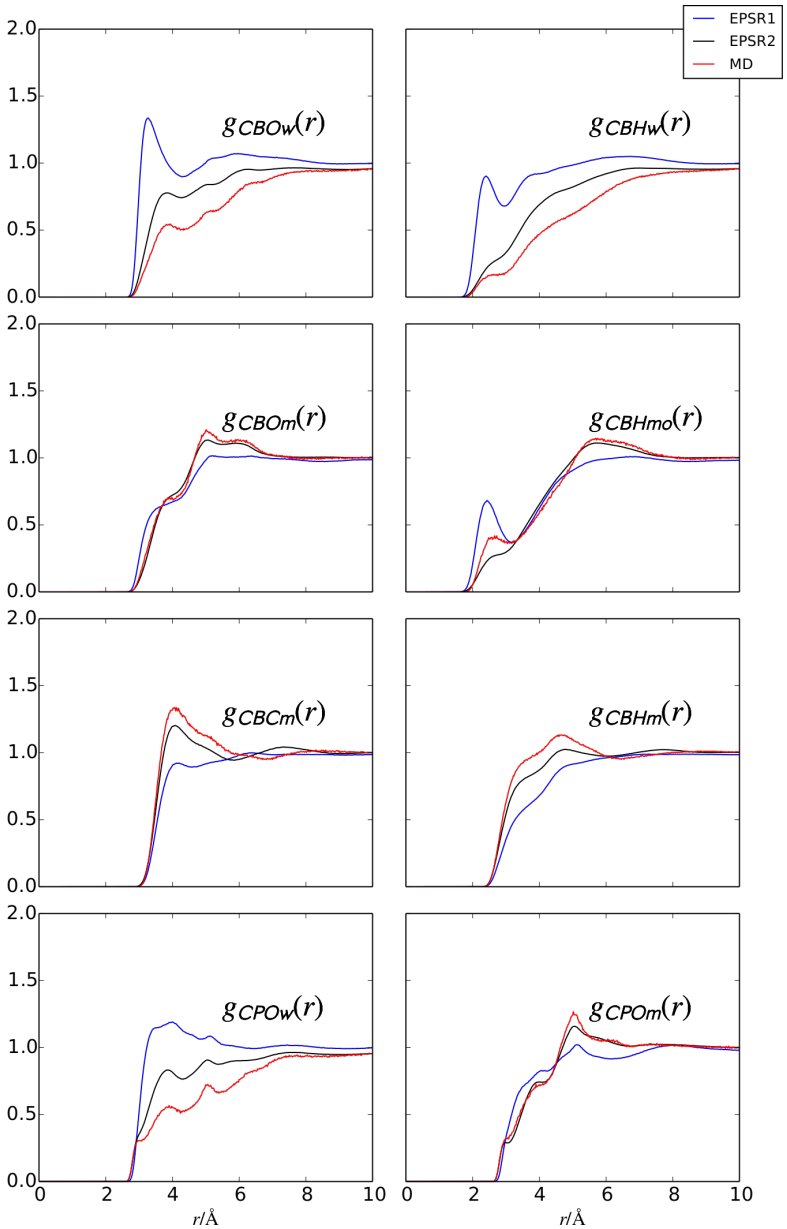


Figure 3.11: RDFs for interactions of the carbon atoms of the benzene ring (C_B - the weighted average of C_{ib}, C_{i1} and C_{i2}) and atoms of the pyrrole ring (C_P - the weighted average of C_{i1}, C_{i2}, C_{i3}, C_{i4} and N_i) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR1 (blue line), EPSR2 (black line) and MD (red line).

Computing the respective CNs for $g_{C_BO_w}(r)$, $g_{C_BO_m}(r)$ and $g_{C_BC_m}(r)$ at 4 Å yields

n_{α}^{β}	**EPSR1**	**EPSR2**	**MD**	r**/Å**
$n_{C_B}^{O_w}$	2.74	1.68	1.16	4.44
$n_{C_B}^{H_w}$	1.16	0.36	0.28	2.97
$n_{C_B}^{O_m}$	10.54	11.49	11.53	6.81
$n_{C_B}^{H_{mo}}$	0.47	0.24	0.32	3.15
$n_{C_B}^{C_m}$	10.06	10.76	11.07	6.69
$n_{C_B}^{H_m}$	22.22	25.34	27.10	6.24
$n_{C_P}^{O_w}$	3.03	2.30	1.63	4.65
$n_{C_P}^{O_m}$	10.24	11.02	11.05	6.75

Table 3.8: CNs of the indole/methanol/water RDFs shown in Fig. 3.11.

1.8, 0.9 and 0.7 for EPSR1, 0.7, 0.7 and 1.0 for EPSR2 and 0.6, 0.8 and 1.0 for MD. The total number of hydroxyl interactions for EPSR1 (2.7) is almost double the number of hydroxyl interactions for MD (1.4) at 4 Å. In each simulation solvent -OH groups are closer to the benzene ring of indole than the -CH_3 group of methanol, though these interactions are more common in EPSR1 and EPSR2 than MD. In MD benzene-CH_3 interactions are more dominant than benzene-OH interactions, but interactions of the benzene ring with water and methanol oxygens still occur over shorter distances than interactions of the benzene ring with C_m.

The greater sharpness of the peaks in $g_{C_{ib}O_w}(r)$ and $g_{C_BO_w}(r)$ compared with $g_{C_PO_w}(r)$ points towards a more well defined orientation of water molecules about the benzene ring than the pyrrole ring. In fact, it may be suggested that some of the density of water molecules in $g_{C_PO_w}(r)$ peak is merely a reflection of the water molecules around the attached benzene ring. However, this density is more likely related to the interactions between the amine group, which is closer to the geometrical centre of the pyrrole ring of indole than the benzene ring, and water molecules. The small peak in each of the simulations at ∼5.1 Å could be attributed to the interactions between water molecules and the benzene ring. In $g_{C_PO_w}(r)$, there is minimal evidence of any enhanced ordering after the first hydration shell, a theme revisted in Section 3.3.3.3. Similarly, in $g_{C_PO_m}(r)$ small peaks at ∼5.1 Å could be due to interactions between the benzene ring and methanol molecules, with density at shorter distances being related to interactions between methanol and the indole nitrogen.

3.3.3.2 Orientation

Fig. 3.12 shows the average orientations of methanol and water molecules around the benzene ring of indole for EPSR1 and MD, which, based on the preceding analysis of indole hydration, encompasses the full range of solvent behaviour in the present simulations. Corresponding RDFs ($g_{\mathrm{C_BH_w}}(r)$, $g_{\mathrm{C_BH_{mo}}}(r)$ and $g_{\mathrm{C_BH_m}}(r)$) are shown in Fig. 3.11.

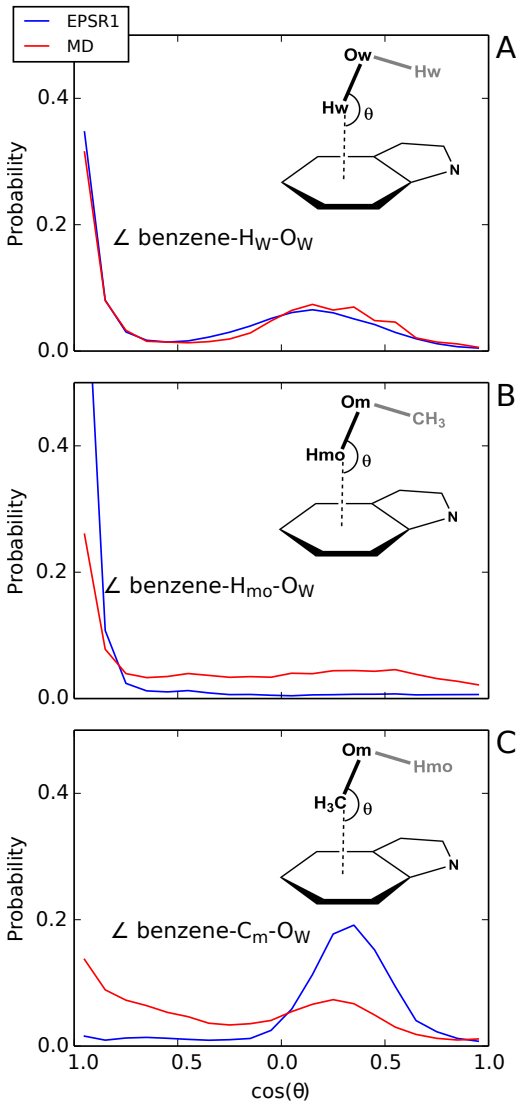


Figure 3.12: Angular orientation of solvent molecules within 5 Å of the origin of indole, where (A) depicts the benzene-H_w-O_w angle, (B) the benzene-H_{mo}-O_m angle and (C) the benzene-O_m-C_m angle for EPSR1 (blue) and MD (red).[125]

In both EPSR1 and MD, the first peak in the benzene-H_w-O_w angle distribution (Fig. 3.12 (A)) is almost linear, while a second peak reaches a maximum at $\sim \cos(0.25)$ ($\sim 75^o$). This broad peak is related to the second O_w-H_w interaction from the nearest

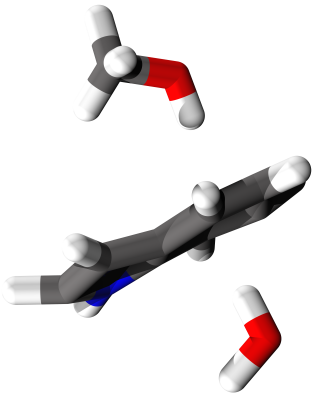


Figure 3.13: A representative snapshot of the orientation of methanol and water molecules relative to indole selected from the EPSR1 simulation box.

neighbour waters and suggests that this second covalently bound hydrogen is freely rotating around the hydrogen bond formed by the benzene $\cdots$ O_w-H_w motif. The first peak in the benzene-H_{mo}-O_m angle distribution (Fig. 3.12 (B)) follows the form of the first peak in the benzene-O_w-H_w angle distribution for both EPSR1 and MD. This similarly indicates a benzene $\cdots$ O_m-H_{mo} motif. Analogous to the second O_w-H_w interaction, the first peak benzene-C_m-O_m angle distribution (Fig. 3.12 (C)) indicates that the CH_3 group of methanol has a range of orientations relative to the benzene ring in EPSR1 and MD.

Informed by the angular orientation information shown in Fig. 3.12, a representative snapshot of both water and methanols solvating indole in the EPSR1 simulation box was chosen. This is displayed in Fig. 3.13 and illustrates how the methanol and water molecules solvate the indole molecule through interactions between their hydroxyl groups and the centre of the benzene ring of indole.

3.3.3.3 Spatial Density Maps

SDMs showing the distribution of nearest neighbour water molecules relative to indole in Fig. 3.14, plotted using the coordinate systems assigned to each of the molecules as described in Section 3.2.5, emphasise deductions made in Section 3.3.3.1 about the hydration of the benzene ring. Explicitly, Fig. 3.14 underlines the greater density of

water molecules above and below the plane benzene ring in EPSR1 compared with EPSR2 and MD, as is reflected in Fig. 3.10, Fig. 3.11, Table 3.7 and Table 3.8. In each simulation, a similar level of density is seen around the amine group, consistent with the RDFs displayed in Fig. 3.9 and the CNs in Table 3.6. The maximum density of water oxygens in the nearest neighbour SDMs generated from EPSR1, EPSR2 and MD is 4.02 $Å^{-3}$, 0.42 $Å^{-3}$ and 0.21 $Å^{-3}$, respectively. Despite differences in the number of water molecules interacting with indole across each of the simulations, the location of the water molecules is consistent.

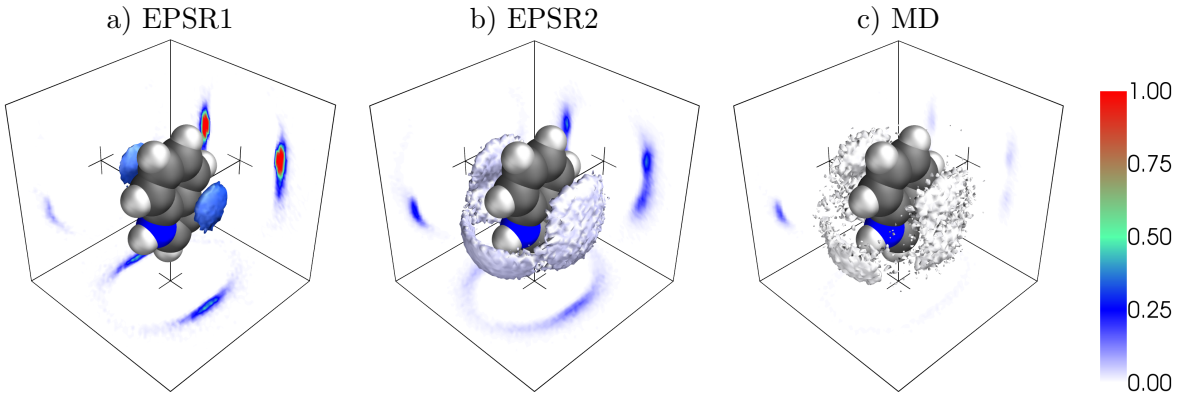


Figure 3.14: SDMs showing the distribution of nearest neighbour water molecules relative to indole generated from EPSR1 (a), EPSR2 (b) and MD (c).

In Fig. 3.15, the similar tendency for methanol to interact with the amine group of indole as water that was previously deduced from the RDFs extracted from EPSR1, EPSR2 and MD (Fig. 3.9) and the corresponding CNs (Table 3.6) is also reflected in the indole-methanol SDMs generated from each of the simulations. Similarly, the density of methanol around the benzene ring of indole, where there is little difference between EPSR and MD simulations, and the nature of the solvation of the molecule, appear largely unaffected by changes in charge distribution. As with the hydration of the benzene ring, methanol solvation seems to occur above and below the plane of the benzene ring of indole. These findings are in broad agreement with Fig. 3.11, where, although indole-methanol interactions vary in their frequency, differences are not as pronounced for methanol solvation as hydration. The narrower range of maxmimum methanol oxygen density compared with maximum water oxygen density - 1.24 $Å^{-3}$ (EPSR1), 0.53 $Å^{-3}$ (EPSR2) and 1.10 $Å^{-3}$ (MD) - emphasises the less discernible differences between each of the SDMs in Fig. 3.15 compared with Fig. 3.14.

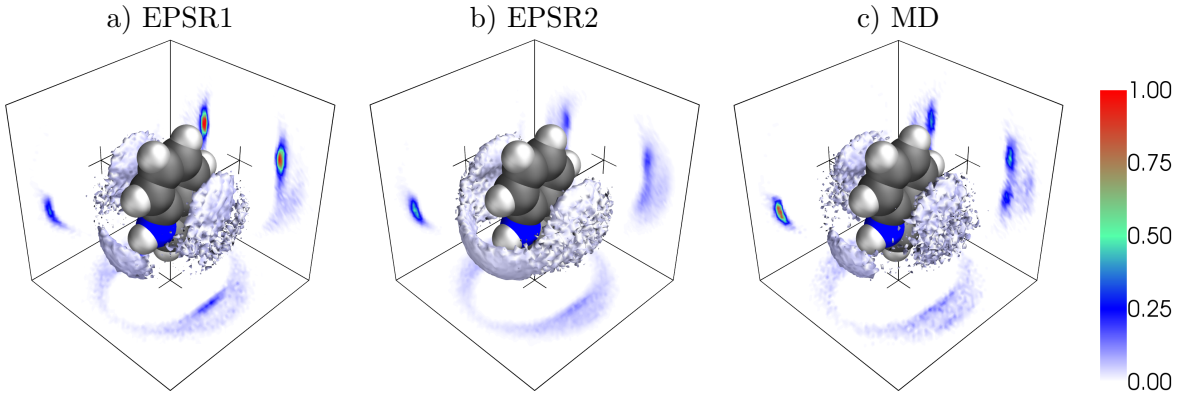


Figure 3.15: SDMs showing the distribution of nearest neighbour methanol molecules relative to indole generated from EPSR1 (a), EPSR2 (b) and MD (c).

Additional to the SDMs in Fig. 3.14 and Fig. 3.15 showing the nearest neighbour water and methanol molecules of indole, Fig. 3.16 and Fig. 3.17 show SDMs generated from EPSR1 and MD, respectively, using a distance range analysis. Having established the general trend for the solvation of indole by inspection of SDMs in Fig. 3.14 and Fig. 3.15, Fig. 3.16 and Fig. 3.17 are limited to showing the distribution of water and methanol around indole for EPSR1 and MD. As well as conducting this analysis with the same coordinate systems assigned to each of the molecules as in Fig. 3.14 and Fig. 3.15, also included in Fig. 3.16 and Fig. 3.17 are SDMs where the origin of the methanol molecule was the carbon atom. A comparison between the distribution of methyl groups around indole and the hydroxyl groups around indole can thus be undertaken.

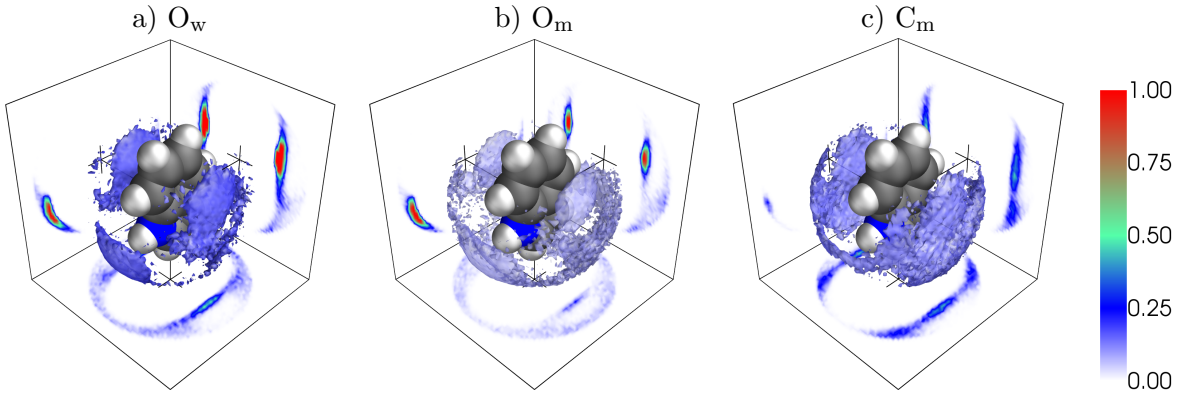


Figure 3.16: SDMs showing the distribution of solvent molecules within 5 Å of indole generated from EPSR1. The origin of the solvent molecules was defined as the water oxygen (a), the methanol oxygen (b) and the methanol carbon (c).

Unsurprisingly given the nearest neighbour indole-water SDMs, the distance range SDMs showing the distribution of water molecules relative to indole reveal a propensity

for water to hydrogen bond to the amine group of indole. These distance range SDMs further emphasise that water is preferentially located around the middle of benzene ring, above and below its plane. Differences in the degree of hydration of indole between EPSR1 and MD seen for a nearest neighbour analysis are also apparent for the SDMs produced using a distance range analysis: the hydration of the benzene ring is greater in EPSR1 (Fig. 3.16 (a)) than in MD (Fig. 3.17 (a)). In fact, the maximum density of water oxygens in Fig. 3.16 (a) is 4.65 $Å^{-3}$ compared with 2.58 $Å^{-3}$ in Fig. 3.17 (a).

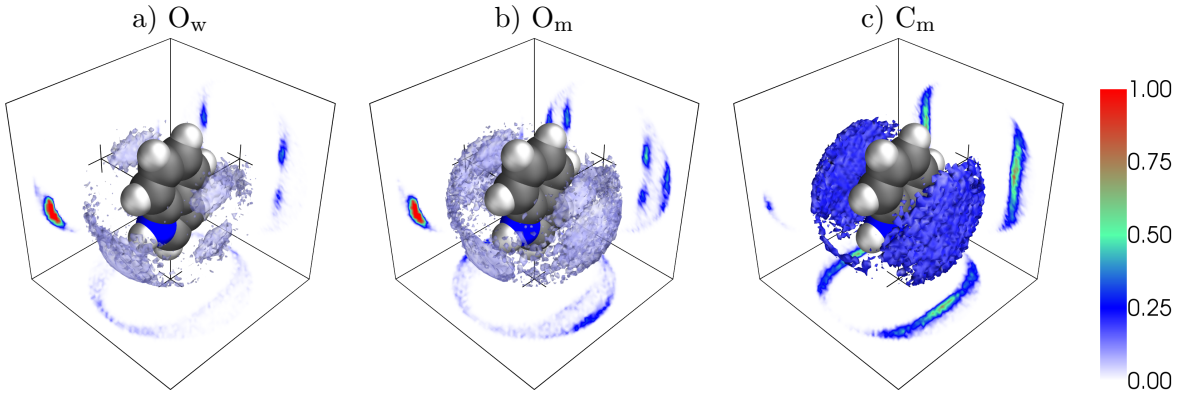


Figure 3.17: SDMs showing the distribution of solvent molecules within 5 Å of indole generated from MD. The origin of the solvent molecules was defined as the water oxygen (a), the methanol oxygen (b) and the methanol carbon (c).

Analogous SDMs for O_m around indole in EPSR1 (Fig. 3.16 (b)) and MD (Fig. 3.17 (b)) similarly show a high propensity for the hydroxyl group of methanol to be located around the benzene ring, in addition to forming hydrogen bonds to the amine group. The difference in the degree of methanol solvation of indole is not as pronounced between EPSR1 and MD as the difference in the hydration of the the molecule (Fig. 3.16 (a) and Fig. 3.17 (a)). However, the maximum density of methanol oxygens in MD (2.73 $Å^{-3}$) is higher than the density in EPSR1 (1.41 $Å^{-3}$).

SDMs showing the distribution of C_m relative to indole, which allow an assessment of the interaction of the hydrophobic group of methanol with indole, show lower density above and below the plane of the benzene ring for both EPSR1 (Fig. 3.16 (c)) and MD (Fig. 3.17 (c)) when compared to their counterparts showing the distribution of O_m relative to indole (Fig. 3.16 (b) and Fig. 3.17 (b)). Differences between the degree of interaction between the benzene ring of indole and the -OH and -CH_3 groups of methanol is made particularly apparent by looking the cut-throughs. The maxmimum

density of C_m atoms also underlines this difference, as for both EPSR1 ($0.61\,Å^{-3}$) and MD ($0.95\,Å^{-3}$) the maximum density of C_m atoms is is lower than the corresponding maximum density of methanol oxygens. These findings are consistent with Fig. 3.12, where the methyl groups have a less well defined orientation relative to the benzene ring. Moreover, the SDMs support the theorised interaction between the solvent molecules and indole illustrated by Fig. 3.12.

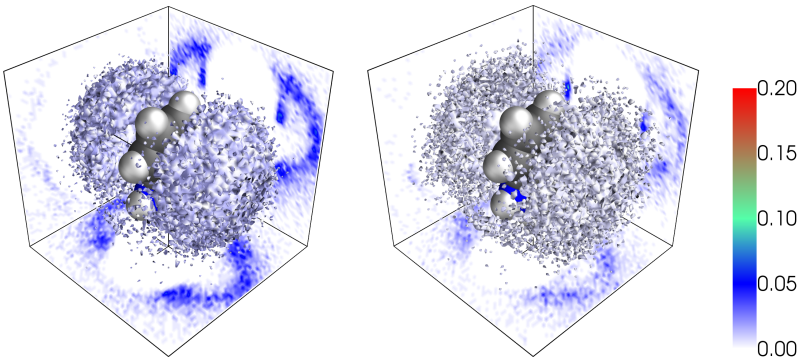


Figure 3.18: SDMs showing the distribution of nearest neighbour water molecules of nearest neighbour water molecules relative to indole (left) and nearest neighbour methanol molecules of nearest neighbour water molecules relative to indole (right) generated from the EPSR1.

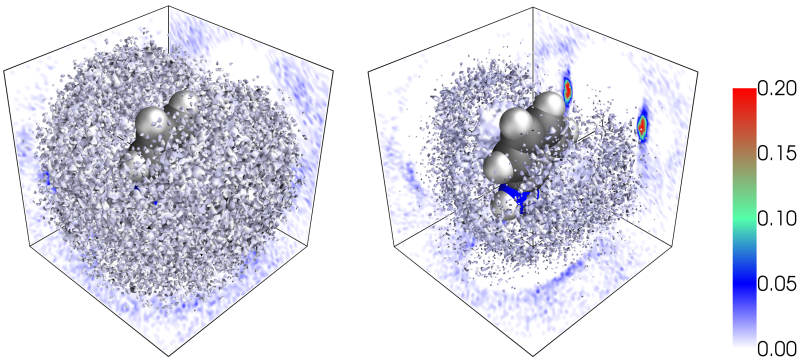


Figure 3.19: SDMs showing the distribution of nearest neighbour methanol molecules of nearest neighbour methanol molecules relative to indole (left) and nearest neighbour water molecules of nearest neighbour methanol molecules relative to indole (right) generated from the EPSR1.

By looking at SDMs that show the distribution of the nearest neighbour solvent molecules of the nearest neighbour solvent molecules relative to indole for EPSR1, as shown in Fig. 3.18 and Fig. 3.19, observations can be made about the solvent structure over a longer range. What is clearest from Fig. 3.18 and Fig. 3.19 is that the long range solvent structure is diffuse in this system, as these SDMs show low density. Not only do these two figures show comparably low densities, but the distribution of the solvent

molecules is also very similar. Given the similarities between the differently charged systems, as embodied by the RDFs and nearest neighbour SDMs, it is unsurprising that in each of EPSR1, EPSR2, and MD this diffuse long range structure is found. That the long range structure is relatively diffuse is in agreement with a number of studies, cited earlier with regards to the solvent-solvent RDFs (Section 3.3.2), which indicate that most methanol and water molecules in methanol/water solutions persist in structures similar to the ones found in the pure liquids, and there no evidence of structural enhancement of either liquid when the two are mixed.[6, 56, 91, 266, 286]

3.3.4 Simulation Checks

3.3.4.1 The Effect of Dummy Atoms

Fig. 3.20 shows the effect on select RDFs of the placing of dummy atoms above and below the centres of each of the constituent rings of indole in order to keep the molecules planar in EPSR1. From this figure, it can be seen there there is no significant difference between simulations with and without dummy atoms. Peak positions are consistent across the two simulations, and the general picture of solvation is maintained. RDFs in Fig. 3.20 showing the interaction between methanol atoms and the benzene ring of indole are identical for EPSR simulations run with and without dummy atoms. Although there are some minor differences in the height of the peaks in the RDFs showing the hydration of the benzene ring, these differences are slight and clearly not due to the presence of dummy atoms, as the indole-methanol RDFs are unchanged by their presence. Therefore, although not influencing the behaviour of the system in terms of the RDFs, the affixing of dummy atoms can be said to have effected a more physically realistic model by ensuring a planar indole molecule.

3.3.4.2 The Effect of Fitting

As discussed in Section 2.2, the key feature of EPSR is that it produces a 3D structural model using experimental data as a constraint. Thus, in order to compare the effect of the fitting process on the $F(Q)$s, a simulation of indole in methanol/water solutions was

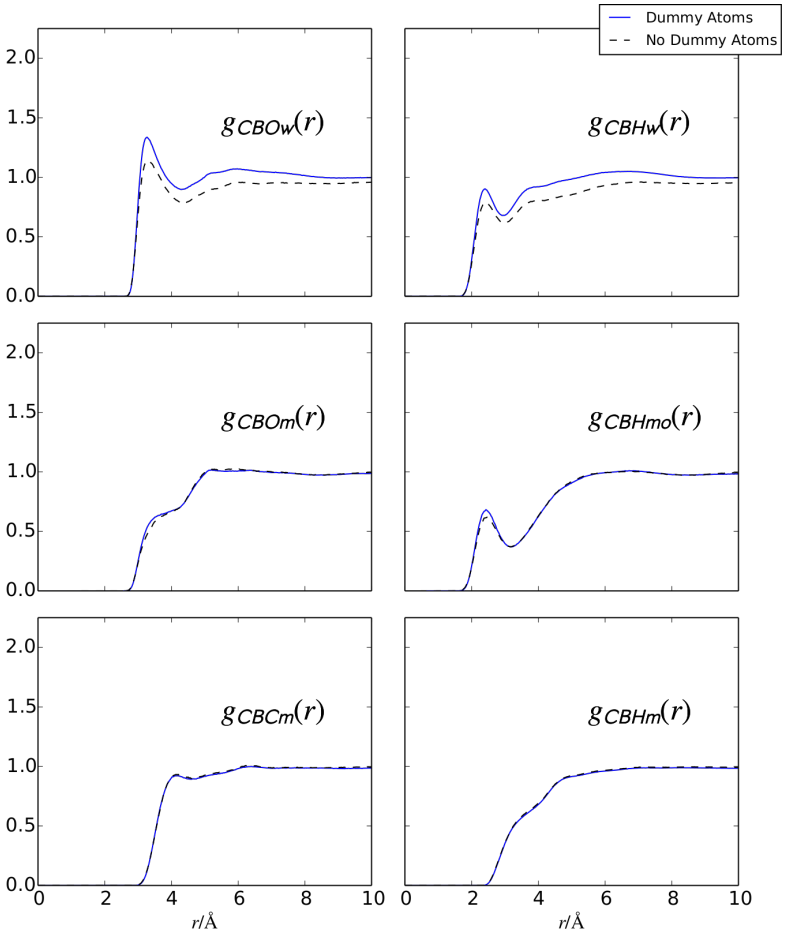


Figure 3.20: The effect of dummy atoms on the RDFs for interactions between the benzene ring of indole and solvent atoms.

run without fitting the simulated data to the experimental diffraction data. All other parameters remained the same between the two simulations.

The $F(Q)$s and $G(r)$s for the EPSR simulation run without fitting are shown in Fig. 3.21. The $F(Q)$s and $G(r)$s for when the fitting process was applied are those which are shown in Fig. 3.5 and Fig. 3.6. That the $F(Q)$s and $G(r)$s look so similar before fitting means that other constraints imposed on the simulation before fitting well characterised the system. These constraints include the experimental bulk density, atomic concentrations and the intramolecular structures of the molecules. For example, for a methanol-benzene ring correlation, there are 6 carbon atoms and 4 hydrogen atoms on the benzene ring of indole and 6 distinct atoms of methanol. This means that a resultant C_B-H_{mo} correlation is actually constrained by 60 different distances and that the model must be consistent with all of these individual correlations in order to fit to the diffraction data. In short, because the initial system before fitting is accurately

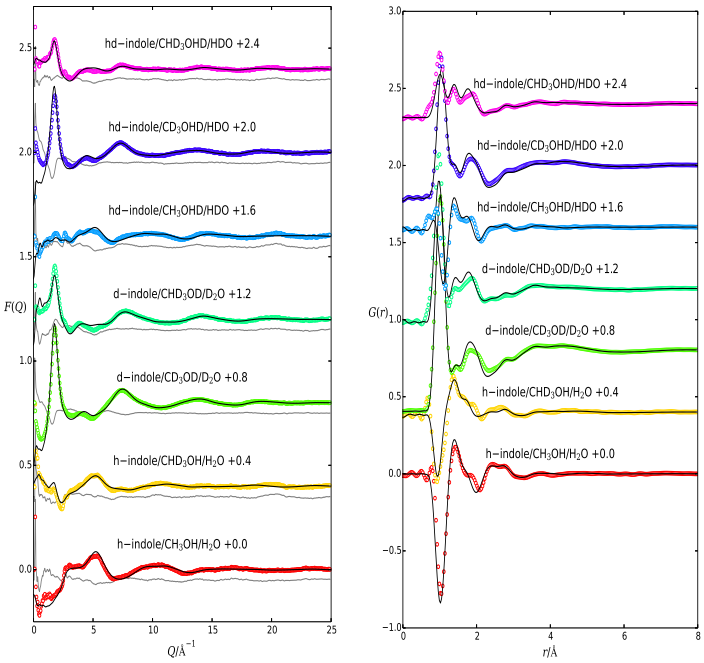


Figure 3.21: Measured neutron diffraction data (coloured lines) and simulated EPSR data that has not been fit (black lines) in reciprocal space (left) and real space (right). The difference curves (grey lines) have been shifted by 0.05 for clarity.

defined, with respect to the experimental data, differences between the $F(Q)$s and $G(r)$s before and after fitting to the experimental data is small.

3.3.4.3 Uncharged Systems

The importance of starting the EPSR simulation with a charge balanced system is highlighted by Fig. 3.22. This figure shows select RDFs for solvent-solvent and indole-solvent interactions for EPSR1, a completely uncharged system and a system with an uncharged indole molecule. The system with an uncharged indole molecule has solvent molecules with the same parameters as in EPSR1.

RDFs for both the completely uncharged system and the sytem with an uncharged indole molecule show much well less defined peaks when there is an absence of the charges that contribute to the association between molecules in EPSR1. This is particularly true of the solvent-solvent RDFs, the solvent atoms being the most prevalent in the simulation.

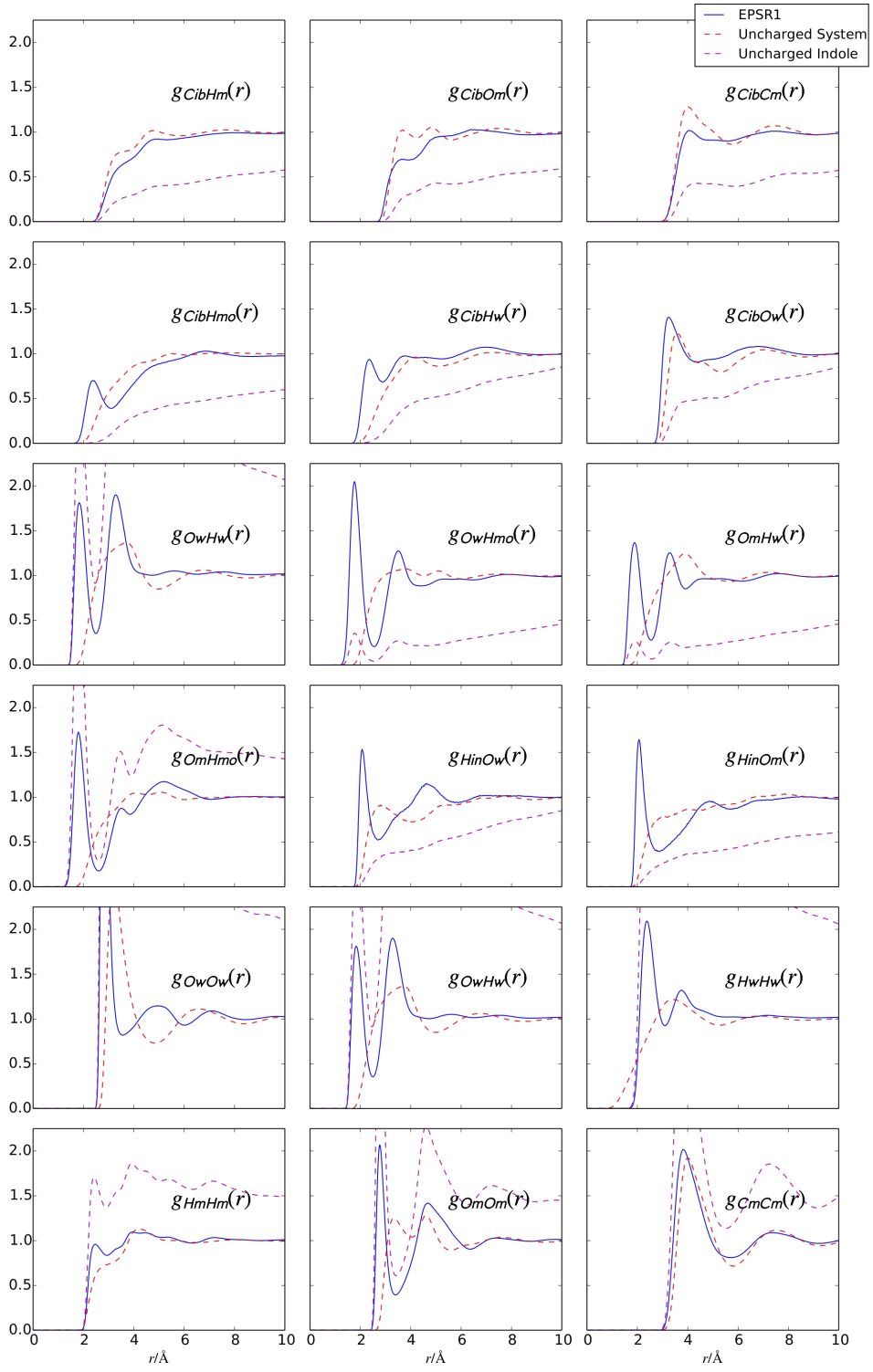


Figure 3.22: A selection of RDFs showing the effect of charges in the indole/methanol/water system. EPSR1 (blue line) is compared to a completely uncharged system (dashed red line) and a system where just the indole molecule is uncharged (dashed magenta line).

3.3.5 Indole in Different Methanol/water Concentrations

Nearest neighbour SDMs generated from the MD simulations of indole dissolved in methanol/water solutions run at different ratios of methanol:water for a ratio of 1 indole molecule to 1000 solvent molecules (infinite dilution) to assess the effect of the methanol:water ratio on indole solvation are shown in Fig. 3.23. The percentage of nearest neighbour solvent atoms which were water and the percentage which were methanol was recorded. This information is summarised in Table 3.9 and presented graphically in Fig. 3.24.

Name	Moles of Methanol/Water	Percent Methanol/Water
M000W100	0/1000	0.00/99.9
M001W099	1/998	0.10/99.8
M010W090	10/989	1.00/98.9
M025W075	250/749	25.00/74.9
M050W050	509/491	50.85/49.1
M075W025	750/249	75.00/24.9
M100W000	1000/0	99.90/0.0

Table 3.9: The percentage of nearest neighbour solvent molecules which are either water or methanol in a solution of 1 indole molecule and 1000 solvent molecules.

As can be seen from Table 3.9 and Fig. 3.24, the decrease in the percentage of nearest neighbour solvent molecules which are water decreases almost linearly with increasing methanol concentration. The same characteristics of the distribution of solvent atoms around indole as seen in previous SDMs, such as the position of solvent atoms above and below the plane of the benzene ring of indole, is seen throughout Fig. 3.23. The density of solvent molecules, either methanol or water, in these regions varies following Table 3.9. It has been previously found that when a solute is dissolved in a mixed solvent the specific solvation effects determine that the proportion of the solvent in the solvation sphere of the solute can differ from the proportion in the bulk solvent.[232] These data indicate that the preferential solvation of indole in the present solutions is not a function of the methanol:water ratio. However, it should be borne in mind when considering these data that these MD simulations were run at an infinite dilution at a concentration of indole $\sim$17 times lower than that at which the neutron diffraction data were collected.

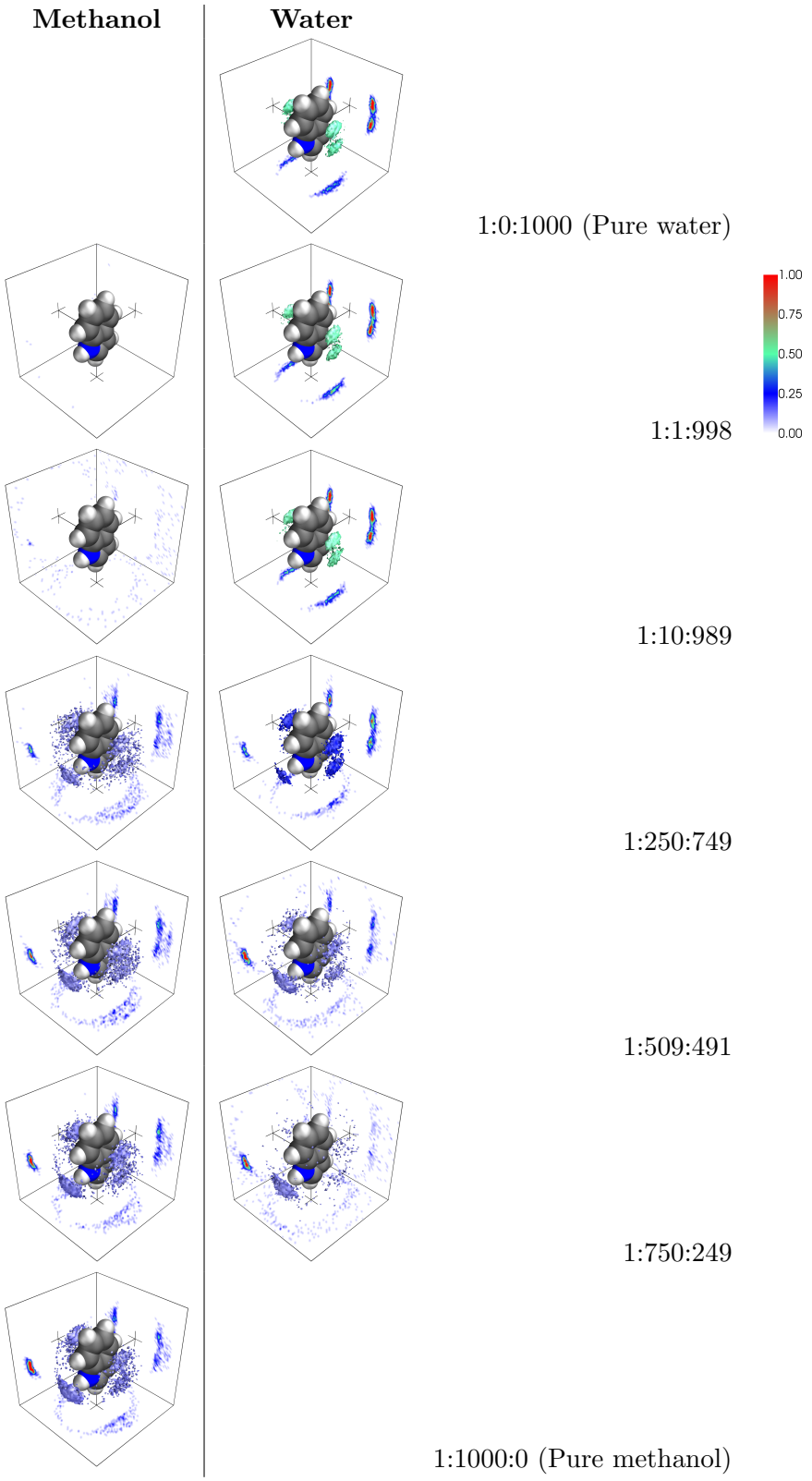


Figure 3.23: SDMs showing the distribution of nearest neighbour solvent molecules relative to indole generated from MD simulations run at different molar ratios.

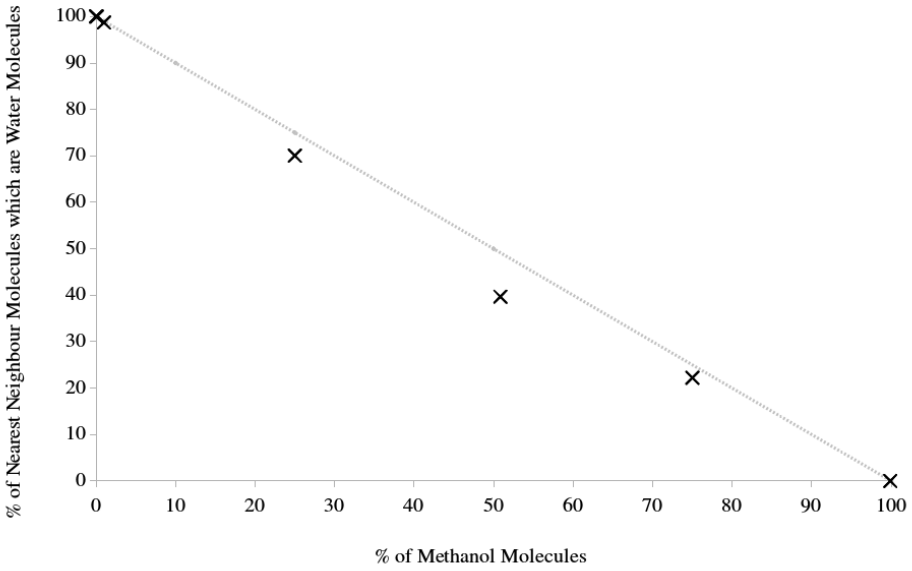


Figure 3.24: The percentage of nearest neighbour solvent molecules which are water as the percentage of methanol molecules is increased for 0% to 100%. The dashed line represents a linear change. SDMs corresponding to each point plotted are shown in Fig. 3.23.

3.3.6 NMR

3.3.6.1 Assignment in Chloroform

Indole peaks were assigned using ^{13}C HSQC and ^{1}H-^{1}H COSY spectra of indole dissolved in $CDCl_3$ at a concentration of 1 indole: 150 $CDCl_3$. Previously measured data for the benzene ring of tryptophan were also used in the assignment. A full assignment is displayed in Figure 3.25. ^{13}C HSQC and ^{1}H-^{1}H COSY spectra are included in Appendix B.1.

Indole peaks were assigned by first identifying the carbons, using the ^{13}C HSQC spectrum of indole dissolved in $CDCl_3$, associated with protons characterised by chemical shifts of 7.42ppm and 7.66ppm, by referring to previous NMR data for the benzene ring of tryptophan. Thus, indole peaks for indole dissolved in $CDCl_3$ at 7.42ppm and 7.66ppm were assigned as H_7 and H_4, respectively. As H_7 is the closest proton to H_6 and H_4 is the closest proton to H_5, using the couplings measured in the corresponding ^{1}H-^{1}H COSY spectrum, H_4 and H_5 were then assigned. H_2 and H_3 were assigned by analogous

reasoning. The most downfield peak at 8.15ppm was identified as the exchangeable indole hydrogen, H_1; its ability to exchange manifesting itself as a broader peak in the NMR spectrum.

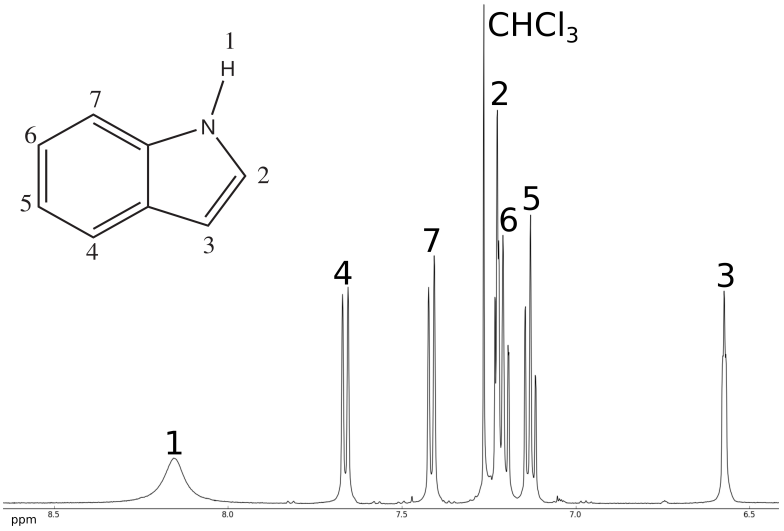


Figure 3.25: A ^{1}H spectrum of indole dissolved in $CDCl_3$ at a concentration of 1 indole: 150 $CDCl_3$. The peaks have been asssigned with labels that correspond to those of the inset indole molecule.

3.3.6.2 Indole in Different Solvents

Spectra of indole dissolved in CD_3OD, CD_3OD/D_2O and D_2O each show peaks attributable to the same protons identified in Fig. 3.25, with the exception of H_1. H_1 can exchange with the deuterium atoms of D_2O and CD_3OD. Due to differences in the solubility of indole in methanol and water, the spectra shown in Fig. 3.26 were not all recorded at the same concentration.

Comparing spectra measured at different concentrations was deemed valid, as measurements of indole dissolved in CD_3OD/D_2O at 3 different concentrations (not shown), 1:29:30 - the concentration at which neutron diffraction measurements were performed - 1:145:150 and 1:290:300, showed the same peak positions and couplings, implying that the conformation of indole is concentration independent at such orders of magnitude of concentration.

Chemical shifts of the indole peaks are mainly related to the different properties of the solvents. This is highlighted by the Table 3.10 showing the chemical shift of the

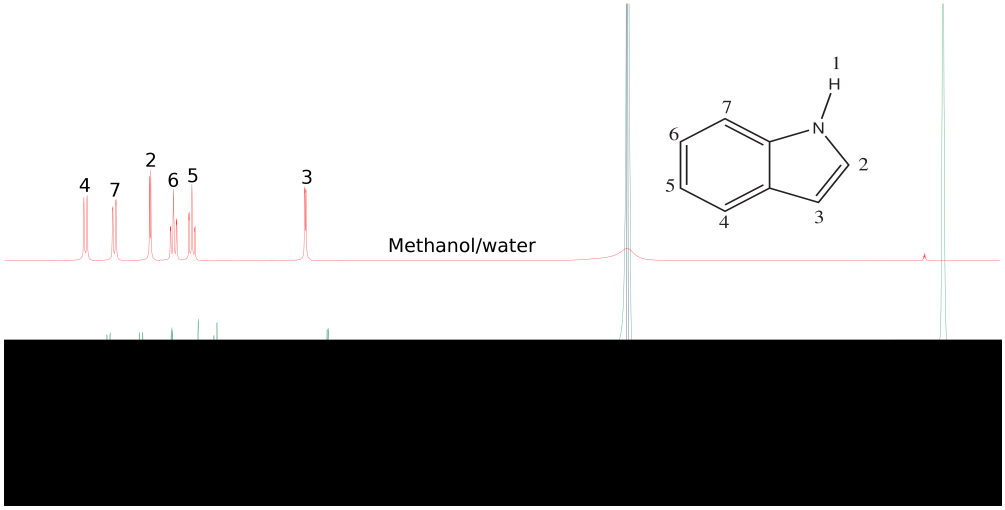


Figure 3.26: ^{1}H spectra of indole dissolved in CD_3OD (1:150), CD_3OD/D_2O (1:150) and D_2O (1:1500). Methanol and water OH peaks are at 4.8ppm. The methanol CH_3 peak is at 3.3ppm.

peaks, apart from H_1, in each of the four solvents relative to H_3, which had the smallest range of peak position across the four solvents. From this table it can be seen that there is little change in the position of the indole peaks relative to H_3. Furthermore, the coupling constants, J, were also consistent across the different solvents. For example, H_7, a doublet, has a mean coupling constant across the four solvents of (8.25±0.08)MHz, the error being calculated from the standard deviation. The mean chemical shift range for all peaks across all four solvents was calculated to be 0.34ppm. Peaks were consistently measured more downfield in D_2O, followed by $CDCl_3$, CD_3OD/D_2O, CD_3OD. The similar chemical shifts of the indole peaks, as underlined by Table 3.10, together with the same couplings and carbon shifts in the COSY and HSQC spectra, respectively, imply that the structure of indole is not significantly influenced by the solvent in which it is dissolved.

	Chemical Shift of Peak Relative to H_3 (ppm)				
Solvent ↓	**H_2**	**H_4**	**H_5**	**H_6**	**H_7**
$CDCl_3$	0.66	1.09	0.56	0.64	0.85
CD_3OH	0.78	1.11	0.56	0.65	0.94
CD_3OH/H_2O	0.78	1.11	0.57	0.66	0.96
H_2O	0.83	1.14	0.58	0.67	0.97

Table 3.10: The chemical shift of each peak relative to H_3 in $CDCl_3$, CD_3OD,CD_3OD/D_2O and D_2O, corresponding to spectra in Fig. 3.25 and Fig. 3.26.

3.4 Conclusions

Differences in the solvation of indole are seen with a change in the polarity of the charges of indole carbon atoms (Fig. 3.2). Specifically, EPSR1 shows greater hydration of the benzene ring of indole than EPSR2, whereas EPSR2 shows greater benzene-methanol interactions than EPSR1. These differences are seen most clearly both through the RDFs and CNs corresponding to benzene-solvent interactions, despite the comparable extent to which EPSR1 and EPSR2 fit the neutron diffraction data. However, there are also differences seen in the solvation of indole when comparing EPSR2 to MD, both of which have the same charge distribution. MD shows less hydration of the benzene ring of indole than EPSR2 and the greatest solvation by methanol of the benzene ring. The differences in solvation across the three simulations are more pronounced for indole-water interactions than indole-methanol interactions but perceptible in both scenarios.

Despite differences between the degree of solvation by water and methanol experienced by indole, which correspond to changes in charge distribution and simulation technique, the nature of these interactions remains consistent between the three simulations presented here. This pattern is particularly evident in the SDMs and plots showing the orientation of solvent relative to indole. Moreover, contrary to what might be predicted by some models of hydrophobicity, hydrogen bonding configurations appear to be consistently preferred over hydrophobic-hydrophobic interactions for indole in this model amphiphilic solution.

Although it might be expected that some benzene-water interactions would be present, given the presence of π-electrons on benzene, and that in crystal structures there is evidence of hydrogen bond-like interactions of water or other side chains to the benzene ring of Trp (indole),[274, 275] that the solution structure of indole is dominated by these electrostatic interactions is surprising. The dominance of these interactions is particularly unexpected here, as Trp is sometimes ranked as the most hydrophobic of amino acids.[195]

The standard theory of the hydrophobic effect predicts indole to be preferentially solvated by hydrophobic motifs, as methanol microsegregates in solution,[56] offering

a near continous surface of hydrophobicity in which indole could, in principle, be partitioned. However, it could be that the hydrophobic surfaces in the present solutions are too small, as it has been suggested that a large enough hydrophobic cavity must be present for the hydrophobic effect to drive assembly.[42] Notwithstanding this observation, the data here show that electrostatic interactions between the benzene ring of indole and either methanol or water comprise most of the nearest neighbour solvation of indole. The preponderance of electrostatic interactions is evidenced by the high density of H_{mo} and H_w atoms at short distances in the RDFs. As discussed, while MD shows relatively fewer hydrogen bonding configurations, these configurations are still dominant. At least 40-50% of the nearest neighbour interactions are -OH groups. It has previously been noted that structures which are deficient in hydrogen-bond acceptors will make use of the π-systems of aromatic rings.[173]

Taken overall, these data imply that electrostatic forces are far more powerful in dictating amphiphilic solvation of a mostly hydrophobic molecule. The electrostatic interactions between the benzene ring of indole and -OH groups of methanol appear to place the -CH_3 groups close enough to benzene ring to participate in stabilising van der Waals interactions between indole and methanol. Water, in contrast, has a broad range of distances, indicated by the broad angular range for the benzene-H_w-O_w motif in Fig. 3.12 (A) for the water hydrogens that do not participate in the benzene-hydrogen bonding interactions. For indole, it seems that it is the amphiphilic nature of the solvent, rather than the hydrophobic regions present in the solvent, which dictates nearest neighbour solvation.

The importance of the amphiphilic solution in facilitating the solubility of indole is underlined if it is considered that the solubility of benzene and pyrrole in water expressed as a molar ratio is 1: 2619 and 1:81, respectively. The solubility of indole in water, rather than having a value between the solubilities of its two constituent structures, is 1:3400. However, in an approximate 50/50 methanol/water solution indole is soluble. The greater than 10 fold increase in the solubility of indole in water on the addition of methanol adds credence to the aforementioned theory of how the interplay between hydrophobic and hydrophilic motifs may precipitate an increase in the solubility of a hydrophobic molecule. Furthermore, the propensity of indole to interact with water

molecules only when in an amphiphilic environment is consistent with experimental data showing a greater percentage of Trp in membrane proteins compared with water soluble proteins.[109, 245]

These results also emphasise the importance of electrostatic interactions in the behaviour of proteins and their interactions with amphiphilic membrane interfaces, suggesting that electrostatics help drive the behaviour of proteins, rather than, for example, indole being excluded from the hydrocarbon core due to entropic reasons,[313]. There are many previous previous studies which show hydrogen bonding interactions with π electrons from aromatic systems,[274, 275] where the X–H$\cdots\pi$ interactions can be considered as weak hydrogen bonding interactions.[173] Although cation-π interactions are rare in proteins compared with other electrostatic interactions, such as from salt bridges, these interactions have been proposed as important factors in determining the 3D structure and recognition properties of proteins and have been observed previously in crystal structures of small molecules and proteins, as well as in gas-phase spectroscopic measurements. [2, 162, 168, 274, 275, 212]

In a biological membrane, a steep polarity gradient exists across the interfacial region from the highly polar aqueous phase to the highly apolar hydrocarbon core [140], where it is not clear how molecules, traditionally classed as polar or non-polar, will interact with a membrane interface. Additionally, it has been found that in proteins and other small molecules the classification of certain molecular motifs as polar or non-polar does not necessarily dictate their specific interactions in protein structures or by analogy how they stabilise proteins at the protein membrane interface.[30, 70, 199] It has rather been noted that molecular association in amphiphilic solutions is more complex on the atomic scale.[30]

In the more complex membrane environment, it is possible that both electrostatic interactions and hydrophobic interactions between the indole and water, in conjunction with the amphiphilic lipid molecules, may allow for Trp to be stabilised at the membrane bilayer interface, although the orientation of Trp relative to the membrane bilayer cannot be deduced from this study. The atomic scale structural data here show that indole is likely interact with charges present at the interface of a bilayer, predominantly through

its benzene ring motif. It is clear from the present work that electrostatic-π interactions play a significant role in the solvation of indole and these interactions are likely to dictate the role Trp plays in anchoring membrane proteins into their native environments.

Chapter 4

Cocaine in Aqueous Solution

4.1 Introduction

As mentioned in Section 3.1, cocaine belongs to a class of drugs called alkaloids. Found in 20% of plant species, alkaloids are naturally occurring, organic, low molecular weight nitrogenous compounds. In plants they are mainly involved in defence against herbivores and pathogens.[233] Most alkaloids with biological activity in humans affect the nervous system, particularly the action of chemical transmitters (e.g. acetylcholine, dopamine and serotonin), leading to their popularity as recreational drugs. However, alkaloids also have a number of medicinal properties (e.g. antihypertensive effects, antiarrhthymic effects, antimalarial activity and anticancer actions).[233]

Cocaine easily crosses the BBB (Section 1.3) and is believed to function once it crosses this barrier by inhibiting dopamine reuptake through binding to the dopamine transporter. As a result, dopamine cannot leave the synapse and produces an amplified signal to the receiving neurons.[97, 230, 314] Cocaine is highly lipophilic, so much so that treatment of cocaine overdose includes the administration of lipid emulsions to dampen its effects.[121] Cocaine is also hydrophilic, the hydrochloride salt being readily soluble in aqueous solution. Administered as the hydrochloride salt and in its freebase form, both forms of cocaine are very addictive and have similar pharmacological effects.[100, 210] It has been stated that cocaine is active only in its protonated form and that cocaine crosses

the BBB only as a freebase (*viz* its deprotonated form).[316] However, crystallographic investigations of cocaine molecules bound into a protein receptor show that the cocaine molecule is deprotonated,[97] and measurement of molecules as they cross the BBB *in vivo* is not yet possible.

To date there is limited information concerning the atomic structure of BBB penetrating molecules such as cocaine in solution, the physical milieu in which they operate *in vivo*. In this Chapter, the structure and conformation of cocaine hydrochloride is investigated in aqueous solution on the atomic scale. Both how cocaine interacts with water and its average conformation are assessed in detail, as estimates of the ability of molecules to cross the BBB have been linked to molecular shape and hydrogen bonding descriptors, such as the number of hydrogen bond donors and acceptors.[301] The work presented in this Chapter also builds on the results of indole in methanol/water solutions presented in Chapter 3, by further comparing the solvation of hydrophilic and hydrophobic chemical groups.

4.2 Methods

4.2.1 Neutron Diffraction Sample Preparation

Cocaine hydrochloride was purchased from Sigma Aldrich (CAS 53-21-4) and used without further purification. The molecular structure of the protonated form of cocaine is shown in Fig. 4.1, which is distinguished from the freebase form of cocaine by an exchangeable proton (H_P) attached to its nitrogen atom. Following the procedure described in Section 3.2.1, 3 g of cocaine·HCl was deuterated. Deuteration of H_P was verified by measuring cocaine·HCl dissolved in $CDCl_3$ using ^1H NMR. The NMR sample was prepared and measured as outlined in Section 3.2.6. Integration indicated that H_P was ∼93% deuterated. Water used in neutron diffraction sample preparation was as is detailed in Section 3.2.1. All isotopomeric solutions of cocaine·HCl dissolved in water were prepared by mass in an N_2 atmosphere, as in Section 3.2.1.

Six different isotopomers of cocaine·HCl dissolved in water were prepared at a rel-

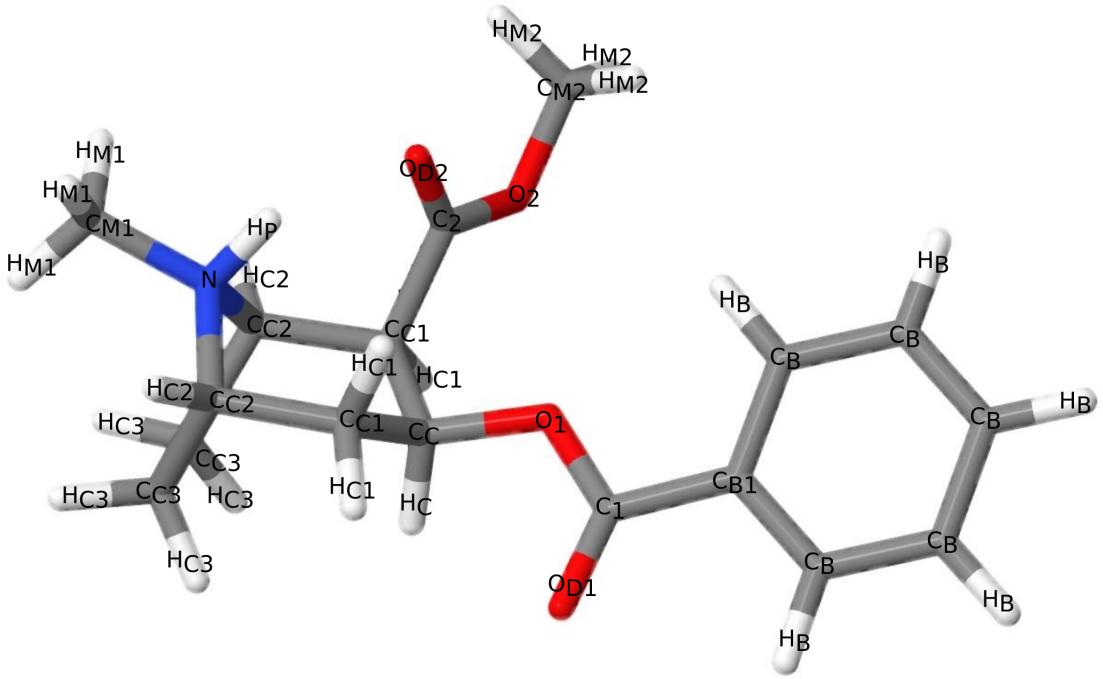


Figure 4.1: The molecular structure and labelling scheme of the protonated form of cocaine.

Sample Name	Cocaine Hydrochloride	Water
100h	h-cocaine	H_2O
75h/25d	$h_{0.75}d_{0.25}$-cocaine	$H_{1.5}D_{0.5}O$
50h/50d	$h_{0.5}d_{0.5}$-cocaine	HDO
64h/36d	$h_{0.64}d_{0.36}$-cocaine	$H_{1.28}D_{0.72}O$
100d	d-cocaine	D_2O
25h/75d	$h_{0.25}d_{0.75}$-cocaine	$H_{0.5}D_{1.5}O$

Table 4.1: The components of the six samples studied. The molar ratio used was 1 cocaine hydrochloride: 65 water. Subscripts indicate isotopic proportions.

ative molar ratio of 1 cocaine hydrochloride: 65 waters (∼0.8 M). The samples measured by neutron diffraction are listed in Table 4.1, where h-cocaine and d-cocaine refer cocaine·HCl with an H atom and a D atom attached to N, respectively.

4.2.2 Neutron Diffraction Measurements

Diffraction data were collected at 298 K using the SANDALS instrument (Section 2.1.2). The samples were transferred from the vials to the SiO_2 cells described in Section 2.1.3.1.2 using gastight syringes. Only the flat face of the main body of each cell was placed in the neutron beam, an area of 25 mm × 25 mm being illuminated. The SiO_2

Sample Name	Measurement Time (μA hrs)	DCS Level
100h	1406	98%
75h/25d	1337	91%
50h/50d	1506	95%
64h/36d	1409	96%
100d	1378	94%
25h/75d	1601	95%

Table 4.2: The measurement time and DCS level for each sample. The average current was ∼170 μA.

cells were cleaned in the same fashion as the Ti/Zr canisters. Diffraction data were also collected from the empty cells, empty instrument and for a vanadium standard for reasons outlined in Section 2.1.3.1.

The μA hrs each sample was measured for is shown in Table 4.2. The empty cells were each measured for 1285-1313 μA hrs. SiO_2 cells were measured for longer than the Ti/Zr canisters used in Section 3.2.2 because of the greater scattering signal from SiO_2 compared with Ti/Zr. The empty instrument and vanadium standard were measured for 394 μA hrs and 1327 μA hrs, respectively.

Data were corrected for absorption, multiple scattering and inelasticity effects, as described in Section 2.1.3.2. All samples had a corresponding DCS level within ∼10% of their theoretical values (Table 4.2), with the mean DCS level of the samples ∼95% of the theoretical value. Each of the SiO_2 cells had a DCS level within ∼1% of the theoretical value. The vanadium standard had a DCS level of between 99% and 100% of the theoretical value.

4.2.3 X-ray Diffraction Measurements

Sample 100h was also measured by X-ray diffraction. The sample was loaded into a borosilicate glass capillary with a diameter of 1 mm using a gastight syringe. Borosilicate is used because of its relatively small X-ray scattering cross section. The capillary was flame sealed and measured at 298 K on a Panalytical X'pert Pro X-ray diffractometer located at the ISIS facility (STFC, UK), which has a maximum power of 35 mA. This instrument uses a silver anode to produce X-rays and has a Q range of 0.7-20 $Å^{-1}$. The

sample was measured for 24 hrs. X-ray diffraction data were also collected for the empty capillary and the empty instrument and were corrected using GudrunX.[264]

4.2.4 Empirical Potential Structural Refinement

The 39.0 Å cubic EPSR simulation box contained 25 protonated cocaine molecules, 25 Cl^- ions and 1625 water molecules at the measured solution density ($\rho = 0.101$ atoms Å^{-3}) at 298 K. The SPC/E water model was used.[18]. The water labelling scheme is the same as in Fig. 3.1.

The molecular structure of cocaine was generated and energetically optimised using Avogadro,[98] and cocaine and Cl^- ion parameters were generated with ANTECHAMBER. [299, 300] Subsequently, collections of atoms were grouped together depending on the similarity of their charges. For example, benzene ring hydrogens, whose charges ranged from 0.145e-0.161 e, were grouped together and collectively labelled as H_B. Sometimes atoms with similar parameters were not grouped together so that their individual behaviour could be understood, such as the carbonyl oxygens. A full list of the parameters used in EPSR is provided in Table 4.3, with the corresponding labelling scheme in Fig. 4.1. This simulation was accumulated for 24623 steps.

Atom Label	ϵ **(KJ mol^{-1})**	σ**(Å)**	**Atomic Mass (u)†**	$\mathbf{q}_{\mathrm{EPSR}}$**(e)**	$\mathbf{q}_{\mathrm{MD}}$**(e)**
C_{B1}	0.4577	3.3997	12.011	-0.1536	-0.153600
C_B	0.4577	3.3997	12.011	-0.0972	-0.064500
C_B	0.4577	3.3997	12.011	-0.0972	-0.141500
C_B	0.4577	3.3997	12.011	-0.0972	-0.064500
C_B	0.4577	3.3997	12.011	-0.0972	-0.074000
C_B	0.4577	3.3997	12.011	-0.0972	-0.141500
H_B	0.0628	2.5996	2.000	0.1516	0.153000
H_B	0.0628	2.5996	2.000	0.1516	0.150500
H_B	0.0628	2.5996	2.000	0.1516	0.153000
H_B	0.0628	2.5996	2.000	0.1516	0.151000
H_B	0.0628	2.5996	2.000	0.1516	0.150500

C_1	0.3598	3.3997	12.011	0.6697	0.669701
O_{D1}	0.8786	2.9559	16.000	-0.5370	-0.537001
O_1	0.7113	3.0000	16.000	-0.4489	-0.448900
N	0.7113	3.2500	14.007	-0.7115	-0.711401
H_P	0.0657	1.0691	2.000	0.4898	0.489800
C_{M1}	0.4577	3.3997	12.011	0.0911	0.091100
H_{M1}	0.0657	1.9600	2.000	0.1010	0.101033
H_{M1}	0.0657	1.9600	2.000	0.1010	0.101033
H_{M1}	0.0657	1.9600	2.000	0.1010	0.101033
C_2	0.3598	3.3997	12.011	0.6471	0.647101
O_{D2}	0.8786	2.9599	16.000	-0.5770	-0.577001
O_2	0.7113	3.0000	16.000	-0.3939	-0.393900
C_{M2}	0.4577	3.3997	12.011	0.1197	0.119700
H_{M2}	0.0657	2.4741	2.000	0.0697	0.069700
H_{M2}	0.0657	2.4741	2.000	0.0697	0.069700
H_{M2}	0.0657	2.4741	2.000	0.0697	0.069700
C_C	0.4577	3.3997	12.011	0.1611	0.161100
C_{C1}	0.4577	3.3997	12.011	-0.1431	-0.141400
C_{C1}	0.4577	3.3997	12.011	-0.1431	-0.144700
C_{C2}	0.4577	3.3997	12.011	0.1460	0.144500
C_{C2}	0.4577	3.3997	12.011	0.1460	0.147500
C_{C3}	0.4577	3.3997	12.011	-0.1039	-0.105400
C_{C3}	0.4577	3.3997	12.011	-0.1039	-0.102400
H_C	0.0657	2.4714	2.000	0.0947	0.094700
H_{C1}	0.0657	2.6495	2.000	0.1174	0.108200
H_{C1}	0.0657	2.6495	2.000	0.1174	0.108200
H_{C1}	0.0657	2.6495	2.000	0.1174	0.135700
H_{C2}	0.0657	1.9600	2.000	0.1192	0.117700
H_{C2}	0.0657	1.9600	2.000	0.1192	0.120700
H_{C3}	0.0657	2.6495	2.000	0.0940	0.095200

H_{C3}	0.0657	2.6495	2.000	0.0940	0.095200
H_{C3}	0.0657	2.6495	2.000	0.0940	0.092700
H_{C3}	0.0657	2.6495	2.000	0.0940	0.092700
Cl	0.4184	4.4010	35.000	-1.0000	-1.0000
O_w	0.6500	3.1660	16.000	-0.8476	-0.8476
H_w	0.0000	0.0000	2.000	0.4238	0.4238

Table 4.3: The parameters used in the EPSR and MD simulations of cocaine hydrochloride/water. † The atomic mass of hydrogen used in MD was 1.000 u. The corresponding labelling scheme for cocaine is shown in Fig. 4.1.

4.2.5 Molecular Dynamics

MD simulations were run[a] using Gromacs 4.5.5[220] with the OPLS-AA force field.[133] To provide an MD simulation complementary to EPSR, the SPC/E water model was used.[18] A simulation was also performed using the TIP3P water model, whose parameters are listed in Table 4.4, for comparison.[129, 174]. The most notable difference between the two water models is that the TIP3P water model has too little structure beyond the first peak in $g_{O_wO_w}(r)$.[174] As mentioned, the topology and parameters of the cocaine molecule and Cl^- ion were generated in AMBER format with ANTECHAMBER[299, 300] and converted to Gromacs format using the amb2gmx.pl script.[186]

The initial system was generated by inserting 64 cocaine molecules, 64 chloride ions and 4160 water molecules into a simulation box. The system was energy minimised using the steepest descents gradient integrator. A 100 ps simulation to allow water to relax with other components positionally restrained was performed using the NVT ensemble. Production simulations of 100 ps used the NPT ensemble. A timestep of

[a]These simulations, including the MD simulation run at infinite dilution, were performed by Akshay Sridhar. They were analysed and coverted to $F(Q)$s/$G(r)$s by myself.

Atom Label	ϵ **(kJ mol^{-1})**	σ **(Å)**	**q(e)**
O_w	0.63639	3.15070	-0.83400
H_w	0.00000	0.00000	0.41700

Table 4.4: The parameters of TIP3P water.

2 fs was used, and bond lengths were constrained using the LINCS algorithm.[104] Van der Waals interactions were cut off at 10 Å. Periodic boundary conditions were used with Particle Mesh Ewald electrostatics cut off at 10 Å.[50] Simulations were performed at 298 K coupled to a Berendsen bath. For production simulations isotropic pressure coupling to a Parrinello-Rahman barostat at 1 bar was used.[19]

A snapshot of the MD simulation run using SPC/E water was converted to an EPSR simulation box so that it could be fit to the diffraction data (EPSR2). The potentials defined in Table 4.3 for EPSR were used in this procedure. An MD simulation was also run for cocaine·HCl at infinite dilution (1 cocaine: 1000 water) with the TIP3P water model.

4.2.6 ANGULA

The orthogonal coordinate systems assigned to cocaine in ANGULA are shown in Fig. 4.2. Unlike in the case of indole, where a single coordinate system assigned to the solute was used to analyse the relative position and orientation of the solvent molecules, 4 coordinate systems were assigned to different fragments of the cocaine molecules. Using multiple coordinate systems is necessary for cocaine because of its larger size and greater flexibility. The coordinate system assigned to water is as described in Section 3.2.5.

WMA (Section 2.3.6) was performed for cocaine·HCl dissolved in aqueous solution for a distance range of 0-4 Å. The most probable nearest neighbour (Section 2.3.5) water molecule of cocaine was extracted for H_P and OD_2. IMAPs (Section 2.3.6) were used to assess the nature of the interactions between cocaine molecules in EPSR and MD. Analysis using ANGULA was performed for ∼4000 snapshots of the EPSR and MD simulation boxes.

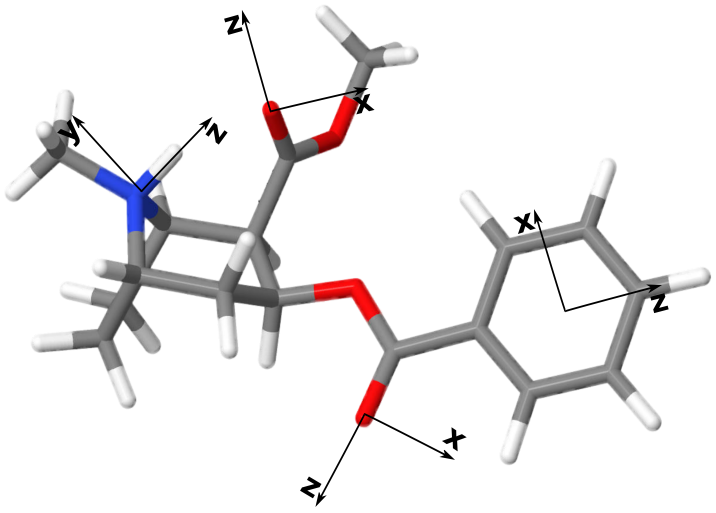


Figure 4.2: The coordinate systems assigned to cocaine using ANGULA.

4.2.7 NMR Sample Preparation and Measurement

NMR samples were prepared and measured as in Section 3.2.6.

In order to confirm that performing neutron diffraction measurements at a temperature of 298 K did not mean that a different structure was measured to that which exists at body temperature, ^{1}H NMR measurements were performed for cocaine·HCl dissolved in D_2O at a concentration of 55 mM for temperatures of 298 K and 310 K.[b]

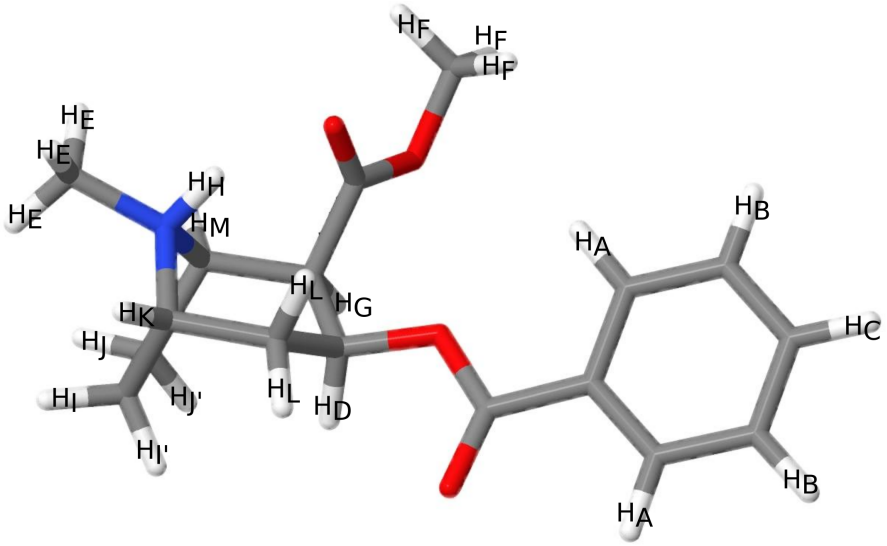


Figure 4.3: The labelling scheme used for cocaine in NMR spectra.

An NMR spectrum of cocaine·HCl dissolved in D_2O was assigned using a combina-

[b]I would like to thank project student Chris Sowden for performing these measurements.

tion of ^{1}H, ^{1}H-^{1}H COSY and ^{13}C-HSQC spectra. The labelling scheme used for this assignment and in subsequent spectra is shown in Fig. 4.3. Cocaine·HCl was measured through a range of concentrations in different solvents to assess if there were any concentration related effects on the spectra and to compare the influence of aqueous and organic environments. Similarly, a pH titration was performed by adding 0.01 M DCl to a sample of known concentration of cocaine·HCl with a pippette. The effect of ionic strength was also investigated by adding NaCl to a NMR sample of cocaine·HCl dissolved in water. The samples listed in Table 4.1 were measured by NMR following their measurement by neutron diffraction to ensure they had not been contaminated. 3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt (TMS), which was purchased from Sigma Aldrich (CAS 2039-96-5), was used as a standard.

4.3 Results

4.3.1 Diffraction Data

Measured neutron and X-ray diffraction data along with the EPSR fits to these data are shown in reciprocal space in Fig. 4.4 (a). In this figure, the EPSR fits to the neutron data are generally good, with some small differences at $Q < 5\,\text{Å}^{-1}$ due to difficulty correcting for inelastic scattering in this region. The X-ray data fits are also good, with again some differences at lower Q values. This difference is likely because polarizability effects are not taken into account, as the weighted sum of atomic form factors is currently used to obtain the structure factor.[261] The Fourier transformations of these data and EPSR fits are shown in Fig. 4.4 (b), where a similar level of agreement between the experimental and simulated data is observed as in reciprocal space (Fig. 4.4 (a)).

Fig. 4.5 shows the $F(Q)$s generated from the MD trajectories run using the different water models (SPC/E and TIP3P) and the EPSR2 fits in reciprocal space alongside the same neutron and X-ray diffraction data shown in Fig. 4.4. The corresponding $G(r)$s generated from the MD trajectories and the EPSR2 fits in real space are shown in Fig. 4.6.

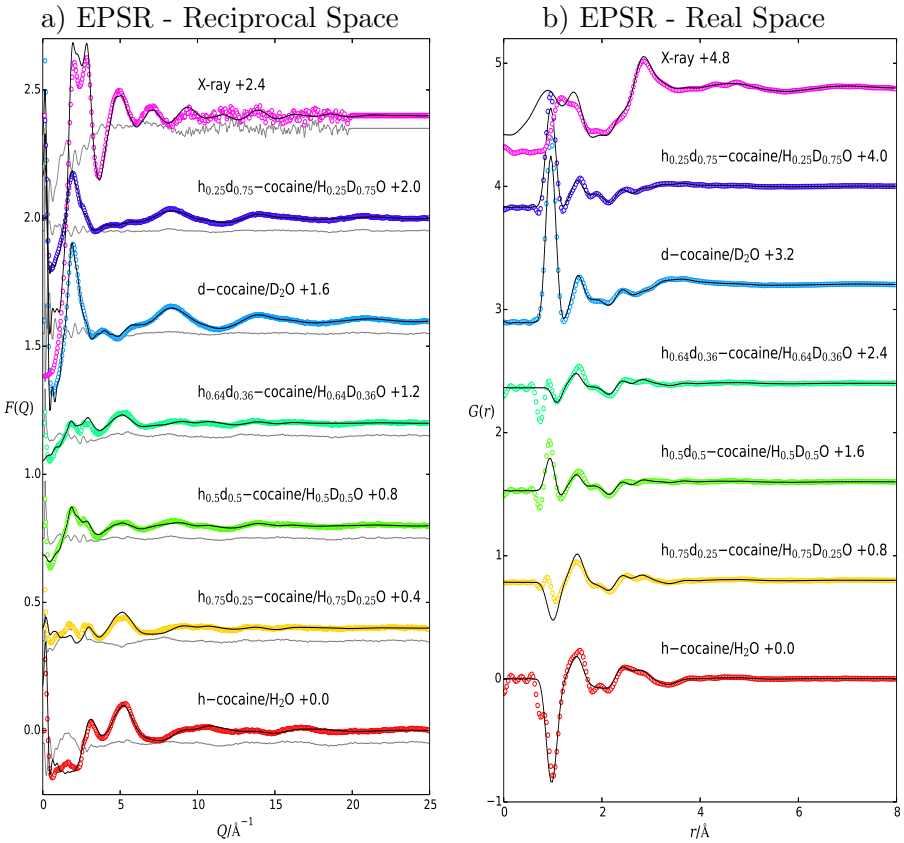


Figure 4.4: Measured neutron and X-ray diffraction data (coloured lines) and EPSR fits to the data (black lines) in reciprocal space (a) and real space (b). The difference curves (grey lines) have been shifted by 0.05 for clarity.

Comparing Fig. 4.5 (a) and Fig. 4.5 (b) reveals no obvious difference between the extent of the agreement of the experimental and simulated diffraction data when a different water model is used, the difference curves being practically identical for MD simulations run using SPC/E and TIP3P water. Accordingly, $G(r)$s in Fig. 4.6 (a) and Fig. 4.6 (b) match the experimental data to similar extent as in Fig. 4.5 (a) and Fig. 4.5 (b). Although MD trajectories using both the SPC/E and TIP3P water models result in $F(Q)$s which follow the same general form as the experimental data, at higher Q the difference curves are not as flat in Fig. 4.5 (a) or Fig. 4.5 (b) as in Fig. 4.4 (a). This is most apparent for the fully protonated sample (h-cocaine/H_2O) and the most deuterated sample (d-cocaine/D_2O), because the opposite signs of the scattering lengths of the hydrogen and deuterium solvent atoms, which comprise the majority ($\sim$51%) of the scattering signal, mean that their correlations are less visible in other $F(Q)$s.

EPSR2, which is effectively a snapshot of the MD trajectory corresponding to Fig. 4.5

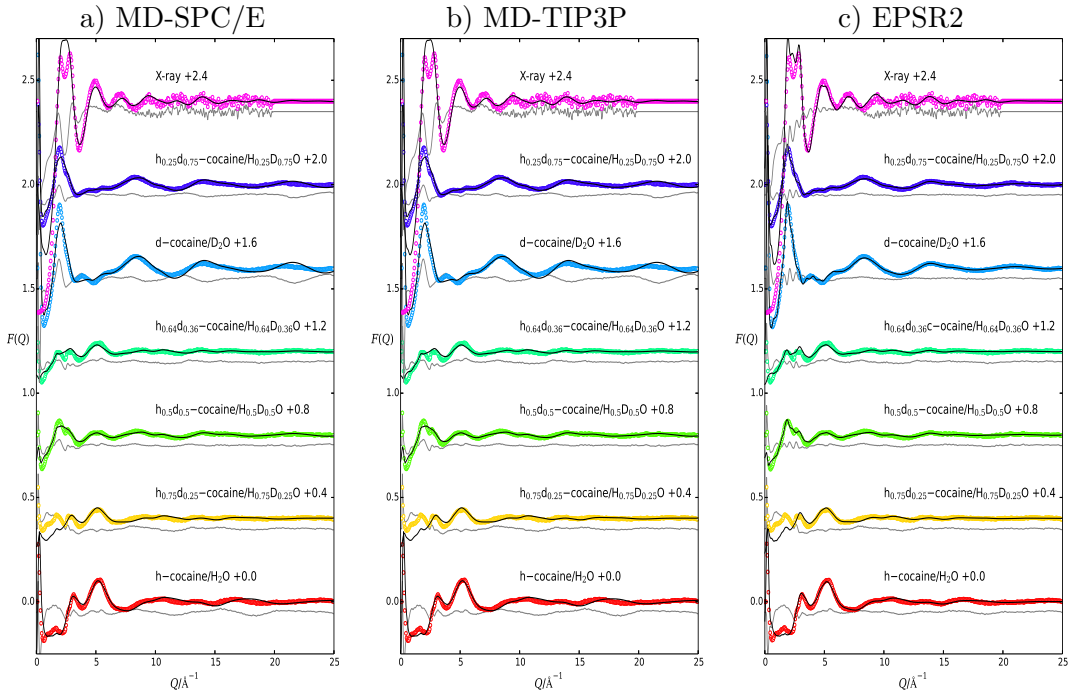


Figure 4.5: Measured neutron and X-ray diffraction data (coloured lines) and the $F(Q)$s generated from the MD simulations (black lines) in reciprocal space for simulations using the SPC/E water model (a), the TIP3P water model (b) and the SPC/E water model with fitting to the experimental data by EPSR (c). The difference curves (grey lines) have been shifted by 0.05 for clarity.

(a) that has been fit to the measured diffraction data with EPSR, shows an improvement in the agreement of the experimental and simulated data compared with MD simulations run using either water model. This improvement is particularly noticeable for the most deuterated sample, where the difference curve is flat in Fig. 4.5 (c). The agreement between the experimental and simulated data is also greatly improved for the first peak of the X-ray diffraction data in Fig. 4.5 (c), for example, and also for the trough in the fully protonated sample in real space (Fig. 4.6 (c)). Comparison of EPSR2 with EPSR shows that the two sets of $F(Q)$s fit the data to a similar extent, despite the different starting structures of the two simulations.

4.3.2 Cocaine-cocaine Interactions

Initial inspection of the MD simulation boxes for which the $F(Q)$s and $G(r)$s are plotted in Fig. 4.5 and Fig. 4.6 revealed a number of cocaine molecules within close proximity of each other. If they are not representative of the measured diffraction data,

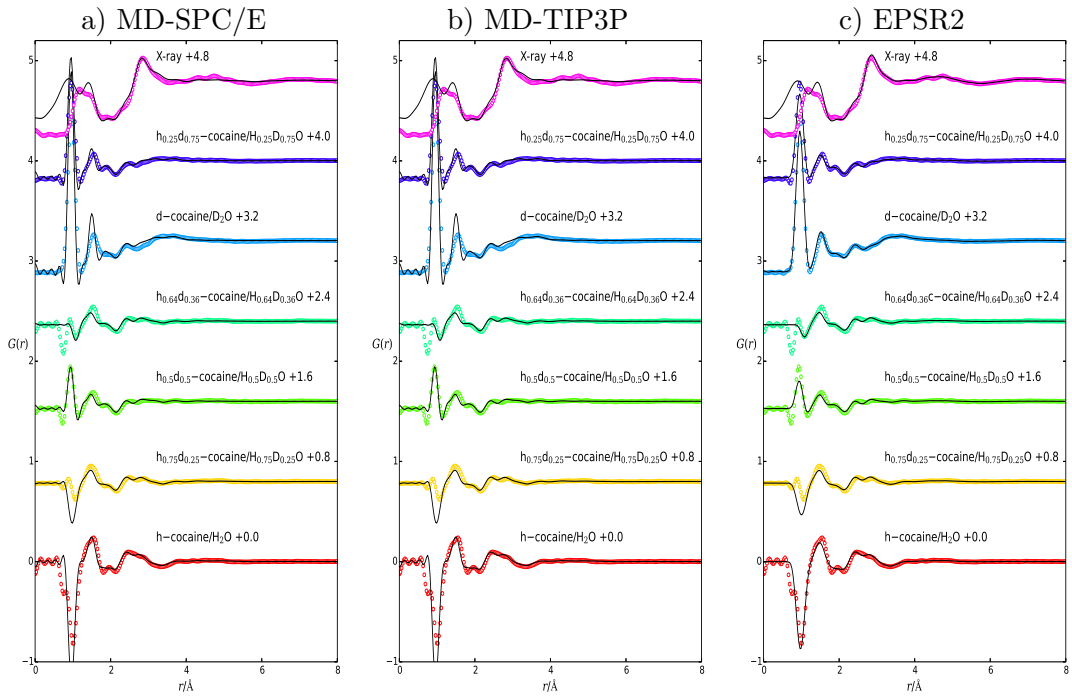


Figure 4.6: Measured neutron and X-ray diffraction data (coloured lines) and the $G(r)$s generated from the MD simulations (black lines) in real space for a simulations using the SPC/E water model (a), the TIP3P water model (b) and the SPC/E water with fitting to the experimental data by EPSR (c).

cocaine-cocaine interactions may partially explain why difference curves in Fig. 4.5 (a) and Fig. 4.5 (b) are not flat as difference curves in Fig. 4.4 (a). Furthermore, fitting of a snapshot of the MD trajectory in Fig. 4.5 (a) to the diffraction data (EPSR2) could be influenced by the frequency of these interactions. A selection of cocaine-cocaine RDFs is shown in Fig. 4.7, where RDFs have been extracted from the MD simulations run using both the SPC/E and TIP3P water models, EPSR2 and EPSR. Also shown in Fig. 4.7 are RDFs extracted from EPSR2 following a temperature increase to 398 K and subsequent decrease to the temperature at which samples were measured by diffraction (298 K).

As might be anticipated given Fig. 4.5 and Fig. 4.6, the water model used in MD seems to have no effect on cocaine-cocaine interactions. Both the location of the peaks in the RDFs shown in Fig. 4.7 and their height is preserved between MD simulations run using the SPC/E water model and the TIP3P water model. There are, however, some minor differences between RDFs extracted from the MD simulations and the RDFs extracted from EPSR2. For example, peaks at ∼2.9 Å and ∼2.8 Å in $g_{\mathrm{NO_{D2}}}(r)$ and $g_{\mathrm{O_{D2}O_{D2}}}(r)$, respectively, are larger for EPSR2 than for either MD simulation. Other

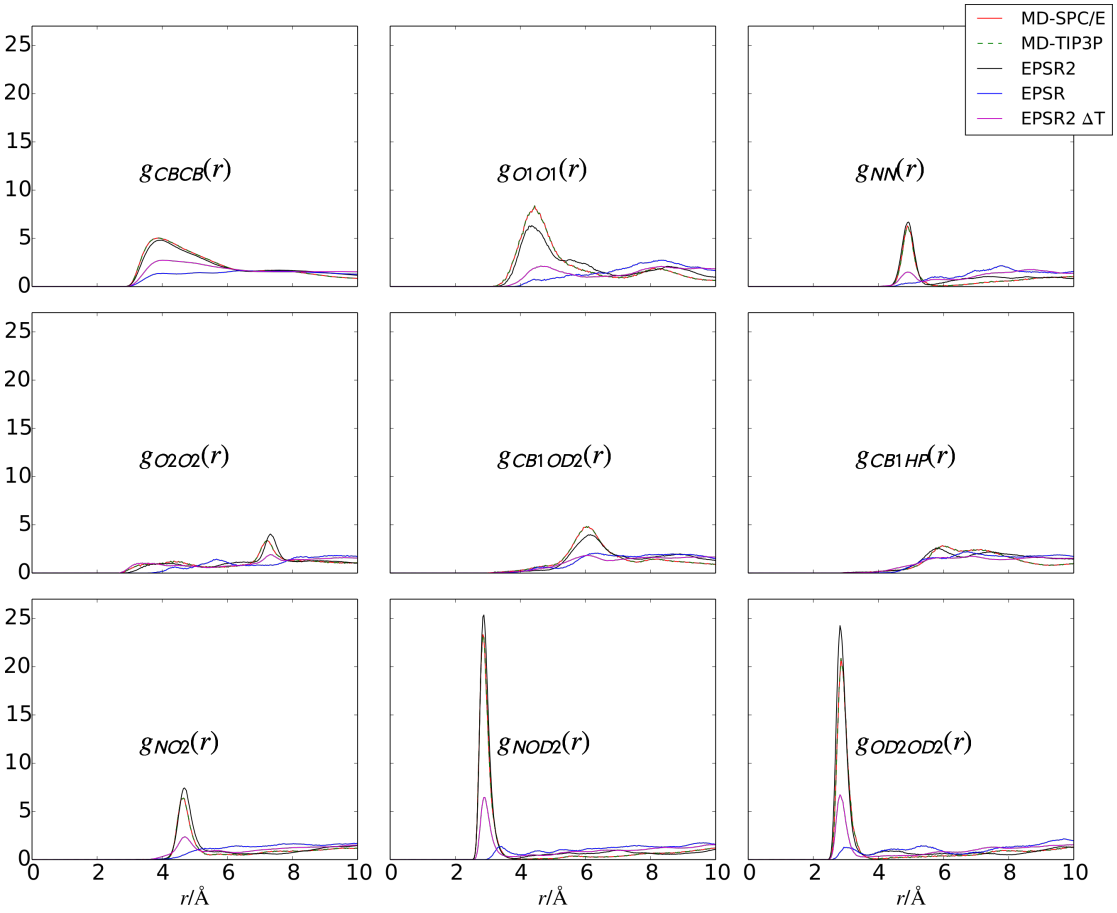


Figure 4.7: RDFs for cocaine-cocaine interactions extracted from MD simulations run using SPC/E (red line) and TIP3P (dashed green line) water models, EPSR2 (black line) and EPSR (blue line). Additionally, cocaine-cocaine RDFs extracted from EPSR2 with changes in temperature (magenta line) are included.

peaks, such as the peak in $g_{\mathrm{NN}}(r)$ at ∼4.9 Å, are also marginally larger in EPSR2. Peaks at ∼4.4 Å and ∼6.2 Å in $g_{\mathrm{O_1O_1}}(r)$ and $g_{\mathrm{C_{B1}O_{D2}}}(r)$, respectively, decrease in height in EPSR2 compared with the RDFs extracted from MD simulations run using the SPC/E and TIP3P water models. Molecules in both MD simulations and in EPSR2 also seem to exhibit a similar level of benzene ring stacking, as indicated by the peaks in $g_{\mathrm{C_BC_B}}(r)$ at ∼4.0 Å.

Minor differences in the height of the RDFs extracted from MD and EPSR2 aside, it is immediately clear from $g_{\mathrm{NO_{D2}}}(r)$ and $g_{\mathrm{O_{D2}O_{D2}}}(r)$ that there is a highly favoured intermolecular interaction between the amine group and the carbomethoxy group of cocaine. Examining the corresponding CNs reveals that ∼40% of cocaine molecules associate *via* N-$\mathrm{O_{D2}}$ interactions in the MD simulations run using SPC/E and TIP3P water models and in EPSR2. Previous neutron diffraction studies of proline in aqueous

solutions at high concentrations (1 proline: 5 water) suggested dimerisation *via* CO_2^- and NH_2^+ groups on neighbouring proline molecules,[33] though these solutions were around 10 times more concentrated than the ones presented here.

Comparing RDFs extracted from MD and EPSR2 with RDFs extracted from EPSR highlights differences in the number of cocaine-cocaine interactions. These differences are starkest for $g_{\mathrm{NO_{D2}}}(r)$ and $g_{\mathrm{O_{D2}O_{D2}}}(r)$, where 800 times fewer cocaine molecules, on average, associate *via* N-O_{D2} interactions in EPSR compared with MD and EPSR2. Although there is a peak visible for EPSR in $g_{\mathrm{NO_{D2}}}(r)$, besides having a much smaller associated CN, the postion of this peak is at ∼3.0 Å as opposed to ∼2.8 Å for MD, indicating that the relatively small number of N-O_{D2} interactions that do occur in EPSR take place over a longer distance than in either MD or EPSR2. Other RDFs, for instance, $g_{\mathrm{O_1O_1}}(r)$ and $g_{\mathrm{O_NO_N}}(r)$, which have peaks at ∼4.4 Å and ∼4.9 Å, respectively, for MD and EPSR2, are almost flat for EPSR. In general, cocaine-cocaine interactions are insignificant in EPSR compared with MD and EPSR2. Fig. 4.8, which shows nearest neighbour benzene-benzene interactions, exemplifies the reduced number of cocaine-cocaine interactions in EPSR compared with MD.

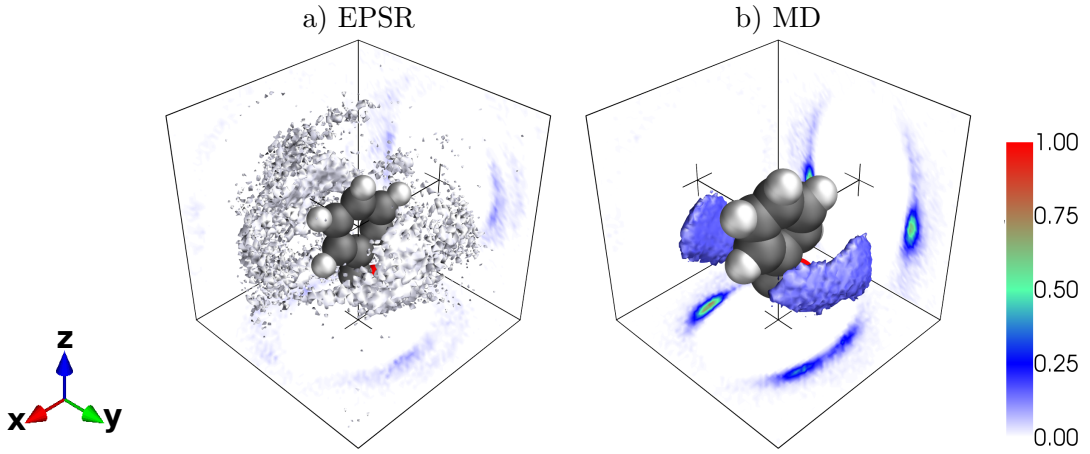
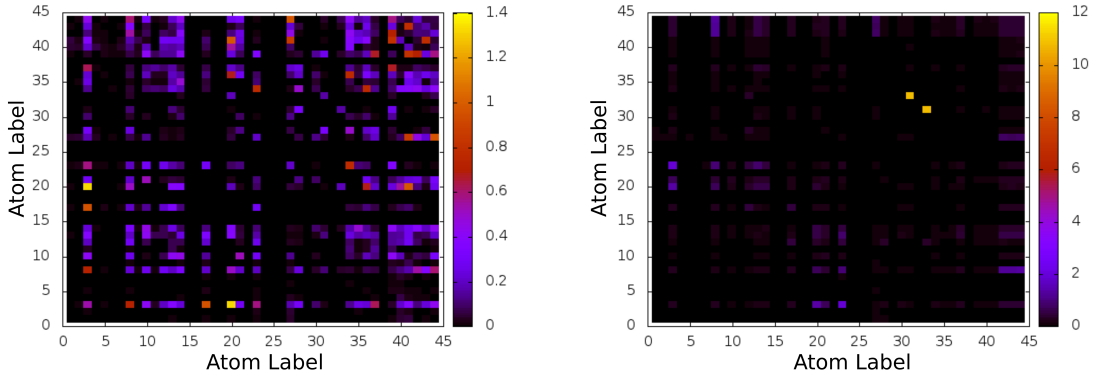


Figure 4.8: SDMs showing the distribution of nearest neighbour benzene rings of cocaine relative to the benzene ring of cocaine generated from EPSR and MD (SPC/E water).

RDFs in Fig. 4.7 for EPSR2 with varying temperature (EPSR2 ΔT) show that these temperature changes resulted in reduced cocaine-cocaine interactions. The peaks in the RDFs extracted from EPSR2 are located at the same positions as the previously described peaks in the RDFs extracted from MD (SPC/E and TIP3P water), but the size of these peaks is in each case smaller for EPSR2 ΔT compared with MD and EPSR2.



a) EPSR b) MD

Figure 4.9: IMAPs generated from EPSR (a) and MD (SPC/E water) (b) that show the favoured intermolecular interactions between cocaine atoms.

That the cocaine molecules associate with each other less frequently at elevated temperatures is unsurprising, but that they remain apart when the temperature is returned to the temperature of meaurement indicates that it is principally the influence of the starting structure from the MD trajectory that drives these molecular associations. Despite the reduction in cocaine-cocaine interactions in EPSR2 ΔT compared with MD and EPSR2 resulting from temperature increase, cocaine-cocaine interactions still remain more prominent in this simulation (11% of cocaine molecules associate *via* N-O_{D2} interactions) than in EPSR.

IMAPs corresponding to cocaine-cocaine interactions in EPSR and MD run using the SPC/E water model are shown in Fig. 4.9. The IMAP corresponding to MD shows that interactions between cocaine atoms occur over a wider range of probability densities in MD compared with EPSR. While in EPSR the most favoured atomic interaction is only roughly double the median density, in MD the most favoured interaction is almost an order of magnitude larger than the median density. The most favoured interaction for MD occurs between atoms 31 and 33 in the matrix. These atoms correspond to O_{D2} and H_P, respectively. This is in agreement with the RDFs presented in Fig. 4.7, where the largest peak is seen in $g_{NO_{D2}}(r)$ for MD at $\sim$2.9 Å.

Some aggregation of cocaine molecules may be anticipated given that caffeine, a fellow alkaloid, has been documented as undergoing considerable self-association in an aqueous environment, and molecular modelling and NMR studies have both pointed to-

wards dimerisation of caffeine molecules in aqueous solution.[282] However, the cocaine-cocaine interactions seen in Fig. 4.7 and the worse agreement between the simulated and experimental data in Fig. 4.5 compared with Fig. 4.4 suggest that the structure produced by EPSR is more physically realistic. Therefore, a comprehensive analysis of the behaviour of the cocaine molecules in aqueous solution was only conducted for EPSR. NMR was used (Section 4.3.6) to further investigate cocaine-cocaine interactions in attempts to experimentally verify which model was most physically realistic.

4.3.3 Cocaine Conformation

Due to the importance of the conformation and structure of molecules in understanding how drugs interact with the hydrophilic and hydrophobic components of the BBB, the conformation of the cocaine molecules in the EPSR simulation box was analysed. The 3D structure of molecules has been linked to BBB permeation and led to structure-based drug design (Section 1.1). To understand the conformation of the cocaine, which, unlike indole (Chapter 3), is a flexible molecule, angular and intraatomic information was extracted from the EPSR simulation box. Fig. 4.10 shows the torsional angles distributions of selected fragments of cocaine extracted from the EPSR simulation box, which were chosen as they correspond to fragments of the cocaine molecule able to rotate.

$\phi_{O_1-C_1-C_{B1}-C_B}$ indicates that the benzene ring of cocaine is within $\sim 50^\circ$ of the plane of O_1. Relative to the other parts of the cocaine molecule, the benzene ring is the most mobile in EPSR, showing a broad torsional angle distribution compared with other fragments of cocaine. In contrast, $\phi_{CC-O_1-C_1-C_{B1}}$ shows that the piperidine ring is at $\sim 30^\circ$ relative to the plane of the benzene ring. $\phi_{CC-O_1-C_1-C_{B1}}$ implies that cocaine rarely folds back on itself, such that any hydration of the benzene ring which may be deduced from the RDFs does not result from associations of water with other parts of the molecule. A similar distribution of angles is seen for the second methyl group. The angular distribution corresponding to this second methyl group places CM_2 above the ester linkage. $\phi_{C_C-C_{C1}-C_2-O_2}$ is $\sim 100^\circ$ and $\phi_{C_{C2}-C_{C1}-C_2-O_2}$ is $\sim 125^\circ$, which results in O_{D2} facing H_P.

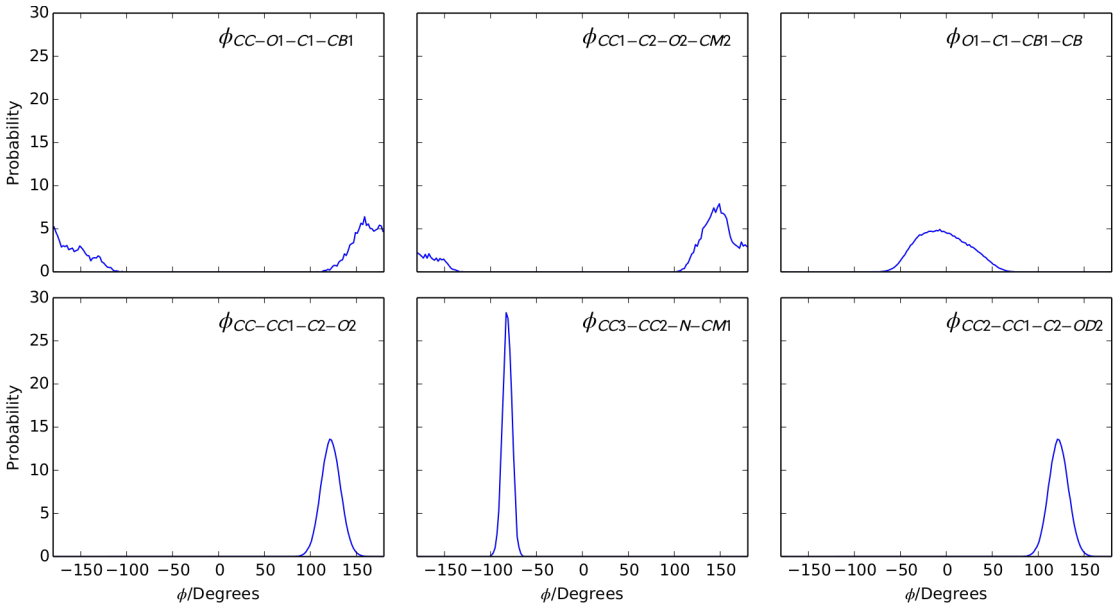


Figure 4.10: Selected torsional angles of the cocaine molecules in the EPSR simulation box.

Fig. 4.11 shows some of the relevant intramolecular RDFs for cocaine. The most prominent peak is found in $g_{\mathrm{O_{D2}H_P}}(r)$ at ∼1.83 Å. That this intramolecular peak is so well defined suggests a highly directed intramolecular hydrogen bond (IHB) between these two groups. The presence of this IHB confers conformational stability to cocaine, as it prevents rotation of the carbomethoxy group, in accordance with Fig. 4.10. The constraints afforded by IHBs are thought to influence the function of molecules.[123] The benzoate ester oxygen-amine hydrogen RDF ($g_{\mathrm{O_1H_P}}(r)$) also shows a prominent peak at ∼3.8 Å. While this distance is too long to be formally considered as a hydrogen bond, that this peak is relatively sharp indicates that this portion of the molecule is additionally constrained. It might be expected that a broader range of distances would be observed for $g_{\mathrm{O_1H_P}}(r)$, such as the range observed for the carbomethoxy ester oxygen ($g_{\mathrm{O_2H_P}}(r)$).

A depiction of the conformation of cocaine deemed the most probable based on torsional angle distributions and constrained to reflect the average distances in Fig. 4.11 is shown in Fig. 4.12. Fig. 4.12 illustrates that the amine group, the carbomethoxy carbonyl and the benzoate ester linkage form an almost ring-like stucture towards interior of the cocaine molecule. These groups are orientated in such a way that leaves them less solvent accessible compared with, for instance, the carbonyl oxygen on the benzoate

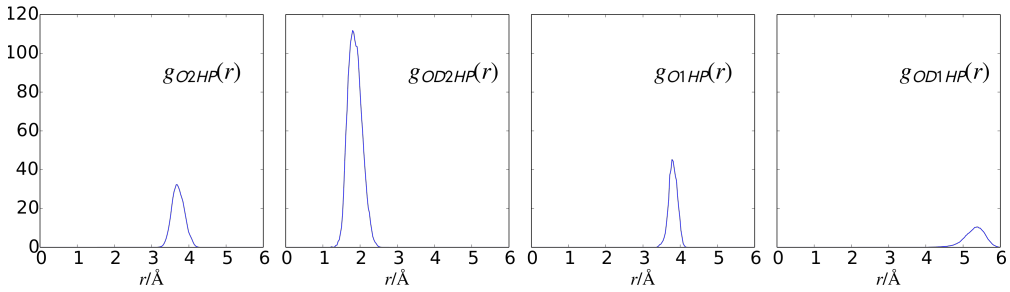


Figure 4.11: Intramolecular RDFs for cocaine oxygen atoms with the amide hydrogen atom.

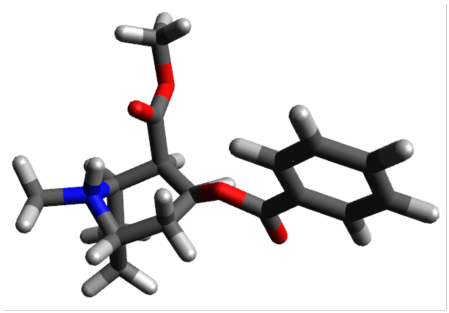


Figure 4.12: The most probable conformation of cocaine based on intramolecular distances (Fig. 4.11) and the angular distributions (Fig. 4.10) extracted from EPSR.

group.

4.3.4 Solvent Structure

Fig. 4.13 shows RDFs for the interactions between water molecules in the system. Corresponding RDFs extracted from an EPSR simulation of pure water are shown for comparison.[266] As can be seen from Fig. 4.13, the water structure is largely unaffected by the addition of cocaine molecules, the position and form of the peaks in the cocaine/water system being broadly consistent with those of pure water. For example, the first peak and second peaks in $g_{\mathrm{O_w O_w}}(r)$ are located at ∼2.8 Å and ∼4.8 Å, respectively, for cocaine/water and pure water. Similarly, the first and second peaks in $g_{\mathrm{H_w H_w}}(r)$ located at ∼2.4 Å and ∼3.9 Å, respectively, for cocaine/water and pure water. Similar results have been reported for other small molecules in solution that show that the water structure is unaffected by their presence.[34]

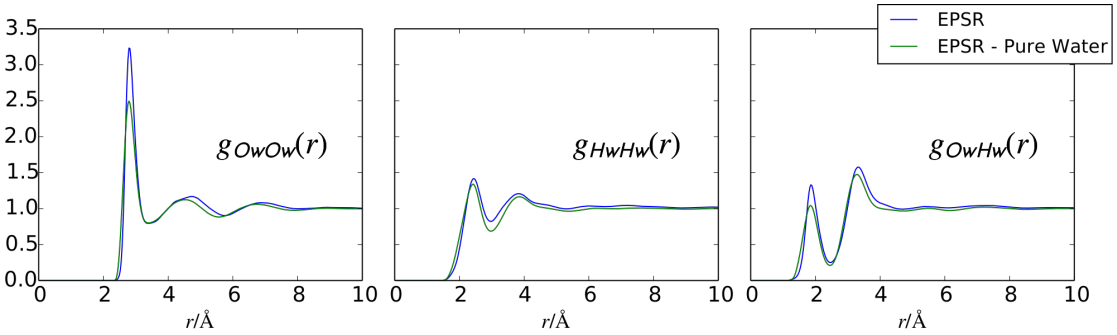


Figure 4.13: RDFs for water-water interactions extracted from EPSR (blue line) simulations of cocaine hydrochloride/water. Water-water RDFs extracted from EPSR simulations of pure water are also included (green line).[266]

4.3.5 Cocaine Hydration

4.3.5.1 Radial Distribution Functions

Carbonyl oxygen-water RDFs in Fig. 4.14 indicate that water is highly orientated around the carbonyl groups of cocaine. Both $g_{O_{D1}O_w}(r)$ and $g_{O_{D2}O_w}(r)$ have sharp peaks at ∼2.8 Å and both $g_{O_{D1}H_w}(r)$ and $g_{O_{D2}H_w}(r)$ show first sharp peaks at ∼1.77 Å, well within the accepted length range of a short, strong hydrogen bond.[274] The carbonyl oxygens are slightly undersaturated compared with pure water[266] and for carbonyl groups which are largely solvent accessible,[32, 76, 112] as indicated by the CNs in Table 4.5. A preference for water to interact with the carbonyl oxygens is expected given the electronegativity of O_{D1} and O_{D2}. CNs show the benzoate carbonyl oxygen (O_{D1}) has slightly more waters present compared with the carbomethoxy carbonyl oxygen (O_{D2}). This difference is mirrored in the CNs showing the number of water hydrogens around the carbonyl oxygens. The greater number of water molecules around O_{D1} compared with O_{D2} is consistent with O_{D2} being involved in IHB formation, as hypothesised in Section 4.3.3.

n_α^β	**EPSR**	r/Å	r_{peak}/Å
$n_{O_{D1}}^{O_w}$	4.15	3.93	2.76
$n_{O_{D1}}^{H_w}$	1.29	2.46	1.77
$n_{O_{D2}}^{O_w}$	3.63	3.93	2.85
$n_{O_{D2}}^{H_w}$	0.98	2.46	1.77

Table 4.5: CNs of RDFs shown in Fig. 4.14.

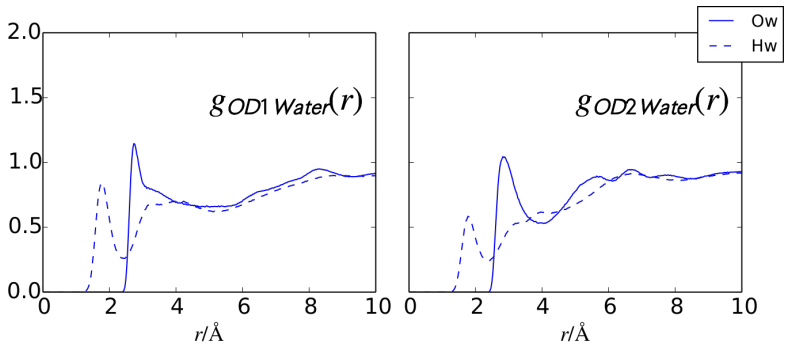


Figure 4.14: RDFs for the interactions between the carbonyl oxygens of cocaine (O_{D1} and O_{D2}) and water molecules extracted from EPSR.

Fig. 4.15 shows RDFs for the interactions between the ester oxygens of cocaine and water. The broad peaks in $g_{O_1O_w}(r)$ and $g_{O_2O_w}(r)$ are indicative of water being less orientated around these groups compared with the carbonyl oyxgens in Fig. 4.14. This is clearest for the benzoate ester oxygen (O_1), which has a broader function between ∼2 Å and ∼6 Å than the carbomethoxy ester oxygen (O_2). CNs for $g_{O_1O_w}(r)$ and $g_{O_2O_w}(r)$ in Table 4.6 show that there are fewer waters around O_1 compared with O_{D1}. In contrast, there are slightly more waters around O_2 compared with O_{D2}. This hydration is consistent with the observation from the most probable cocaine conformation depicted in Fig. 4.12 that H_P, O_{D2} and O_1, given the average conformation of cocaine in the EPSR simulation box, are less solvent accessible.

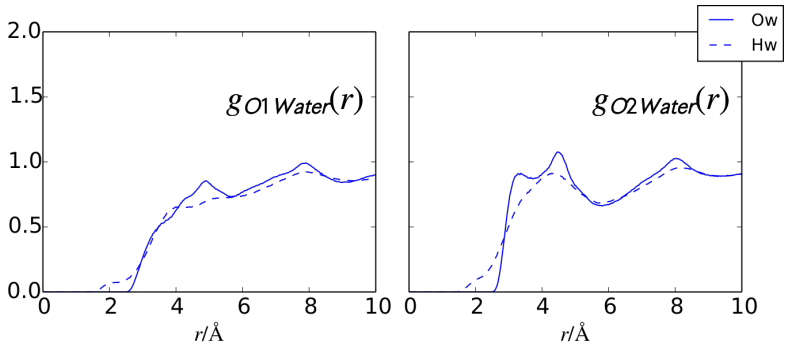


Figure 4.15: RDFs for the interactions between the ester oxygens of cocaine (O_1 and O_2) and water molecules extracted from EPSR.

$g_{H_PO_w}(r)$ in Fig. 4.16 has a clear peak at ∼2.19 Å. The corresponding CN in Table 4.7 suggests that ∼0.9 water molecules are bound to H_P. The bond length indicated by the peak in $g_{H_PO_w}(r)$ is consistent with a moderately strong hydrogen bond[274] from the amine hydrogen (H_P) to the surrounding water molecules. The pK_a for this proton was calculated from pH measurements of cocaine in water at different concentrations

n_{α}^{β}	**EPSR**	r/Å	r_{peak}/Å
$n_{O_1}^{O_w}$	2.07	3.93	4.86
$n_{O_1}^{H_w}$	0.16	2.46	-
$n_{O_2}^{O_w}$	3.80	3.93	3.33
$n_{O_2}^{H_w}$	0.28	2.46	4.38

Table 4.6: CNs of RDFs shown in Fig. 4.15.

as 8.5. This value is in agreement with other values in the literature[314] and implies that all of the cocaine molecules are fully protonated, as solutions measured here have a pH of ∼3.3. As a reference, the pH of neuronal cells of the vertebrate CNS ranges from around 6 to 8,[44] so that, similar to the current work, cocaine would exist in its protonated form over this range of different pHs within the human CNS. Hydrogen bonding from the amine group to water is also visible in $g_{NO_w}(r)$ at ∼3.0 Å. The peaks at larger distances (∼4.4 Å and ∼5.5 Å) are likely related to a second and third hydration shell around cocaine.

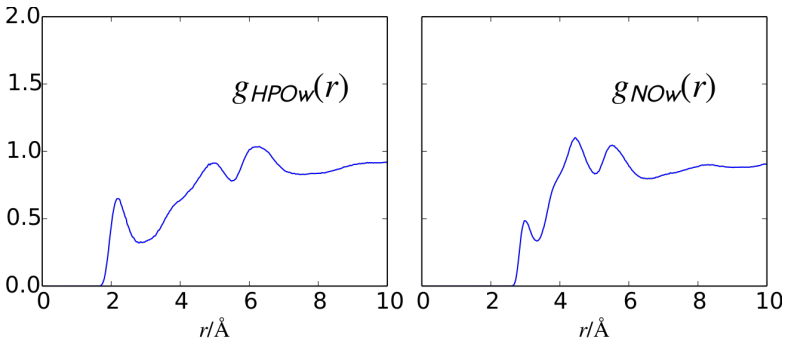


Figure 4.16: RDFs for the interaction between the amine group of cocaine (N and H_P) and water molecules extracted from EPSR.

n_{α}^{β}	**EPSR**	r/Å	r_{peak}/Å
$n_{H_P}^{O_w}$	0.90	2.85	2.19
$n_{N}^{O_w}$	0.73	3.33	3.00

Table 4.7: CNs of RDFs shown in Fig. 4.16.

Fig. 4.17 shows the interactions between the benzene ring of cocaine and water molecules. The broad peak at ∼3.6 Å in $g_{BO_w}(r)$ suggests that the waters surrounding benzene might have some preferred orientation. However, the much broader peak in $g_{BH_w}(r)$ indicates that this is not the case. Although water hydrogens are within a closer distance range than water oxygens, these hydrogens have a broad distribution of

possible nearest neigbour distances around benzene, which is indicative of disordered interactions around this part of the cocaine molecule. This observation is at variance with the results presented in Section 3.3 and also with other investigations of molecules whose structures include benzene rings, which suggest more directed interactions between benzene and water molecules.[278, 315] It may be that the waters around benzene are not as localised, because cocaine affords more direct electrostatic hydrogen bonding opportunities compared with indole or methanol in solution.

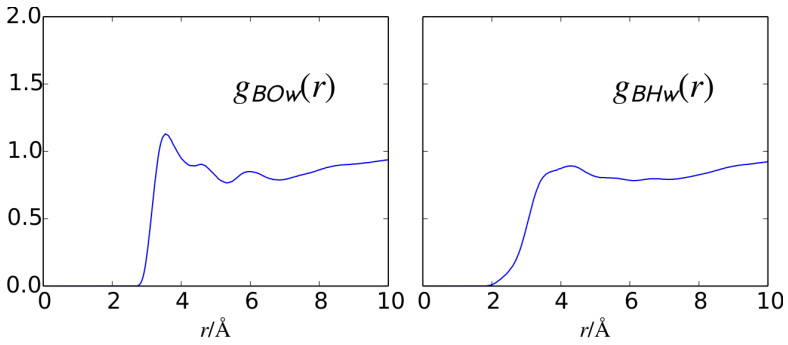


Figure 4.17: RDFs for the interactions between the benzene ring of cocaine and the water molecules (i.e. the weighted average of interactions between C_B and C_{B1} with O_w or H_w) extracted from EPSR.

n_α^β	**EPSR**	r/Å	r_{peak}/Å
$n_B^{O_w}$	6.04	4.32	3.54
$n_B^{H_w}$	21.69	5.19	4.29

Table 4.8: CNs of RDFs shown in Fig. 4.17.

4.3.5.2 Spatial Density Maps

SDMs showing the distribution of the nearest neighbour waters relative to the carbonyl oxygens of cocaine are shown in Fig. 4.18. Comparing these SDMs indicates that the areas of highest density of water molecules around O_{D1} (0.85 Å^{-3}) are less dense than the areas of highest density of water molecules around O_{D2} (1.79 Å^{-3}). Moreover, the waters around the carbomethoxy carbonyl oxygen exhibit a more well defined, localised distribution than the waters around O_{D1}. The highest density of waters around O_{D2} is observed in the -**x** direction, as seen in the cut-throughs of this plane, towards the tropane ring of cocaine, where O_{D2} is proposed to be involved in an IHB with the amine hydrogen. The SDM showing the distribution of the nearest neighbour waters relative

to O_{D2} also implies a comparatively low probablity of waters being directly above O_{D2} in the **z** direction. In contrast, the distribution of nearest neighour waters relative to O_{D1} is broader and symmetric around this atomic site.

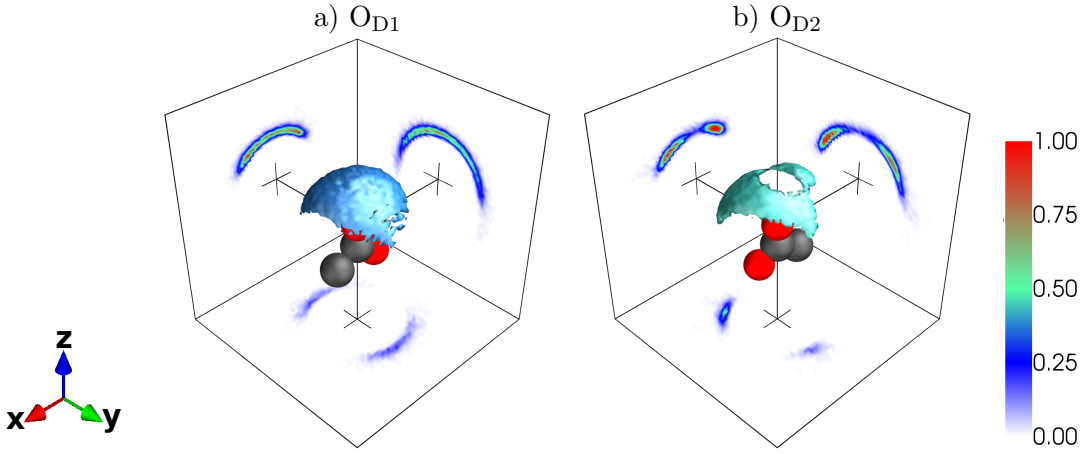


Figure 4.18: SDMs showing the distribution of nearest neighbour water oxygens around O_{D1} and O_{D2}.

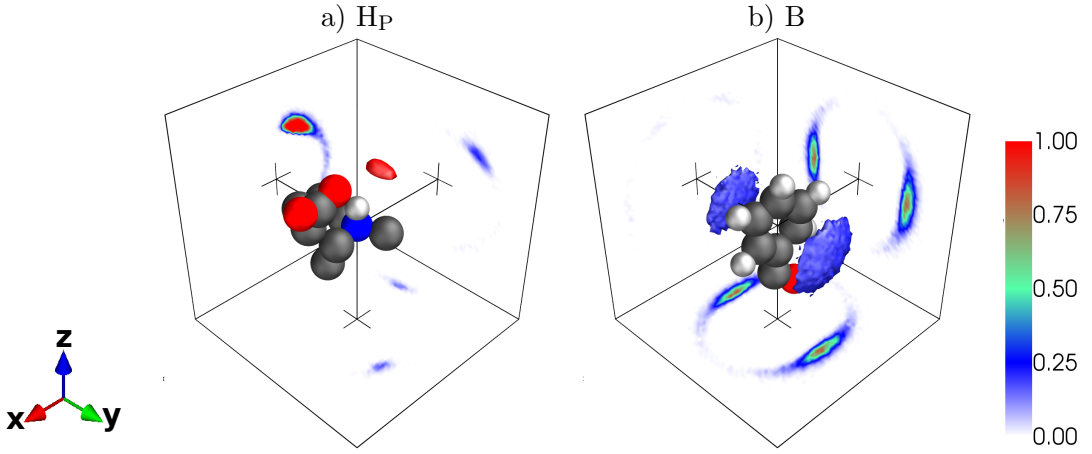


Figure 4.19: SDMs showing the distribution of nearest neighbour water oxygens around H_P and the benzene ring.

Fig. 4.19 shows the distribution of nearest neighbour water molecules relative to H_P and the benzene ring of cocaine. In Fig. 4.19 (a) the nearest neighbour water molecules are all located above H_P, slightly offset in the $-\mathbf{x}$ direction. This localised water density suggests highly orientated water molecules which bind at very specific locations to the amine group of cocaine. The slightly offset position of the water density in Fig. 4.19 (a) means that, rather than being directly above the NH group, the waters are more likely to be located in towards the proposed IHB formed between the amine group and the carbomethoxy oxygen. This suggests that the IHB in cocaine is either water mediated

to some degree or strongly coordinates with water at specific locations around this hydrogen bonding moiety. This interpretation is consistent with Fig. 4.18 (b), where the most likely location of the nearest neigbour waters indicates that water is more likely to be found between the amine and carbomethoxy carbonyl groups than at other locations.

From the SDM in Fig. 4.19 (a), it is clear that the nearest neighbour waters are preferentially located above and below the plane of the benzene ring. However, the distribution of water around benzene is more diffuse compared with the carbonyl and the amine groups.

4.3.5.3 Whole Molecule Analysis

WMA for water within 4 Å of cocaine (Fig. 4.20) shows that water occupies the space between O_{D2} and H_P in the tropane ring over other parts of the molecule. These water molecules appear to preferentially envelop the cavity where IHB formation is suggested to occur, the top 30% of water molecules being located in this cavity. In this Fig. 4.20, the limited movement of the carbomethoxy group is highlighted by the inclusion of a number of cocaine molecules from the EPSR simulation box being plotted on top of one another. In contrast, these central molecules show that the benzene ring is able to move comparatively freely and samples a wider range of conformations throughout the simulation, as was first indicated by Fig. 4.10 ($\phi_{O_1-C_1-C_{B1}-C_B}$).

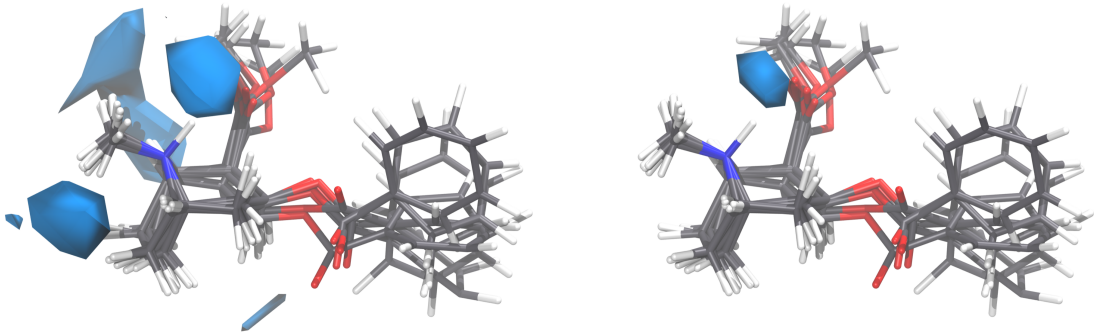


Figure 4.20: WMA for water (blue) within 4 Å of cocaine. The isocontour surface is cut to show the top 50% (left) and top 30% (right) of water molecules. Different cocaine molecules taken from the EPSR simulation box are shown at the origin to display the range of conformations adopted by cocaine.

The most probable position and orientation of the nearest neighbour water around

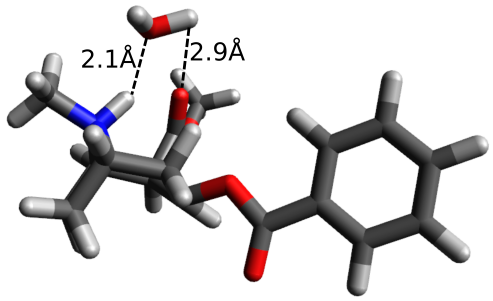


Figure 4.21: The position and orientation of the most probable nearest neighbour water molecule of H_P. The cocaine molecule shown in this figure corresponds to the most probable conformation of cocaine, as shown in Fig. 4.12.

H_P extracted from the EPSR simulation box using the ANGULA is shown in Fig. 4.21. The position and orientation of this water emphasises that the IHB between the amine hydrogen and the carbomethoxy carbonyl group is likely to be mediated by or strongly coordinated with a water molecule, as suggested by the water density shown around this group in Fig. 4.19 (a), where this water is orientated toward the IHB.

On the surface, the results presented in Fig. 4.20 and Fig. 4.21 are seemingly contradictory to the slightly higher solvent accessibility observed for O_{D1} relative to O_{D2} in Fig. 4.14. However, it should be noted that the WMA represents most probable water location around all of the atoms in the entire cocaine molecule within 4 Å rather than being specific to a single site. As stated above, the most probable location of waters around the entire molecule is between the O_{D2} and H_P atoms, so this distribution is not solely representative of the hydration of water around one or the other of these atoms but rather shows that the most favoured location of water is between these atoms. Furthermore, a more highly solvent accessible atom would show a broader distribution of waters around a given group, as was seen when comparing Fig. 4.18 (a) and Fig. 4.18 (b), whereas a lower solvent accessibility indicates that there are fewer locations that the water molecules can bond to around a given atom but does not necessarily mean that there is a lower probability of finding water molecules in this location relative to the entire molecule.

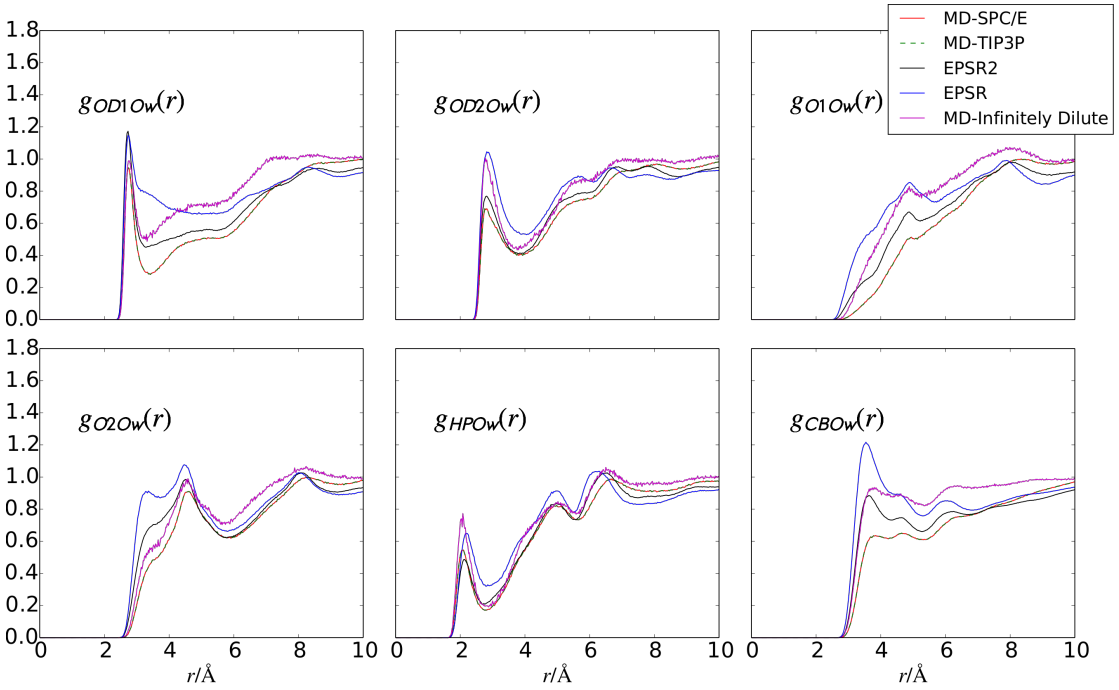


Figure 4.22: RDFs for cocaine-water oxygen interactions extracted from MD simulations run using SPC/E (red line) and TIP3P (dashed green line) water models, EPSR2 (black line) and EPSR (blue line). Additionally, cocaine-water oxygen RDFs extracted from an MD simulation run at an infinite dilution (magenta line) are included.

4.3.5.4 Cocaine Hydration in MD Compared with EPSR

As referred to in Section 4.3.1 and Section 4.3.2, complementary MD simulations that were run in order to provide an independent comparison with EPSR were not analysed to the same extent as EPSR because of the large number of cocaine-cocaine interactions. However, in this Section, for completeness, the hydration experienced by cocaine across the different simulations is compared. This comparison is made for interactions between the atoms of cocaine whose hydration was analysed in Section 4.3.5.1 and water atoms. RDFs extracted from MD (SPC/E and TIP3P water), EPSR2 and EPSR are shown in Fig. 4.22 and Fig. 4.23 for cocaine-water interactions. RDFs extracted from an infinitely dilute (1 cocaine: 1000 water) MD simulation are also plotted to see how the absence of cocaine-cocaine interactions affects the hydration of the molecule.

RDFs in Fig. 4.22 and Fig. 4.22 extracted from the MD simulations run using both the SPC/E and TIP3P water models indicate that the hydration of the cocaine molecule is unaffected by the water model used. The shape of the functions and the height of their peaks is the same for MD simulations run using both water models. This is in agreement

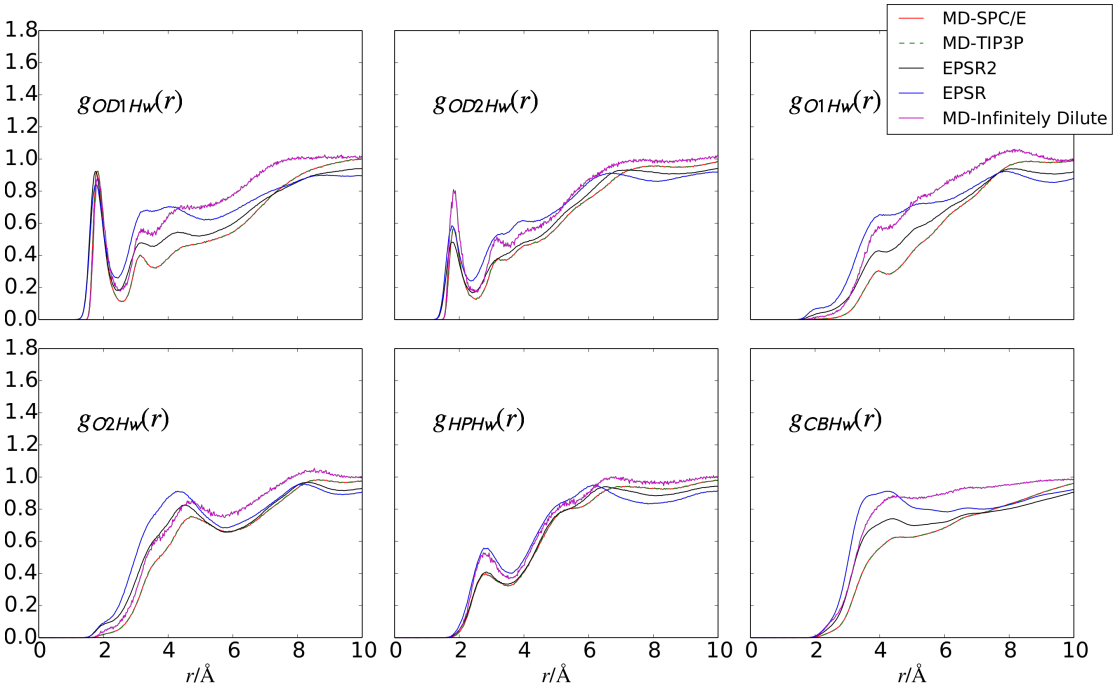


Figure 4.23: RDFs for cocaine-water hydrogen interactions extracted from MD simulations run using SPC/E (red line) and TIP3P (dashed green line) water models, EPSR2 (black line) and EPSR (blue line). Additionally, cocaine-water hydrogen RDFs extracted from an MD simulation run at an infinite dilution (magenta line) are included.

with the $F(Q)$s displayed in Section 4.3.1, where no difference in the agreement of experimental and simulated data is observed that is related to the choice of water model. Furthermore, cocaine-cocaine interactions in Section 4.3.2 remained unchanged with water model. The location of peaks in the RDFs extracted from MD, EPSR2 and EPSR is very similar for the interactions between cocaine atoms and water atoms shown in Fig. 4.22 and Fig. 4.23. For example, peaks at $\sim$1.8 Å and $\sim$2.2 Å are observed for each simulation in RDFs corresponding to carbonyl oxygen-water oxygen and amine hydrogen-water oxygen interactions, respectively. This consistency of peak position means that although there may be changes in the hydration of the cocaine molecule in each of the simulations, the average preferred distance of the water molecules from cocaine is constant. A similar result was found for indole dissolved in methanol/water solutions (Section 3.3), where, although the number of solvent molecules interacting with the solute varied with simulation technique, the nature of the solvation of the solute was consistent across EPSR and MD.

Although the location of the peaks in the RDFs extracted from MD, EPSR2 and EPSR is consistent, there appears to be differences in the degree of hydration experienced

n_{α}^{β}	**EPSR**	**EPSR2**	**MD**	r/Å
$n_{O_{D1}}^{O_w}$	4.15	2.87	2.06	3.93
$n_{O_{D1}}^{H_w}$	1.29	1.21	1.02	2.46
$n_{O_{D2}}^{O_w}$	3.63	2.61	2.41	3.93
$n_{O_{D2}}^{H_w}$	0.98	0.76	0.70	2.46
$n_{H_P}^{O_w}$	0.90	0.58	0.59	2.85
$n_{H_P}^{H_w}$	3.56	2.56	2.49	3.51
$n_{C_B}^{O_w}$	6.67	4.79	3.52	4.41
$n_{C_B}^{H_w}$	23.72	18.30	14.81	5.25

Table 4.9: CNs of the RDFs extracted from EPSR, EPSR2 and MD (SPC/E water) shown in Fig. 4.22 and Fig. 4.23.

by the cocaine molecules in these simulations. These differences can be investigated by considering the CNs (Table 4.9). Table 4.9 does not include CNs for the MD simulation run using the TIP3P water model, as the RDFs extracted from the MD simulation run using the TIP3P water model are practically identical to those extracted from the MD simulation run using the SPC/E water model. From Table 4.9 it can be seen that there is a trend for increased hydration of the cocaine atoms in EPSR2 compared with MD, with cocaine molecules in EPSR experiencing greater hydration than cocaine molecules in either EPSR2 or MD. The largest change appears to occur for $g_{O_{D1}O_w}(r)$, where there is almost, on average, an extra water oxygen interacting with O_{D1} in EPSR2 compared with MD and over twice as many water oyxgens hydrating O_{D1} in EPSR compared with MD. Accordingly, $n_{O_{D1}}^{H_w}$ also increases by ∼20% in EPSR2 compared with MD and decreases by 6% compared with EPSR. CNs relating to the carbomethoxy carbonyl group are also largest in EPSR and larger in EPSR2 compared with MD, although neither $n_{O_{D2}}^{O_w}$ nor $n_{O_{D2}}^{H_w}$ increase to the same extent as $n_{O_{D1}}^{O_w}$. $n_{O_{D2}}^{O_w}$ is ∼50% larger in EPSR compared with MD. Despite a small but observable difference in the height of the peaks located at ∼2.2 Å and ∼2.8 Å in $g_{H_PO_w}(r)$ (Fig. 4.22) and $g_{H_PH_w}(r)$ (Fig. 4.22), respectively, $n_{H_P}^{O_w}$ and $n_{H_P}^{H_w}$ are almost unchanged in EPSR2 compared with MD. In contrast, there is, on average, an extra water hydrogen bound to the amine hydrogen of cocaine in EPSR compared with EPSR2. The benzene ring of cocaine is also significantly more hydrated in EPSR compared with EPSR2 or MD.

RDFs for cocaine-water interactions extracted from an MD simulation of cocaine

dissolved in water at an infinite dilution (Fig. 4.22 and Fig. 4.22) exhibit a similar form to the RDFs extracted from EPSR, EPSR2 and MD, with their peak positions the same as the RDFs extracted from the other simulations. This indicates that the atomic locations at which cocaine is preferentially hydrated in solution are not affected by interactions between cocaine molecules, although cocaine-cocaine interactions vary across the different simulations (Section 4.3.2).

4.3.6 NMR

4.3.6.1 Assignment in Water and Methanol

Fig. 4.24 shows spectra of cocaine hydrochloride dissolved in methanol and in water at the same concentration (1 cocaine: 150 solvent), roughly twice as dilute as the samples measured using the SANDALS instrument. The spectra in Fig. 4.24 were assigned using a combination ^{1}H-^{1}H COSY and ^{13}C-HSCQ spectra (Appendix B.2). As can be seen from Fig. 4.24, there is no change in the chemical shift of the peaks in methanol compared with water, indicating that the molecule adopts the same conformation in aqueous and organic environments at this concentration. The amine hydrogen in not visible in Fig. 4.24 because of exchange with the solvent.

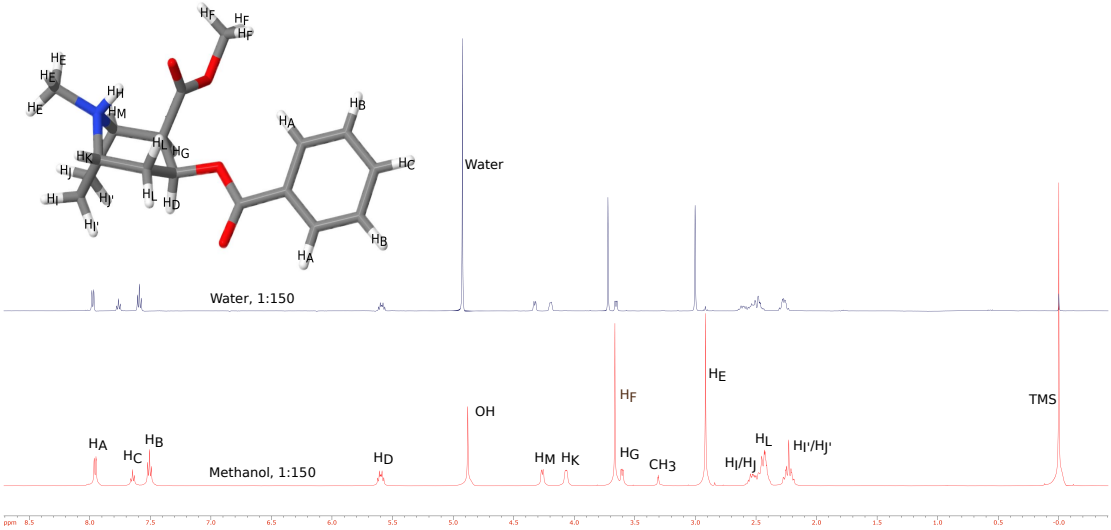


Figure 4.24: ^{1}H NMR spectra of cocaine hydrochloride dissolved in D_2O (blue) and CD_3OD (red) at a molar ratio of 1 cocaine: 150 solvent.

4.3.6.2 The Effect of Temperature

NMR spectra recorded in order to confirm that performing diffraction measurements at a temperature of 298 K did not mean a different stucture was measured to that which exists at body temperature are shown in Fig. 4.25. Fig. 4.25 indicates that there is no change in the structure of cocaine in this temperature range. The only change between the spectra is a chemical small shift observed for the water peak because of the temperature dependence of the dielectric constant of water.

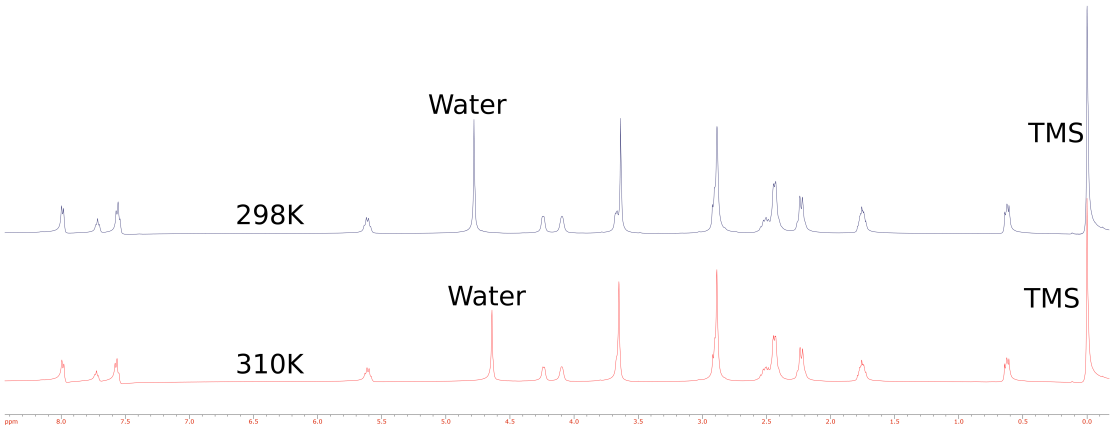


Figure 4.25: ^{1}H NMR spectra of cocaine hydrochloride dissolved in D_2O at a concentration of 1 cocaine:1000 water at 298 K (blue) and 310 K (red).

4.3.6.3 Concentration Effects

The results of the NMR measurements performed to investigate possible concentration related effects are shown in Fig. 4.26. This figure shows spectra of cocaine dissolved in D_2O sequentially diluted from a concentration of 1:200 (275 mM) to a concentration of 1:1000 (55 mM). The pH of the second most concentrated sample and the most diluted sample are noted in Fig. 4.26. A spectrum of the 1:1000 sample measured at a pH reduced to the approximate pH of the second most concentrated sample is also shown in Fig. 4.26, so that comparisons could be made between spectra across both varying concentration and pH.

Comparing the spectra across the different concentrations of cocaine indicates a change in the chemical shift of the peaks with concentration. This is most obvious through the relative positions of H_G and H_F in the spectra, which swap when the initial

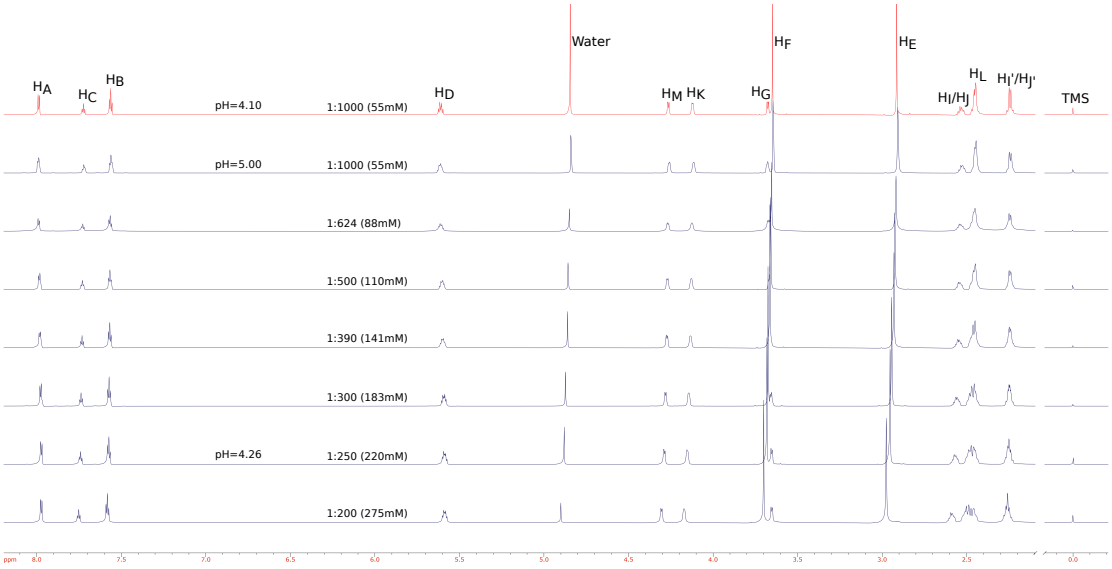


Figure 4.26: ^{1}H NMR spectra of cocaine hydrochloride dissolved in D_2O at different concentrations. The pH of the second most concentrated and the most diluted sample are noted. The spectrum of a sample whose pH was reduced to that of the second most dilute sample is included (red).

sample is diluted by a factor 5. However, the chemical shifts of all of the peaks in the spectrum of the most concentrated sample in Fig. 4.26 change when the sample is diluted. This is made clear by Fig. 4.27, which shows the change in the chemical shift of all of the peaks with concentration relative to the most concentrated sample. H_D, H_G and H_A all shift downfield as the concentration of the sample is decreased, whereas the remaining 6 peaks shift upfield as the concentration decreases. The peak whose chemical shift changes by the most is H_E, which is located at a ppm $\sim$0.07 less at a concentration of 275 mM compared with 55 mM.

Measurements made to test whether these changes were a result of the changing ionic strength of the samples measured in Fig. 4.26 showed constant chemical shifts of the peaks with varying ionic strength. Similarly, the time spent in solution was eliminated as a factor that could contribute to the results in Fig. 4.26. These tests, coupled with the absence of any changes in the spectrum when the pH of the most dilute sample was reduced to that of the most concentrated sample, affirm that the changes in the spectra in Fig. 4.26 are indeed due to changes in the concentration.

The change in the chemical shift of the peaks of cocaine seen in Fig. 4.26 and Fig. 4.27 could perhaps be indirectly related to greater hydration of the cocaine molecule, which

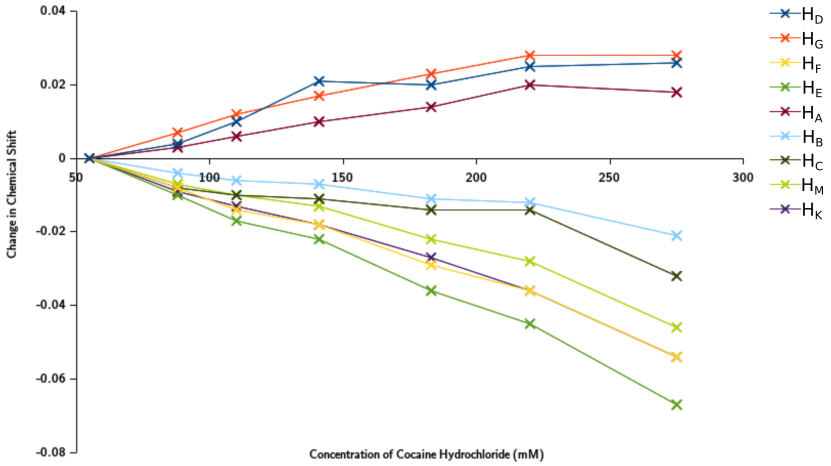


Figure 4.27: The chemical shifts of each of the peaks in Fig. 4.26 relative to the chemical shifts of these same peaks in the most concentrated sample.

is proposed in Section 4.3.5 to influence conformation, at lower concentrations, since the likely reduced cocaine-cocaine interactions would make the solute more solvent accessible, as evidenced by comparison of the hydration experienced by cocaine in EPSR, EPSR2 and MD (Section 4.3.5.4). However, the most concentrated sample in Fig. 4.26 is still more dilute than the samples used in neutron diffraction measurements (Section 4.3.5), where there were significant cocaine-water interactions. Moreover, Section 4.3.5.4 underlines how even when a large proportion of the cocaine molecules associate, the general picture of hydration is preserved.

Consequently, it might be considered that cocaine-cocaine interactions, which are more likely to occur at high concentrations, directly influence that chemical shift of the peaks in NMR spectra. Previous studies have suggested concentration dependencies in the chemical shifts of non-exchangeable protons and eliminated conventional π-π stacking as a factor contributing to this phenomenon, ascribing chemical shifts at high concentrations to molecular aggregation leading to the formation of differently packed supramolecular assemblies.[185] As the concentrations of the sample in Fig. 4.26 is reduced these structures are likely to be disrupted, affecting the chemical shifts of the peaks in the NMR spectrum. Molecular aggregation may be expected to occur *via* the amine-carbomethoxy interaction described in Section 4.3.2, meaning that hydrogens involved in this interaction would be expected to undergo the greatest chemical shift in Fig. 4.27. However, it is difficult to ascertain to what what extent this is the case, as the amine hydrogen is not visible in Fig. 4.26 due to exchange with the solvent. Even

if spectra were recorded where the amine hydrogen were visible, the chemical shift of amine hydrogens, in general, is known to be concentration dependent.

Concentration related chemical shift changes were not observed when cocaine was dissolved in methanol over the same range of concentrations as in Fig. 4.26 and Fig. 4.27. This may suggest that, in fact, water plays a role in coordinating these supramolecular assemblies. This finding is consistent with the reported entropically favourable contributions that arise from the aggregation of π-π systems in aqueous solution that are minimised in organic solvents.[114] Given the supposed involvement of the amine hydrogen in facilitating the aggregation resulting in these chemical shifts, an analogous analysis of the results of the freebase form of cocaine dissolved in water may be anticipated to be useful. However, such an analysis is precluded by the low solubility of the freebase form of cocaine in water. Despite highlighting the role water may play in determining the behaviour of cocaine in solution, no clear evidence of the extent of the cocaine-cocaine interactions in an aqueous environment was obtained from chemical shifts in the NMR spectra.

To verify the effects of the concentration in aqueous solution on the aggregation of cocaine molecules, the diffusion coefficients of cocaine in aqueous solution were determined for samples at different concentrations.[c] These diffusion coefficients are shown in Table 4.10. As the diffusion coefficient of a molecule increases with reduced concentration, the diffusion coefficient of TMS, which was at the same concentration in each solution, was also computed for each solution. The ratio of the diffusion coefficient of cocaine was then calculated relative to the diffusion coefficient of TMS for each solution. The results of this calculation show an absence of any trend for increasing $D_{Cocaine}$: D_{TMS} with increasing cocaine concentration, suggesting a lack of an increase in the frequency of cocaine-cocaine interactions with concentration. However, there are insufficient data points make firm conclusions regarding whether aggregation increases with concentration over this concentration range for cocaine in aqueous solution.

The radius of hydration of a molecule in solution may be determined from the Stokes-Einstein equation[d] if the diffusion coefficient of the molecule and the viscosity of the

[c] I would like to thank Tim Claridge for performing these calculations.

[d] $R = \frac{k_B T}{6\pi\eta D}$, where R is the radius of hydration, k_B is the Boltzmann constant, T is temper-

Moles of Water	**Concentration (M)**	$\mathbf{D}_{\text{Cocaine}}(10^{-10}\text{m}^2\text{s}^{-1})$	$\mathbf{D}_{\text{Cocaine}}$**:** $\mathbf{D}_{\text{TMS}}$
65	1.000	2.73	1.05
200	0.250	4.05	0.89
500	0.100	4.60	0.92
2000	0.025	5.52	0.88

Table 4.10: The diffusion coefficients ($D_{Cocaine}$) for different concentrations of cocaine hydrochloride in aqueous solution as determined from NMR. The ratio of the diffusion coefficient of cocaine to the diffusion coefficient of TMS (D_{TMS}) is also included for each solution.

solution are known.[99] Consequently, changes in this radius can be related to the aggregation of solute molecules in solution, with an increase in radius being indicative of greater aggregation of solute molecules. However, the Stokes-Einstein equation is only valid for spherical solute molecules. Also, the viscosity of the solutions in Table 4.10 are not known. Careful measurement of the viscosity of cocaine hydrochloride in aqueous solution over a range of concentrations would be required in order to accurately compute the radius of hydration and make deductions regarding the extent of cocaine-cocaine interactions and their relationship to the concentration of cocaine in aqueous solution.

4.4 Conclusions

Cocaine is very water soluble, yet water seems to be strongly associated with few of its atoms. Highly directed hydrogen bonds only occur between water and the carbonyl oxygens or the amine group, with very diffuse interactions being observed betweeen water and other regions of the cocaine molecule. The high solubilty of cocaine might mean that it would be expected that water be able to form highly directed hydrogen bonds with the ester oxygens or the benzene ring in solution, as has been previously observed for some molecules.[126, 200] However, these interactions are apparently absent for cocaine. This is not to say that there are no waters present, but rather that the waters present around these groups are disordered. It has previously been suggested that once the cumulative number of hydrogen bonds a drug forms with water exceeds 8-10, the BBB transport of the molecule will probably be pharmacologically insignificant.[205]

ature, η is viscosity and D is the diffusion coefficient.

In this system, significant hydrogen bonds are accepted from water from O_{D1}, O_{D1} and N atoms. Summing the CNs for these atoms gives a value of 5.4, below the reported threhold above which molecules may not cross the BBB. The H_P atom on the average also donates 0.9 hydrogen bonds to the surrounding water solvent.

Torsional angle information and intramolecular RDFs indicate that the carbomethoxy group on cocaine is rotationally restricted, held in place due to an IHB between the carbonyl oxygen and amine hydrogen. The results here also indicate that this IHB is strongly associated with water, where the waters around this hydrogen bonding motif are highly orientated. In addition, the carbomethoxy ester oxygen (O_2) and benzoate carbonyl oxygen (O_{D1}) show significantly higher hydration, compared with the benzoate ester oxygen (O_1) and the carbomethoxy carbonyl oxygen (O_{D2}), respectively. The reason for this is apparent from inspection of the most probable orientation of cocaine in solution (Fig. 4.12), where both of these hydrated oxygen atoms are orientated such that they are easily accessible to the surrounding water solvent. Conversely, the other hydrophilic atoms (H_P, O_{D2} and O_1) are orientated so that they are facing towards each other and largely away from the surrounding water solvent. These groups are comparatively less hydrated, but the waters around these atoms are highly orientated, suggesting a very strong association between water and these more 'internal' hydrophilic atoms. The IHB in cocaine appears to either be mediated by a water molecule whose hydrogen atoms are $\sim$2.8 Å from O_{D2} and H_P or to strongly bind to a water molecule rather than excluding water as might be expected from the formation of an IHB. It should be noted that previous crystallographic investigations of cocaine hydrochloride do not show the presence of an IHB, emphasising the need for investigations in the liquid phase.[79]

The combination of intramolecular bonding and the apparent affinity for cocaine to tightly bind water to the atoms associated with this non-covalent bond provides a possible explanation for the mechanism by which cocaine is able to cross the BBB. In opposition to previous thought that cocaine must be deprotonated in order to effectively passively diffuse into the BBB,[316] the hydration structure observed here suggests another possibility. In solution, the charged regions associated with the IHB present in cocaine appear to be effectively shielded by the conformation of cocaine. In addition, the formation of this internal hydrogen bond also strongly coordinates water molecules,

which may also allow the benzoate ester oxygen to be pulled towards the inside of the cocaine molecules. This water-mediated conformation would effectively shield charges from an external hydrophobic environment such that the cocaine molecule becomes sufficiently lipophilic to cross the BBB. Water does not appear to be tightly bound to either the benzene ring or around the carbomethoxy ester, both of which are comparatively more solvent accessible. This might indicate that water would be easily shed from these regions of the molecule when cocaine is in a lipophilic environment.

It has been previously noted that IHBs can confer lipophilicity to drugs by shielding potential hydrogen bonding motifs from the solvent[153] and that IHBs were most probable when they resulted in the formation of a 5 or 6 membered ring.[116] The most probable conformation of cocaine observed here, coupled with the water mediation interactions, certainly suggest that cocaine could easily adopt a more complex ring structure, where these rings would include water molecules, and would thus help confer lipophilicity to the cocaine molecule. Water association with the IHB in cocaine may also have a more direct biological function than solely aiding diffusion across the BBB. Water-mediated interactions in cocaine have been previously suggested to help facilitate its function, where it has been proposed that water mediates cocaine interactions with antibodies *in vivo*.[158] These conserved waters are in a similar location around the tropane ring to those observed here.

The ability of cocaine to shield several of its potential hydrogen bonding atoms, perhaps even trapping a molecule within its structure, suggests that cocaine is highly lipophilic even when it is protonated. However, contrary to previous suggestions that more conformationally open forms drugs should be relatively more water-soluble,[153] the results here also show that cocaine is still readily water soluble in a more closed conformation. Furthermore, the results here suggest that cocaine can easily achieve both lipophilicity and hydrophilicity simultaneously without having to adopt a different conformation in either physical regime or having to go through a protonation-deprotonation reaction to effectively permeate the BBB.

Chapter 5

Alprazolam in Methanol/water Solutions

5.1 Introduction

Alprazolam (Xanax^{TM}) belongs to a class of pyschoactive molecules called benzodiazepines, whose core chemical structure consists of a fused benzene ring and diazepine ring. Other common benzodiazepines, the most prescribed pharmaceuticals globally,[101, 214] include diazepam (Valium^{TM}) and oxazepam (Serax^{TM}). Benzodiazepines are used to treat a range of disorders including anxiety, insomnia and epilepsy due to their sedative, amnesic and muscle relaxing properties.[105, 231] The major clinical advantages of benzodiazepines are high efficacy rates, rapid onset of action and minimal toxicity.[231]

The pharmacological effects of benzodiazepines are primarily related to the CNS.[113] Electrophysiological investigations have suggested that benzodiazepines enhance the actions of the neurotransmitter γ-aminobutyric acid (GABA) on its receptor[251] after they have crossed the BBB by passive diffusion.[127] The binding of GABA to its receptor produces hyperpolarisation and neuronal inhibition by causing chloride ion gates to open.[231] Benzodiazepines enhance the inhibitory effects of GABA by allosterically binding to the receptor. Consequently, an increase in the frequency of gated channel opening results in greater chloride channel conductance.[218] The subclass of GABA

receptor to which benzodiazepines bind is $GABA_A$ receptors, which are ligand-gated ion channels. Long term adaptive changes in $GABA_A$ receptors resulting from allosteric binding of benzodiazepines can result in chemical dependence by the patient.[155]

Benzodiazepines have also been discovered to bind to proteins belonging to the bromodomain and extraterminal domain (BET) family of proteins.[14, 69, 196] Bromodomains are protein domains, conserved parts of a given protein sequence and structure, that recognise acetylated lysine residues such as those on the N-terminal tails of histones.[189, 196] Histones are important because of the role they play in organising DNA strands into nucleosomes, a basic unit of DNA packaging in eukaryotic cells, by forming molecular complexes around which the DNA winds. Proteins that contain two bromodomains and an extraterminal domain, whose function is unknown,[224] belong to the BET protein family, which includes BRD2, BRD3, BRD4 and BRDT. BET family proteins perform transcription regulatory function and regulate transcription of several oncogenes (genes which can transform a cell into a tumour cell).[78, 189, 242]

Alprazolam is one of the most toxic benzodiazepines and has a long residence time.[118, 246] In addition to binding to $GABA_A$ receptors and BET proteins, spectroscopic studies predict its binding to a range of important biomolecules including haemoglobin, albumin and even DNA.[49, 172, 240, 243] However, limited structural information exists on its conformation and hydration both in bound and free states. A search of the Protein Data Bank[21] reveals only one structure of alprazolam bound to the first human BRD4 (PDB ID: 3U5J).[68] Alprazolam is also of interest because of its low $\log BB$ value (Section 1.3.) in comparison to other benzodiazepines (0.04) given its $\log P$ value (3.13).[45] For comparison, diazepam has $\log BB$ and $\log P$ values of 0.3 and 2.15, respectively.

Measurements of the solution structure of alpazolam presented in this Chapter are performed in methanol/water solutions, as the predominantly hydrophobic surface of alprazolam means that is sparingly soluble in water ($\sim$0.13 mM). This Chapter relates the conformation and solution structure adopted by alprazolam in methanol/water solutions primarily to its interactions with BET proteins.

5.2 Methods

5.2.1 Neutron Diffraction Sample Preparation

Alprazolam (Fig. 5.1) was purchased from Sigma Aldrich (CAS 28981-97-7) and used without further purification. As alprazolam has no labile protons, unlike indole and cocaine·HCl, the molecule was measured only in its protonated form. Anhydrous methanol and degassed water were prepared as described in Section 3.2.1. Samples for neutron diffraction were prepared by mass under an N_2 atmosphere following procedures described in Section 3.2.1. Six different isotopomers of alprazolam/methanol/water solutions (Table 5.1) were prepared at a relative molar ratio of 1 alprazolam: 125 methanol: 125 water (∼0.14 M).

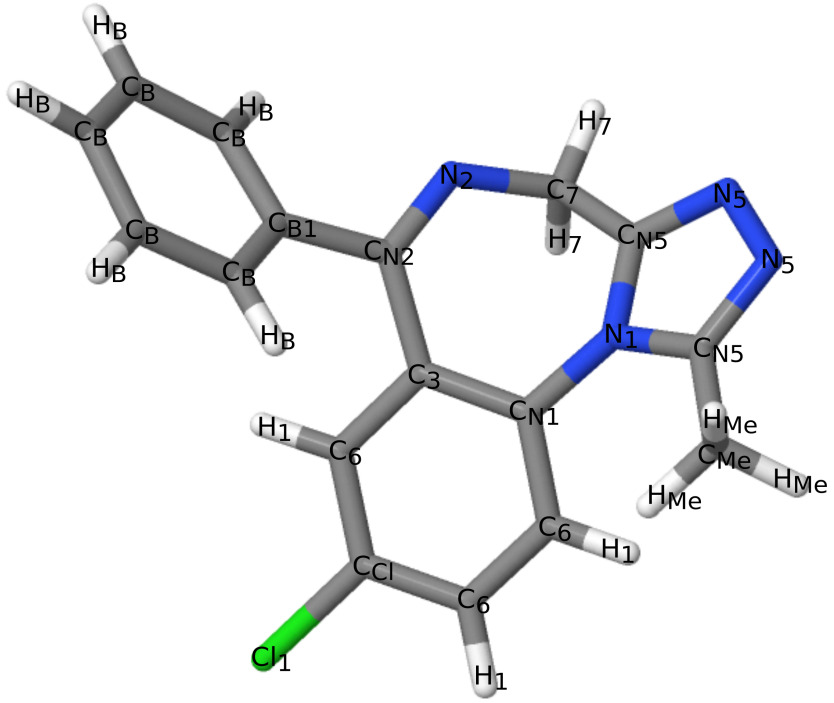


Figure 5.1: The molecular structure and labelling scheme of alprazolam. The rings of alprazolam were also labelled as shown in Fig. 5.2.

5.2.2 Neutron Diffraction Measurements

The samples were measured in Ti/Zr canisters using the SANDALS instrument (Section 2.1.2). The same experimental setup and procedures were used as for indole in Section 3.2.2 save that measurements were performed at 323 K, because alprazolam is not soluble

Sample Name	Alprazolam	Methanol	Water
alprazolam/CH_3OH/H_2O	alprazolam	CH_3OH	H_2O
alprazolam/CHD_3OH/H_2O	alprazolam	$CH_{1.5}D_{1.5}OH$	H_2O
alprazolam/CH_3OD/D_2O	alprazolam	CH_3OD	D_2O
alprazolam/CD_3OD/D_2O	alprazolam	CD_3OD	D_2O
alprazolam/CHD_3OD/D_2O	alprazolam	$CH_{1.5}D_{1.5}OD$	D_2O
alprazolam/CD_3OHD/HDO	alprazolam	$CD_3OH_{0.5}D_{0.5}$	HDO

Table 5.1: The components of the six samples studied. The molar ratio used was 1 alprazolam: 125 methanol: 125 water. Subscripts indicate isotopic proportions.

Sample Name	Measurement Time (μA hrs)	DCS Level
alprazolam/CH_3OH/H_2O	1220	90.2
alprazolam/CHD_3OH/H_2O	1371	89.7
alprazolam/CH_3OD/D_2O	1020	86.8
alprazolam/CD_3OD/D_2O	1030	93.8
alprazolam/CHD_3OD/D_2O	1255	97.7
alprazolam/CD_3OHD/HDO	1280	89.7

Table 5.2: The measurement time and DCS level for each sample. The average current was $\sim$170μA.

at the present concentration at room temperature.

Data were collected for the empty canisters, the empty instrument and a vanadium standard. The μA hrs each sample was measured for is shown in Table 5.2. The empty canisters were each measured for 700 μA hrs. The empty instrument and vanadium standard were measured for 485 μA hrs and 672 μA hrs, respectively.

Diffraction data were corrected as previously described (Section 2.1.3.2). The DCS level of the samples is shown in Table 5.2, where each of the samples has a DCS level within $\sim$14% of their theoretical values, and the mean DCS level of the samples is $\sim$91% of the theoretical value. The vanadium standard and each of the empty Ti/Zr canisters had DCS levels 99%-100% and 94%-100% of their theoretical values, respectively.

5.2.3 Empirical Potential Structure Refinement

The 68.6 Å cubic EPSR simulation box contained 25 alprazolam molecules, 3125 methanol molecules and 3125 water molecules at the measured solution density (ρ = 0.09 atoms

$Å^{-3}$) at 323 K. The SPC/E water model was used.[18] The molecular structure of methanol was generated and energetically optimised using Avogadro,[98] and the same methanol parameters used as in Section 3.2.3. The labelling schemes of methanol and water are shown in Fig. 3.1.

Alprazolam parameters were generated with ANTECHAMBER.[299, 300] Subsequently, collections of atoms were grouped together depending on the similarity of their charges. A full list of the parameters used in EPSR is provided in Table 5.3, with the corresponding labelling scheme in Fig. 5.1. This simulation was accumulated for 6676 steps.

Atom Label	ϵ **(KJ mol^{-1})**	σ**(Å)**	**Atomic Mass (u)†**	$\mathbf{q}_{\mathrm{EPSR}}$**(e)**	$\mathbf{q}_{\mathrm{MD}}$**(e)**
N_5	0.71128	3.2500	14.007	-0.3550	-0.3300
N_5	0.71128	3.2500	14.007	-0.3550	-0.3400
C_{N5}	0.35982	3.9967	12.011	0.3650	0.3500
C_{N5}	0.35982	3.9967	12.011	0.3650	0.3800
N_1	0.71128	3.2500	14.007	-0.2700	-0.2700
C_{Me}	0.45773	0.39967	12.011	-0.1300	-0.1300
H_{Me}	0.06569	2.6495	2.000	0.0650	0.0650
H_{Me}	0.06569	2.6495	2.000	0.0650	0.0650
H_{Me}	0.06569	2.6495	2.000	0.0650	0.0650
C_7	0.45773	3.9967	12.011	0.0850	0.0860
H_7	0.06569	2.4714	2.000	0.0880	0.0890
H_7	0.06569	2.4714	2.000	0.0880	0.0890
N_2	0.71128	3.2500	14.007	-0.5700	-0.5700
C_{N2}	0.35982	3.9967	12.011	0.5300	0.5300
C_3	0.35982	3.9967	12.011	-0.1800	-0.1800
C_{N1}	0.35982	3.9967	12.011	0.0310	0.0330
C_6	0.35982	3.9967	12.011	-0.1070	-0.0820
C_6	0.35982	3.9967	12.011	-0.1070	-0.0980
C_6	0.35982	3.9967	12.011	-0.1070	-0.1400
H_1	0.06276	2.5996	2.000	0.1600	0.1600

H_1	0.06276	2.5996	2.000	0.1600	0.1700
H_1	0.06276	2.5996	2.000	0.1600	0.1500
C_{Cl}	0.35982	3.9967	12.011	0.0060	0.0065
Cl_1	1.10880	3.4709	35.453	-0.0740	-0.0720
C_{B1}	3.5982	3.9967	12.011	-0.1500	-0.1500
C_B	0.35982	3.9967	12.001	-0.1136	-0.0990
C_B	0.35982	3.9967	12.001	-0.1136	-0.1300
C_B	0.35982	3.9967	12.001	-0.1136	-0.1100
C_B	0.35982	3.9967	12.001	-0.1136	-0.1300
C_B	0.35982	3.9967	12.001	-0.1136	-0.0990
H_B	0.06276	2.5996	2.000	0.1400	0.1400
H_B	0.06276	2.5996	2.000	0.1400	0.1400
H_B	0.06276	2.5996	2.000	0.1400	0.1400
H_B	0.06276	2.5996	2.000	0.1400	0.1400
H_B	0.06276	2.5996	2.000	0.1400	0.1400
C_m	0.27614	3.5000	12.011	0.1450	0.1450
H_m	0.12552	2.5000	2.000	0.0400	0.0400
O_m	0.71128	3.1200	16.000	-0.6830	-0.6830
H_{mo}	0.00000	0.0000	2.000	0.4180	0.4180
O_w	0.65000	3.1660	16.000	-0.8476	-0.8476
H_w	0.00000	0.0000	2.000	0.4238	0.4238

Table 5.3: The parameters used in the EPSR and MD simulations of alprazolam/methanol/water. † The atomic mass of hydrogen used in MD was 1.000 u. The corresponding labelling scheme for alprazolam is shown in Fig. 5.1.

5.2.4 Molecular Dynamics

MD simulations were run[a] using Gromacs 5.0.2.[272], with the number of each molecular species the same as was used in EPSR. Alprazolam and methanol parameters (Table 5.3) were generated using AMBER and converted to Gromacs format using the amb2gmx.pl script.[186] The SPC/E water model was used.[18, 165] All of the bonds and angles for the water molecules were constrained using the SHAKE algorithm.[239] Initially, energy minimisation corrected atomic overlaps until the maximum force on any atoms was below $100\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$. Energy minimisation was followed by solute restrained equilibration runs of 1 ns each in the NVT and NPT ensembles. The Berendsen thermostat and barostat were used to maintain temperature and pressure, respectively, in the equilibration runs.[18] One simulation was performed at 323 K, the temperature at which neutron diffraction measurements were made, and a second at 298 K, room temperature, while the pressure was 1 bar in both runs. In this way, the behaviour of the molecules over a temperature range (298 K-323 K) encompassing body temperature was assessed. The production runs in both simulations were carried out in the NPT ensemble for 40 ns with a timestep of 2 fs. Temperature was maintained using the V-rescale thermostat, while pressure was controlled using the Parrinello-Rahman barostat.[206] A cut-off of 14 Å was used for the van der Waals interactions, and long range interactions were calculated using the Particle Mesh Ewald algorithm.[50, 166] Coordinates were saved every 0.8 ps, resulting in 50000 frames.

5.2.5 ANGULA

The orthogonal coordinate systems assigned to alprazolam in ANGULA are shown in Fig. 5.2. Coordinate systems were assigned to each of the 4 rings of alprazolam. The coordinate systems assigned to methanol and water were as described in Section 3.2.5. WMA was performed for a distance range of 0-4 Å. Analysis using ANGULA was performed for ∼3000 snapshots of the EPSR simulation box and ∼5000 frames of the MD trajectory.

[a]Both MD simulations of alprazolam in methanol/water solutions were performed by Akshay Sridhar.

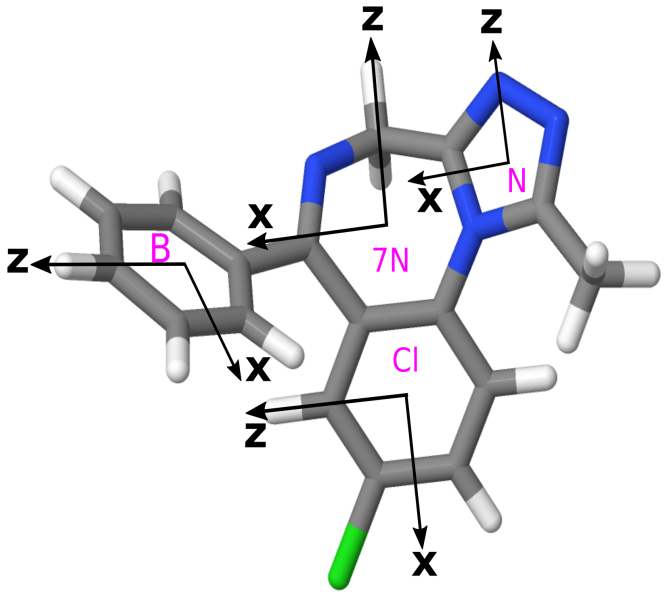


Figure 5.2: The coordinate systems assigned to alprazolam using ANGULA. The **y** axes are coming out of the plane of the page. Labels used to identify the rings of alprazolam to which the coordinate systems were assigned (pink) are also shown. RDFs were plotted which also used this labelling scheme to investigate the interactions between the rings of alprazolam and solvent atoms.

5.3 Results

5.3.1 Diffraction Data

Measured neutron diffraction data for alprazolam dissolved in methanol/water solutions at 323 K and EPSR fits to these data are shown in reciprocal space and real space in Fig. 5.3 (a) and Fig. 5.3 (b), respectively. The EPSR fits to the neutron diffraction data in reciprocal space are generally good, as indicated by the flat difference curves for Q >5 Å^{-1}. There are some small differences between the experimental and simulated data at low Q values ($Q < 5$ Å^{-1}) due to difficulty correcting for inelastic scattering in this region. The EPSR fits to the neutron diffraction data in real space show a similar level of agreement between the experimental and simulated data as in reciprocal space.

The same neutron diffraction data shown in Fig. 5.3 are shown in Fig. 5.4 along with the $F(Q)$s and $G(r)$s generated from the MD trajectory of the MD simulation run at 323 K. The difference curves in Fig. 5.4 (a) are not as flat as those in Fig. 5.3 (a), the MD $F(Q)$s matching the experimental data less closely than the EPSR fits. Similarly, the agreement between the experimental and the simulated data in Fig. 5.4

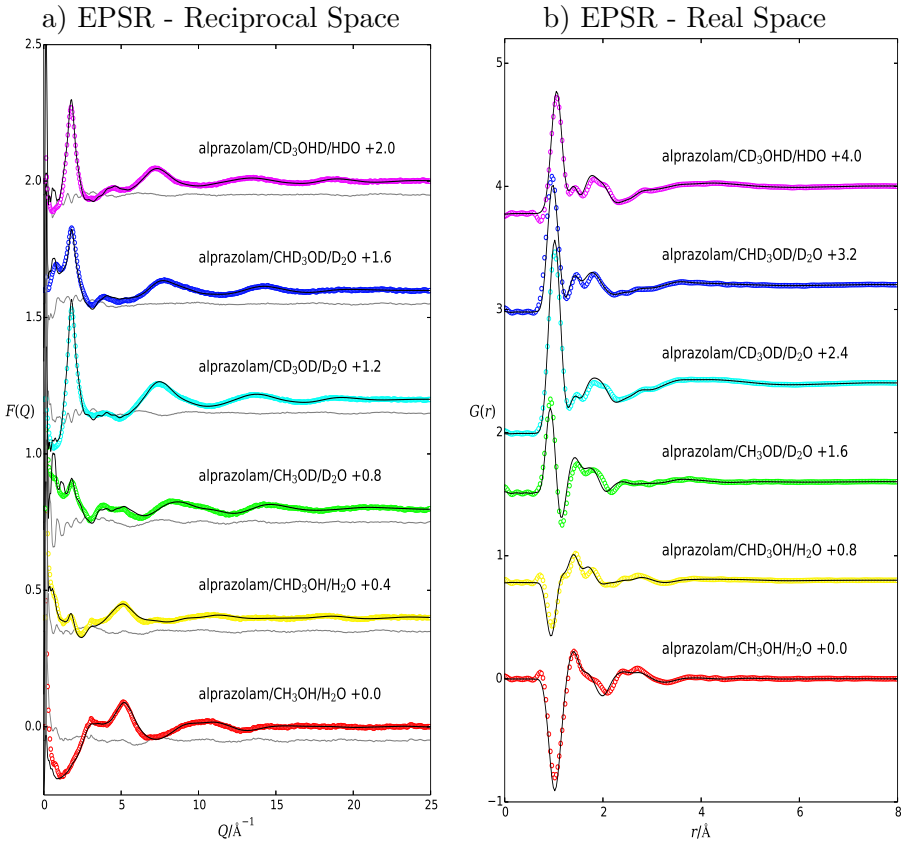


Figure 5.3: Measured neutron diffraction data (coloured lines) and EPSR fits to the data (black lines) in reciprocal space (a) and real space (b). The difference curves (grey lines) have been shifted by 0.05 for clarity.

(b) is worse than the agreement in Fig. 5.3 (b). An example of this worse agreement is seen for the first trough in Fig. 5.4 (b) for sample alprazolam/CH_3OH/H_2O. Samples alprazolam/CD_3OD/D_2O, alprazolam/CHD_3OD/D_2O and alprazolam/CD_3OHD/HDO each show noticeably worse agreement between experimental and simulated data for their first peaks in Fig. 5.4 (b) than is seen for the same peaks in the EPSR fits in Fig. 5.3 (b). As inspection of the simulation box revealed no aggregation of alprazolam molecules, the worse agreement of the MD $F(Q)$s with the experimental data compared with the EPSR fits may be attributed to the larger distribution of intramolecular distances afforded by EPSR, which is more consistent with the measured diffraction data.

5.3.2 Alprazolam Conformation

As previously discussed in relation to cocaine (Section 4.3), understanding the conformation of molecules in solution is important in determining the nature of interactions

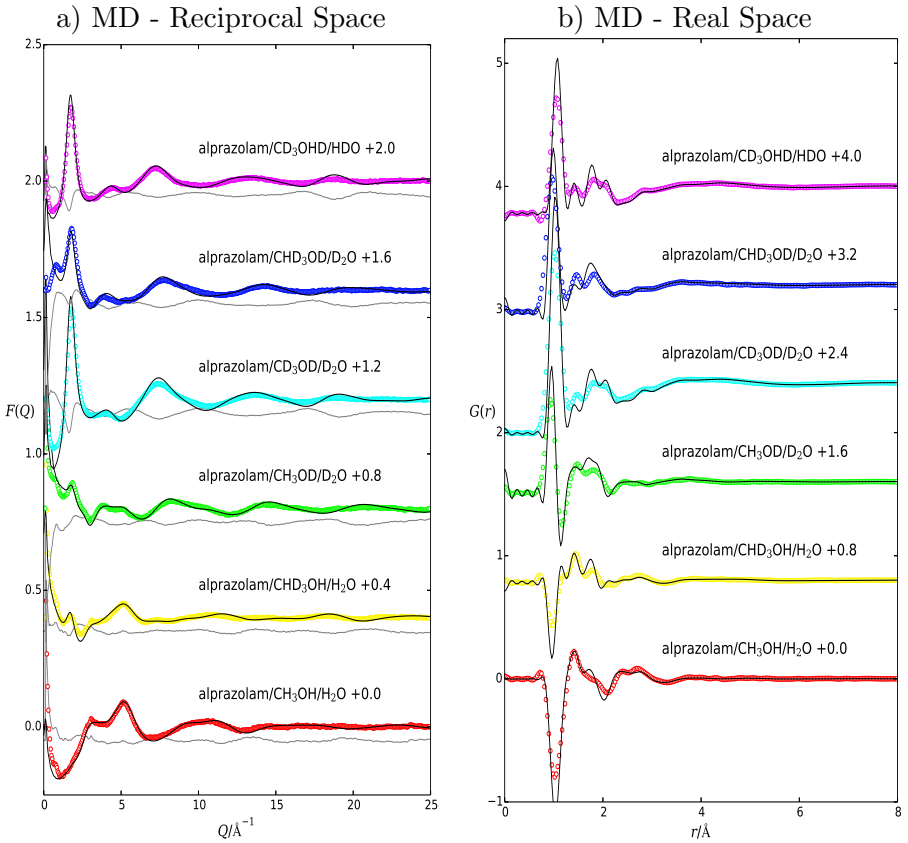


Figure 5.4: Measured neutron diffraction data (coloured lines) and the $F(Q)$s and $G(r)$s generated from the MD simulations run at 323 K (black lines) in reciprocal space (a) and real space (b), respectively. The difference curves (grey lines) have been shifted by 0.05 for clarity.

between a molecule and hydrophobic/hydrophilic groups in a system, such as in the amphiphilic environment of the BBB or with proteins. The fused chlorobenzene, diazepine and triazole rings of alprazolam leave the benzene ring as the principal constituent ring capable of dictating the conformations adopted by the molecule. Therefore, the distribution of the torsional angles corresponding to the rotation of the benzene ring about the C_{N2}-C_{B1} bond was analysed for EPSR and MD (Fig. 5.5 (a)). (Herein, unless stated otherwise, MD refers to the MD simulation run at 323 K.)

The torsional angle distribution for the rotation of the benzene ring of alprazolam about the C_{N2}-C_{B1} bond plotted for the EPSR simulation box in Fig. 5.5 (a) shows peaks at ±40°, with an approximately equal proportion of molecules whose populations are centred on ϕ=40° or ϕ=-40°. This finding indicates that the benzene ring is slightly tilted with respect the molecular plane formed by the chlorobenzene, diazepine and triazole rings. The range of conformations adopted by alprazolam in EPSR is further

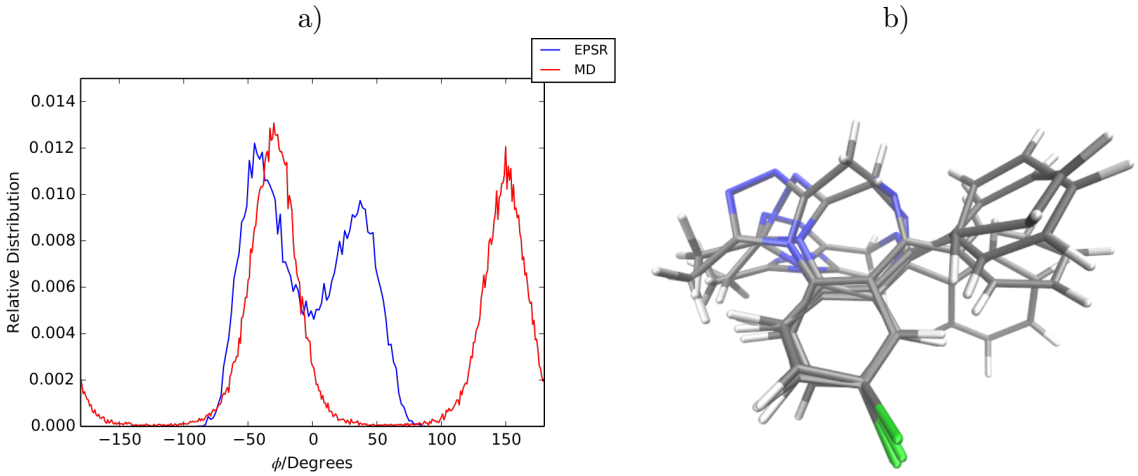


Figure 5.5: The torsional angle distribution for the benzene ring (a) extracted from EPSR (blue line) and MD (red line). An overlay of 4 alprazolam molecules taken from the EPSR simulation box (b), which underlines the mobility of the benzene ring.

illustrated by Fig. 5.5 (b), which shows a number of alprazolam molecules from the EPSR simulation box each centred on the origin of the coordinate system attached to the chlorobenzene ring, where the benzene ring is the most mobile ring of the molecule. The torsional angle distribution plotted for the MD simulation box in Fig. 5.5 (a) shows a peak at -30° in addition to a peak at 150°. The proportion of alprazolam molecules whose populations are centred on either ϕ=-30° or ϕ=150° is roughly equivalent, as it is for the two populations in EPSR. In MD the benzene ring moves through a wider angular range than in EPSR. However, for both EPSR and MD the benzene ring is, on average, tilted <45° from the plane of the rest of the molecule.

Torsional angle distributions with a peak at ∼30° are consistent with crystallographic studies of the structure of alprazolam bound to the BET family protein BRD4, which indicate that the benzene ring is tilted away from the plane of the molecule at an angle of ∼28°.[68] Modelling studies, however, predict the benzene ring to be normal to the plane of the rest of the molecule.[49]

5.3.3 Solvent Structure

Fig. 5.6 shows the water-water and methanol-methanol RDFs extracted from the EPSR and MD simulations, with corresponding CNs being shown in Table 5.4. CNs in this Section are computed at the peak minima (right hand column of CN tables) of the

corresponding RDFs. RDFs extracted from EPSR simulations of pure water[266] and pure methanol[56] are included in Fig. 5.6 for comparison.

The positions of the peaks in the water-water RDFs extracted from both EPSR and MD in Fig. 5.6 match those shown for indole/methanol/water solutions in Section 3.3, with the first peaks in $g_{O_wO_w}(r)$, $g_{O_wH_w}(r)$ and $g_{H_wH_w}(r)$ located at ∼2.8 Å, ∼1.8 Å and ∼2.4 Å, respectively, in both EPSR and MD. Therefore, the peaks of the water-water RDFs in Fig. 5.6 also match the peaks of the RDFs extracted from the EPSR simulations of the pure water. As discussed in detail in Section 3.3, these findings indicate that there is no enhancement of the water structure in methanol/water solutions compared with pure water.[268] Similarly, the positions of the peaks in the methanol-methanol RDFs agree with the peak positions previously reported for pure methanol and methanol/water solutions.[6, 56] The first peaks in both EPSR and MD are located at ∼2.8 Å, ∼3.9 Å and ∼2.5 Å in $g_{O_mO_m}(r)$, $g_{C_mC_m}(r)$ and $g_{H_mH_m}(r)$, respectively. Similar CNs for both simulations are seen in Table 5.4. As in Section 3.3, CNs in Table 5.4 point to a slight preference for water-water interactions in MD compared with EPSR.

n_α^β	**EPSR**	**MD**	***r*/Å**
$n_{O_w}^{O_w}$	2.44	2.63	3.48
$n_{O_w}^{H_w}$	1.03	1.14	2.49
$n_{H_w}^{H_w}$	3.00	3.21	3.15
$n_{H_m}^{H_m}$	2.03	1.72	3.09
$n_{O_m}^{O_m}$	1.02	0.97	3.42
$n_{C_m}^{C_m}$	7.52	7.58	5.70

Table 5.4: CNs of the water-water and methanol-methanol RDFs shown in Fig. 5.6.

Fig. 5.7 shows the solvent-solvent hydrogen bonding interactions between methanol and water in the present solution. As for indole/methanol/water solutions in Section 3.3, the position of the first peak in each of the RDFs corresponding to solvent-solvent hydrogen bonding interactions in Fig. 5.7 indicates that H_w and H_{mo} form hydrogen bonds with O_w and O_m at ∼1.8 Å in both EPSR and MD, with a similar propensity to bind to each other as with themselves underlined by Table 5.5.

Summation of the CNs shown in Table 5.5 indicates that the system is slightly undersaturated with respect to H_w hydrogen bonding in both EPSR and MD: $n_{H_w}^{O_w}+n_{H_w}^{O_m}$ =

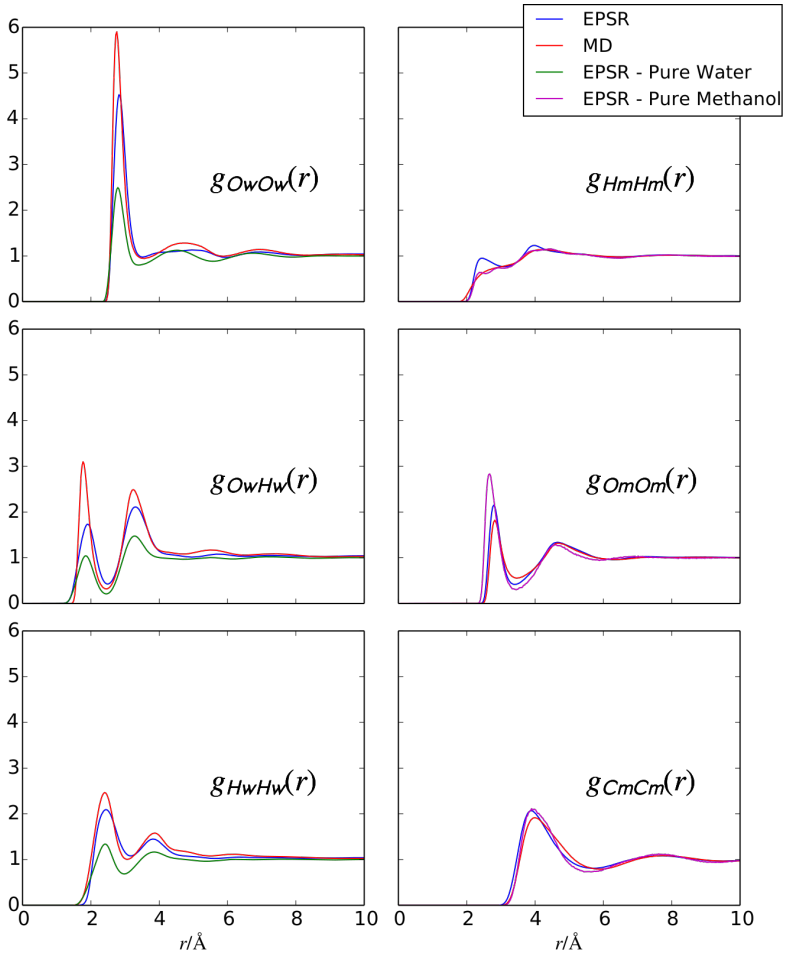


Figure 5.6: RDFs for water-water interactions extracted from EPSR (blue line) and MD (red line) simulations of alprazolam/methanol/water. RDFs extracted from EPSR simulations of pure water[266] (green line) and pure methanol[56] (magenta line) are also included.

$0.51 + 0.39 = 0.90$ for EPSR and $n_{H_w}^{O_w} + n_{H_w}^{O_m} = 0.57 + 0.35 = 0.92$ for MD. Similarly, in the case of H_{mo} hydrogen bonding $n_{H_{mo}}^{O_w} + n_{H_{mo}}^{O_m} = 0.46 + 0.47 = 0.93$ and $n_{H_{mo}}^{O_w} + n_{H_{mo}}^{O_m} = 0.50 + 0.42 = 0.92$ for EPSR and MD, respectively. Differences in the hydrogen bonding potential when comparing both solvent and simulation technique are limited to within 3.5%.

n_α^β	**EPSR**	**MD**	r**/Å**
$n_{H_w}^{O_w}$	0.51	0.57	2.49
$n_{H_w}^{O_m}$	0.39	0.35	2.52
$n_{H_{mo}}^{O_w}$	0.46	0.50	2.55
$n_{H_{mo}}^{O_m}$	0.47	0.42	2.58

Table 5.5: CNs of the hydrogen bonding RDFs shown in Fig. 5.7.

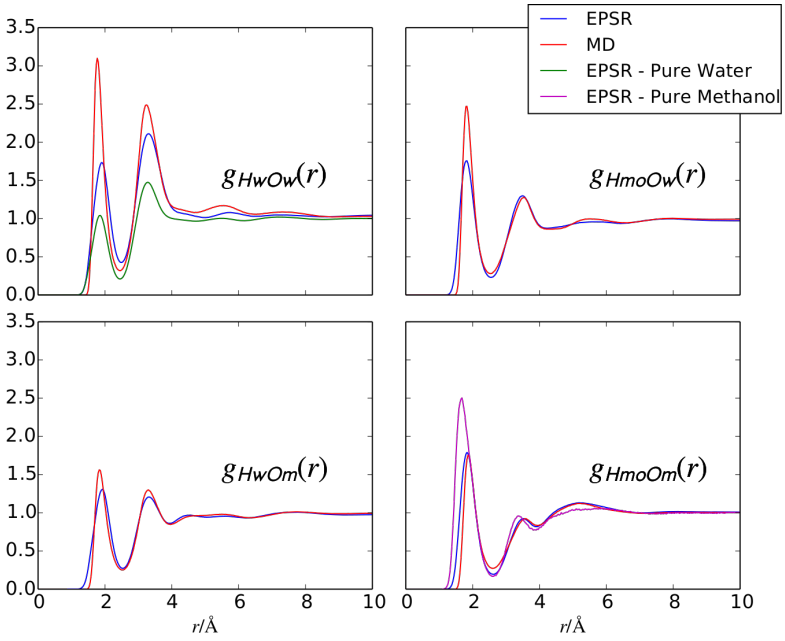


Figure 5.7: RDFs for hydrogen bonding of water (O_w and H_w) and methanol (O_m and H_{mo}) atoms extracted from EPSR (black line) and MD (red line). Corresponding RDFs extracted from EPSR simulations of pure water[266] (green line) and pure methanol[56] (magenta line) are also included.

5.3.4 Alprazolam Solvation

5.3.4.1 Radial Distribution Functions

Perhaps the most natural way to frame discussion of the solvation of alprazolam is in terms of understanding the solvation of its four constituent structures (Fig. 5.2): benzene ring (B), chlorobenzene ring (Cl), triazole ring (N) and diazepine ring (7N).

Fig. 5.8 shows RDFs for the interactions of the benzene ring of alprazolam with water and methanol solvent atoms. The form of the RDFs is broadly consistent between EPSR and MD for each benzene-water interaction ($g_{BO_w}(r)$ and $g_{BH_w}(r)$). These RDFs show an absence of any clear peaks, indicating that there are few highly localised interactions between the benzene ring of alprazolam and water in either EPSR or MD. These findings are in contrast to the RDFs for the interactions between the benzene ring of indole and water atoms in Section 3.3, where water molecules were found to hydrogen bond with the benzene ring of indole. Benzene-water RDFs in Fig. 5.8 also indicate fewer highly localised interactions than were seen between the benzene ring of cocaine and water (Section 4.3). Lack of highly localised interactions between the benzene ring of

alprazolam and water may be due to both an increase in the number of atoms for which interaction with water molecules is favourable in comparison to indole and a decrease in the concentration of alprazolam dissolved in solution (1:250) compared with cocaine (1:65).

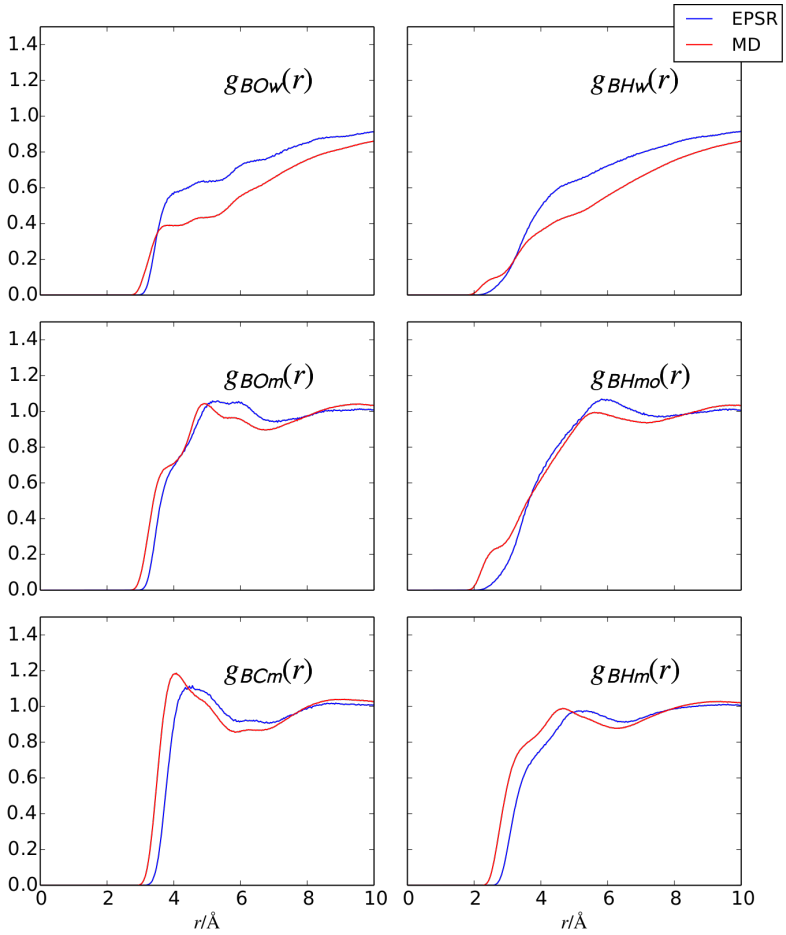


Figure 5.8: RDFs for the interactions of the benzene ring of alprazolam (B - the weighted average of C_B and C_{B1}) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR (blue line) and MD (red line).

n_α^β	**EPSR**	**MD**	***r*/Å**
$n_B^{O_w}$	2.07	1.53	5.00
$n_B^{H_w}$	4.01	3.11	5.00
$n_B^{O_m}$	11.07	11.30	7.00
$n_B^{H_{mo}}$	14.78	14.29	7.47
$n_B^{C_m}$	6.54	6.91	5.94
$n_B^{H_m}$	26.91	28.06	6.51

Table 5.6: CNs of the alprazolam/methanol/water RDFs shown in Fig. 5.8.

Benzene-methanol RDFs in Fig. 5.8 are also similar for the two simulation techniques

employed. However, unlike benzene-water interactions, interactions between the benzene ring and methanol atoms indicate some degree of preferred location of methanol relative to the benzene ring of alprazolam. For instance, $g_{\mathrm{BC_m}}(r)$ has peaks at ∼4.6 Å and ∼4.1 Å in EPSR and MD, respectively. These distances indicate that the -CH_3 group of methanol has a preferred location relative to the benzene ring of alprazolam in both EPSR and MD and that this preferred location is closer to the benzene ring in MD than in EPSR. That the methyl group has a preferred location relative to the benzene ring which is at a shorter distance in MD compared with EPSR is reflected in $g_{\mathrm{BH_m}}(r)$, where peaks are at ∼5.2 Å and ∼4.7 Å in EPSR and MD, respectively. $g_{\mathrm{BO_m}}(r)$ and $g_{\mathrm{BH_{mo}}}(r)$ show peaks, although broader and less well defined than the peaks in $g_{\mathrm{BC_m}}(r)$, at ∼5.6 Å and ∼5.8 Å, respectively, at the same location for EPSR and MD. These peak positions indicate that the hydroxyl group of methanol has the same preferred distance from the benzene ring in EPSR and MD. In contrast to the results of indole dissolved in methanol/water solutions (Section 3.3), -CH_3 groups are closer to the benzene ring than -OH groups in both EPSR and MD.

CNs corresponding to the RDFs displayed in Fig. 5.8 (Table 5.6) indicate that, although there is no clear preferred distance of water relative to the benzene ring in either EPSR or MD, there are a greater number of water molecules at a given distance from the benzene ring in EPSR compared with MD. For example, the number of water hydrogens at a distance of 5 Å from the benzene ring of alprazolam is ∼30% greater in EPSR than in MD. For methanol differences in the CNs associated with solvation of the benzene ring are less pronounced compared with the hydration of the benzene ring. In fact, any differences in the CNs computed from EPSR and MD are within 10%. However, there is a slight trend for the benzene-methanol CNs to be larger in MD compared with EPSR. Benzene-methanol interactions being more frequent in MD compared with EPSR is consistent with the shorter distance at which the first peak is found in $g_{\mathrm{BC_m}}(r)$ for MD compared with EPSR in Fig. 5.8. The only CN which is greater in EPSR than MD is $n_B^{H_{mo}}$.

RDFs for the interactions between the chlorobenzene ring of alprazolam and solvent molecules are shown in Fig. 5.9. Like Fig. 5.8, Fig. 5.9 shows similarities between the RDFs extracted from EPSR and MD. Further like Fig. 5.8, the chlorobenzene-water

RDFs point to an absence of localised interactions between the chlorobenzene ring and water molecules. In Fig. 5.9, there is also an absence of highly ordered methanol molecules about the chlorobenzene ring. The reduction of chlorobenzene-methanol interactions compared with benzene-methanol interactions is most obvious through comparison of the RDFs for interactions between the respective rings and C_m ($g_{BC_m}(r)$ and $g_{ClC_m}(r)$). However, $g_{ClC_m}(r)$ and $g_{ClH_m}(r)$ suggest that chlorobenzene-methanol interactions occur over a shorter distance range in MD than in EPSR, as was the case for benzene-methanol interactions in Fig. 5.8.

CNs in Table 5.7, which correspond to the RDFs in Fig. 5.9, largely follow the trend seen in Table 5.6. Explicitly, there appears to be a greater number of water molecules around the chlorobenzene ring and fewer methanol -CH_3 groups around the chlorobenzene ring in EPSR than in MD. Once again, there is evidence that the hydroxyl group of methanol has a slightly greater propensity to interact with alprazolam in EPSR compared with MD, as the CNs corresponding to chlorobenzene-methanol hydroxyl interactions are fractionally greater for EPSR than for MD.

n_α^β	**EPSR**	**MD**	***r*/Å**
$n_{Cl}^{O_w}$	4.06	2.64	6.00
$n_{Cl}^{H_w}$	8.27	5.47	6.00
$n_{Cl}^{O_m}$	5.94	5.92	6.00
$n_{Cl}^{H_{mo}}$	5.80	5.69	6.00
$n_{Cl}^{C_m}$	4.25	4.65	5.60
$n_{Cl}^{H_m}$	17.14	19.55	6.00

Table 5.7: CNs of the alprazolam/methanol/water RDFs shown in Fig. 5.9.

RDFs corresponding to triazole ring solvation (Fig. 5.10) show more pronounced peaks compared with the functions seen in Fig. 5.8 and Fig. 5.9 for the benzene ring and the chlorobenzene ring, respectively. This contrast is particularly marked for triazole-water interactions, where peaks are seen at ∼3.0 Å and ∼2.1 Å, in $g_{NO_w}(r)$ and $g_{NH_w}(r)$, respectively, for both EPSR and MD. Similarly, $g_{NO_m}(r)$ and $g_{NH_{mo}}(r)$ have peaks at ∼3.0 Å and ∼2.1 Å, respectively, for both EPSR and MD. No clear peaks are visible in either $g_{NC_m}(r)$ and $g_{NH_m}(r)$. These findings illustrate that water molecules and the hydroxyl group of methanol have a similar preferred distance from the triazole ring. The position and sharpness the peaks in $g_{NH_w}(r)$ and $g_{NH_{mo}}(r)$ indicate that water

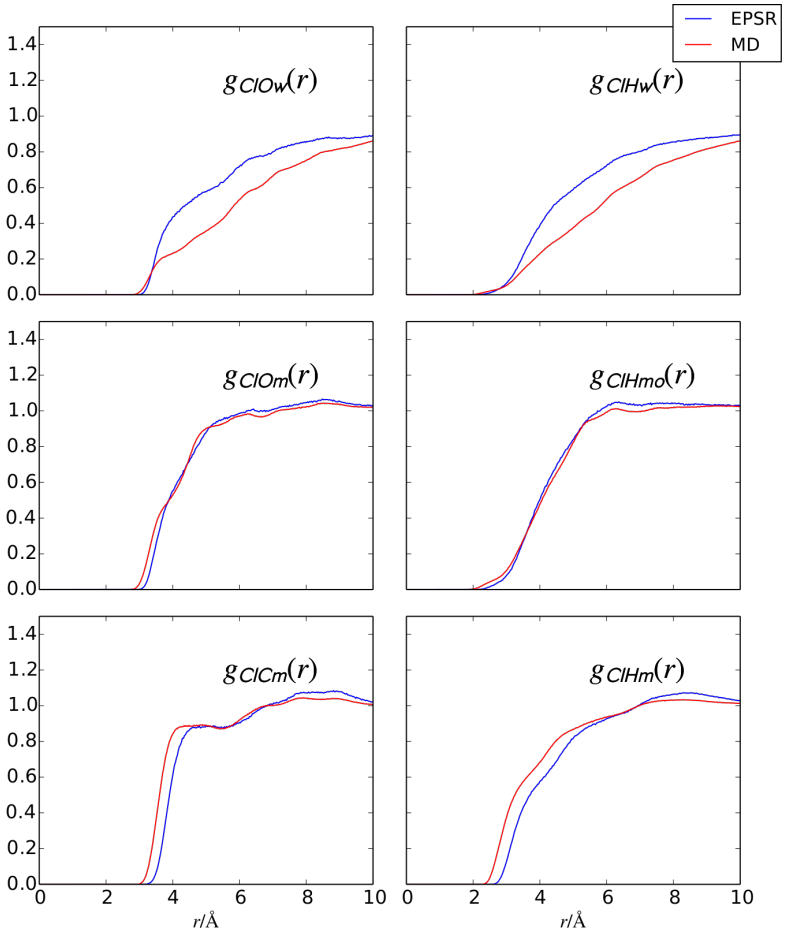


Figure 5.9: RDFs for the interactions of the chlorobenzene ring of alprazolam (Cl - the weighted average of C_{Cl}, C_6, C_3 and C_{N1}) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR (blue line) and MD (red line).

and methanol are forming moderately strong, mostly electrostatic hydrogen bonds with the triazole ring of alprazolam.[274] Differences in the solvation of the triazole ring of alprazolam by water and methanol atoms can be seen in the RDFs, where there is a second peak, corresponding to a second solvation shell, in $g_{NO_m}(r)$ and $g_{NH_{mo}}(r)$ at $\sim$5.0 Å and $\sim$5.7 Å, respectively, for both EPSR and MD that is not clearly seen in $g_{NO_w}(r)$ and $g_{NH_w}(r)$.

CNs in Table 5.8 corresponding to triazole-solvent interaction show that, on average, there are a similar number (within 15%) of water molecules that form hydrogen bonds with the triazole ring in EPSR and MD. CNs for methanol are also very similar across the two simulations. Triazole-water and triazole-methanol interactions considered together, however, hint at a slight preference for hydrogen bonding of the triazole ring with water molecules and the hydroxyl group of methanol in MD over EPSR - the opposite of what

was seen for the benzene ring (Table 5.6) and the chlorobenzene ring (Table 5.7). In contrast, Table 5.8 shows that there are more methanol -CH_3 groups, on average, around the triazole ring in EPSR compared with MD, although these motifs are not as localised relative to the triazole ring (or its constituents) as -OH groups.

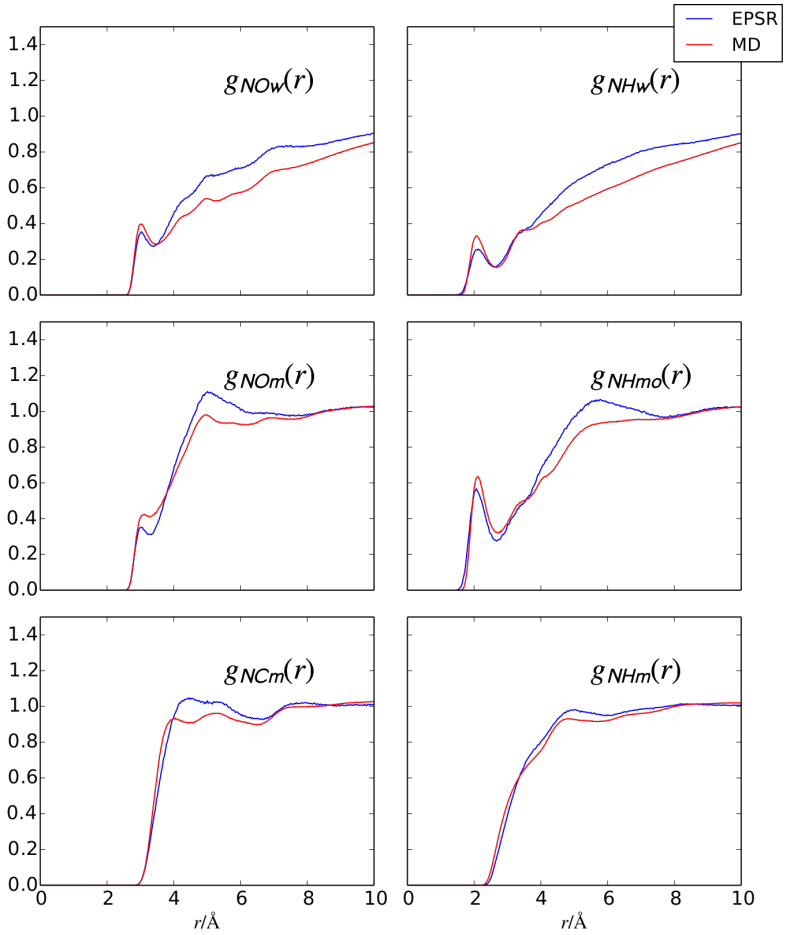


Figure 5.10: RDFs for the interactions of the triazole ring of alprazolam (N - the weighted average of C_{N5}, N_1 and N_5) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR (blue line) and MD (red line).

n_α^β	**EPSR**	**MD**	***r*/Å**
$n_N^{O_w}$	0.23	0.26	3.42
$n_N^{H_w}$	0.21	0.24	2.61
$n_N^{O_m}$	0.19	0.25	3.30
$n_N^{H_{mo}}$	0.23	0.26	2.70
$n_N^{C_m}$	9.69	9.30	6.57
$n_N^{H_m}$	21.02	20.41	6.00

Table 5.8: CNs of the alprazolam/methanol/water RDFs shown in Fig. 5.10.

Peaks in the RDFs in Fig. 5.10 can be largely attributed to interactions between

solvent molecules and the nitrogen atoms that constitute the triazole ring (N_5 and N_1). To illustrate this point, RDFs for the interactions between these nitrogen atoms and solvent atoms are shown in Fig. 5.11, where corresponding RDFs for the remaining nitrogen atom of alprazolam (N_2) are included for comparison. Well defined peaks in $g_{N_2H_w}(r)$, $g_{N_2H_{mo}}(r)$, $g_{N_5H_w}(r)$ and $g_{N_5H_{mo}}(r)$ indicate that H_w and H_{mo} bind to N_5 and N_2 at $\sim$2.1 Å in both EPSR and MD. Interactions between N_1 and solvent molecules appear to occur over a longer distance range, as evidenced by the peaks in $g_{N_1H_w}(r)$ and $g_{N_1H_{mo}}(r)$ located at $\sim$4.1 Å for both EPSR and MD. Peaks in $g_{N_1H_w}(r)$ and $g_{N_1H_{mo}}(r)$ being located at greater distances and being broader and less well defined than the peaks in $g_{N_2H_w}(r)$, $g_{N_2H_{mo}}(r)$, $g_{N_5H_w}(r)$ and $g_{N_5H_{mo}}(r)$ for both EPSR and MD perhaps indicates that they are a reflection of the propensity of nearby atoms (i.e. N_5) to interact with water and methanol rather than being the result of interactions between N_1 and the solvent. The RDFs in Fig. 5.11 suggest that the peaks in the RDFs in Fig. 5.10 are principally related to interactions between the N_5 and solvent molecules.

CNs corresponding to the RDFs in Fig. 5.11 are shown in Table 5.9. The CNs in this table highlight the greater number, on average, of H_w and H_{mo} atoms around both N_2 and N_5 in MD compared with EPSR. There is no clear relationship between either simulation technique or solvent molecule with regards to interactions of N_1 with water and methanol.

n_α^β	**EPSR**	**MD**	r/Å
$n_{N_2}^{H_w}$	0.23	0.24	2.76
$n_{N_2}^{H_{mo}}$	0.14	0.22	2.76
$n_{N_5}^{H_w}$	0.56	0.64	2.76
$n_{N_5}^{H_{mo}}$	0.72	0.77	3.11
$n_{N_1}^{H_w}$	1.26	1.26	4.38
$n_{N_1}^{H_{mo}}$	1.28	1.26	4.41

Table 5.9: CNs of the alprazolam/methanol/water RDFs shown in Fig. 5.11.

Solvation of the diazepine ring of alprazolam is illustrated by the RDFs shown in Fig. 5.12. The solvation of the diazepine ring largely resembles the solvation of the chlorobenzene ring, as there a few well defined peaks clearly visible in the RDFs shown in Fig. 5.12. Like each of the previous RDFs showing the solvation of alprazolam, there

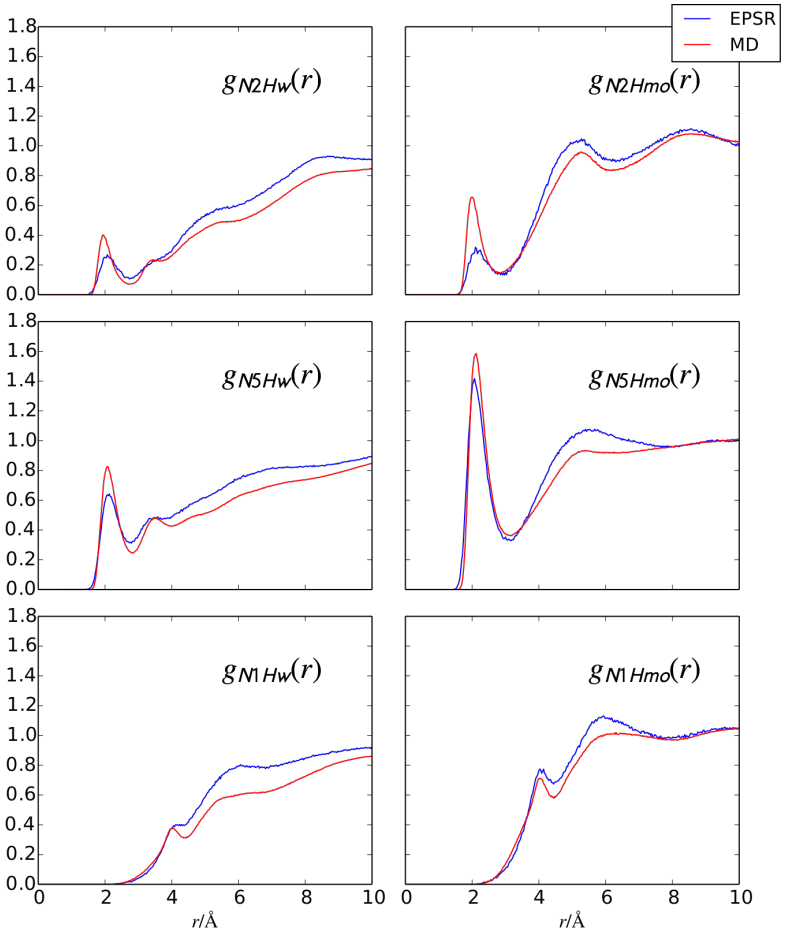


Figure 5.11: RDFs for the interactions of N_2, N_5 and N_1 with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR (blue line) and MD (red line).

is good agreement between the RDFs extracted from EPSR and MD. Small peaks are visible in $g_{7\mathrm{NH_w}}(r)$ and $g_{7\mathrm{NH_{mo}}}(r)$ at ∼2.0 Å that are consistent with hydrogen bonding of the ring with water and methanol oxygens. However, these peaks are significantly less pronounced than those seen for the triazole ring in $g_{\mathrm{NH_w}}(r)$ and $g_{\mathrm{NH_{mo}}}(r)$. The RDFs shown in Fig. 5.11 suggest that diazepine-solvent interactions occur by means of hydrogen bonding between N_2 and solvent molecules, explaining the peaks at ∼2.0 Å in EPSR and MD. A combination of the reduced number of interactions of N_2 with solvent molecules compared with N_5 (Table 5.9) and the averaging of the RDFs in Fig. 5.12 over 7 distinct interactions compared with 5 distinct interactions in Fig. 5.10 means that such peaks are smaller.

CNs for interactions between the diazepine ring of alprazolam and solvent molecules shown in Table 5.10 broadly return to the trends seen for the benzene ring and chloroben-

zene ring, namely that water molecules and the -OH group of methanol have larger CNs in EPSR than MD, and the -CH_3 group of methanol has a larger CN in MD than EPSR. The CNs computed for the small peaks seen in $g_{7\mathrm{NH_w}}(r)$ and $g_{7\mathrm{NH_{mo}}}(r)$ at ∼2.0 Å do not follow this trend but are perhaps too small for an assessment to be made.

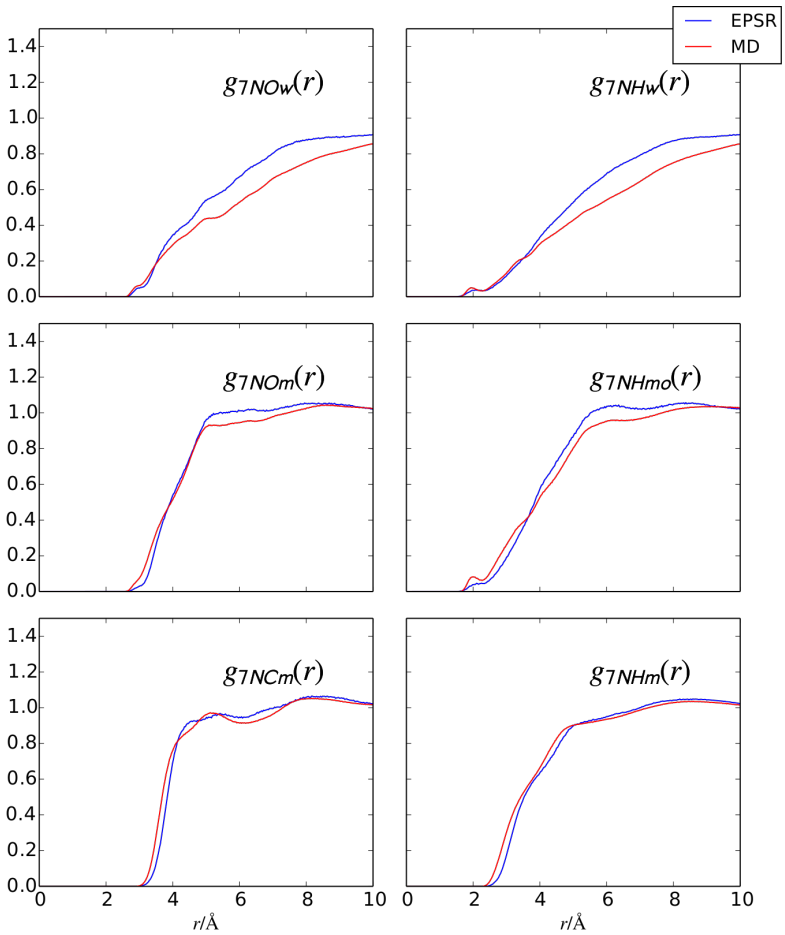


Figure 5.12: RDFs for the interactions of the diazepine ring of alprazolam (7N - the weighted average of C_3, C_{N1}, C_{N2}, C_{N5}, C_7, N_1 and N_2) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR (blue line) and MD (red line).

n_{α}^{β}	**EPSR**	**MD**	***r*/Å**
$n_{7N}^{O_w}$	1.45	1.17	5.00
$n_{7N}^{H_w}$	0.02	0.03	2.34
$n_{7N}^{O_m}$	6.09	5.85	6.00
$n_{7N}^{H_{mo}}$	0.01	0.03	2.34
$n_{7N}^{C_m}$	6.18	6.25	6.00
$n_{7N}^{H_m}$	18.55	19.05	6.00

Table 5.10: CNs of the alprazolam/methanol/water RDFs shown in Fig. 5.12.

5.3.4.2 Spatial Density Maps

SDMs generated from EPSR showing the distribution of nearest neighbour water and methanol oxygens relative to each of the constituent rings of alprazolam are shown in Fig. 5.13. The SDMs showing the distribution of nearest neighbour solvent molecules relative to the benzene ring of alprazolam show the same maximum density of nearest neighbour water and methanol oxygens ($0.26\,\text{Å}^{-3}$). In the cases of both water and methanol, the density that is visible around the benzene ring is found above and below ($\pm\mathbf{y}$ direction) the plane of the benzene ring, as was the case for indole (Section 3.3.3.3) and cocaine (Section 4.3.5.2) in solution.

Like the benzene ring, SDMs corresponding to the distribution of nearest neighbour solvent molecules relative to the chlorobenzene ring of alprazolam show a low density of nearest neighbour water and methanol molecules, with a maximum nearest neighbour solvent oxygen density of $0.28\,\text{Å}^{-3}$ for both water and methanol. The similar low density of methanol and water molecules around the chlorobenzene is consistent with the RDFs in Fig. 5.9, where there is an absence of localised solvent molecules around the chlorobenzene ring.

Unsurprisingly given the RDFs in Fig. 5.10, SDMs generated from EPSR that show the distribution of nearest neighbour solvent molecules relative to the triazole ring of alprazolam display regions of high solvent density. The maximum density in these SDMs is $0.62\,\text{Å}^{-3}$ and $0.54\,\text{Å}^{-3}$ for water and methanol oxygens, respectively. The maxmimum density of nearest neighbour solvent oxygens around the triazole ring is roughly double the maximum density of nearest neighbour solvent oxygens around either the benzene or chlorobenzene rings. SDMs corresponding to both water and methanol molecules have areas of high solvent density in the **xz** plane. The position of these spots of density in the cut-throughs of the SMDs is consistent with hydrogen bonding of N_5 with both water and methanol discussed in relation to Fig. 5.11 and Table 5.9.

SDMs in Fig. 5.13 that show the distribution of nearest neighbour solvent molecules relative to the diazepine ring of alprazolam also imply that the favoured locations of the solvent molecules relative to this fragment of alprazolam are around the nitrogen atom of the ring (N_2). This is evidenced by the spot of density in the **xz** plane of the SDMs

in the bottom row of Fig. 5.13. The density of solvent molecules in this region is not as great as is seen around the nitrogen atom of the triazole ring (N_5), and the maximum density of nearest neighbour solvent oxygens is $0.36\,\text{Å}^{-3}$ and $0.29\,\text{Å}^{-3}$ for water and methanol, respectively. This lower density is unsurprising given the smaller peak in the ring-solvent oxyen RDFs in Fig. 5.12 compared with the corresponding RDFs in Fig. 5.10.

SDMs corresponding to those shown in Fig. 5.13, which were generated from EPSR, are shown for the same distributions generated from MD in Fig. 5.14. SMDs showing the distribution of solvent molecules relative to the benzene ring of alprazolam in Fig. 5.14 display a similar pattern of solvation as was seen in the SDMs generated from EPSR, the preferred regions of location being either side of the plane of the benzene ring. However, the density of the cut-throughs in the **xy** and **yz** planes of the SDMs in Fig. 5.14 point to a greater density of solvent molecules around the benzene ring of alprazolam in MD than in EPSR, with the maximum density of methanol oxygens ($0.55\,\text{Å}^{-3}$) again being greater than the maximum density of water oxygens ($0.51\,\text{Å}^{-3}$).

SDMs generated from MD showing the distribution of solvent molecules relative to the chlorobenzene ring of alprazolam are very similar to the corresponding SDMs generated from EPSR (Fig. 5.13) in terms of both the distribution of the methanol and water molecules and the density of the solvent molecules. In fact, the maxmimum density of both water and methanol around the chlorobenzene ring in MD is the same as was found in EPSR ($0.28\,\text{Å}^{-3}$).

Differences between the density of solvent molecules around the ring structures of alprazolam in the SDMs generated from EPSR and MD become more apparent when comparing the SDMs showing the distribution of solvent molecules relative to the triazole ring of alprazolam. The distribution of water and methanol molecules in the triazole-solvent SDMs in Fig. 5.14 is very similar to the distribution in their counterparts in Fig. 5.13, with the highest density of solvent molecules around N_5 in each case. However, in agreement with Table 5.8, the density of water and methanol molecules around N_5 is greater for MD than for EPSR. The maxmimum density of solvent oxygens around the triazole ring is greater in MD compared with EPSR for both water ($0.76\,\text{Å}^{-3}$) and

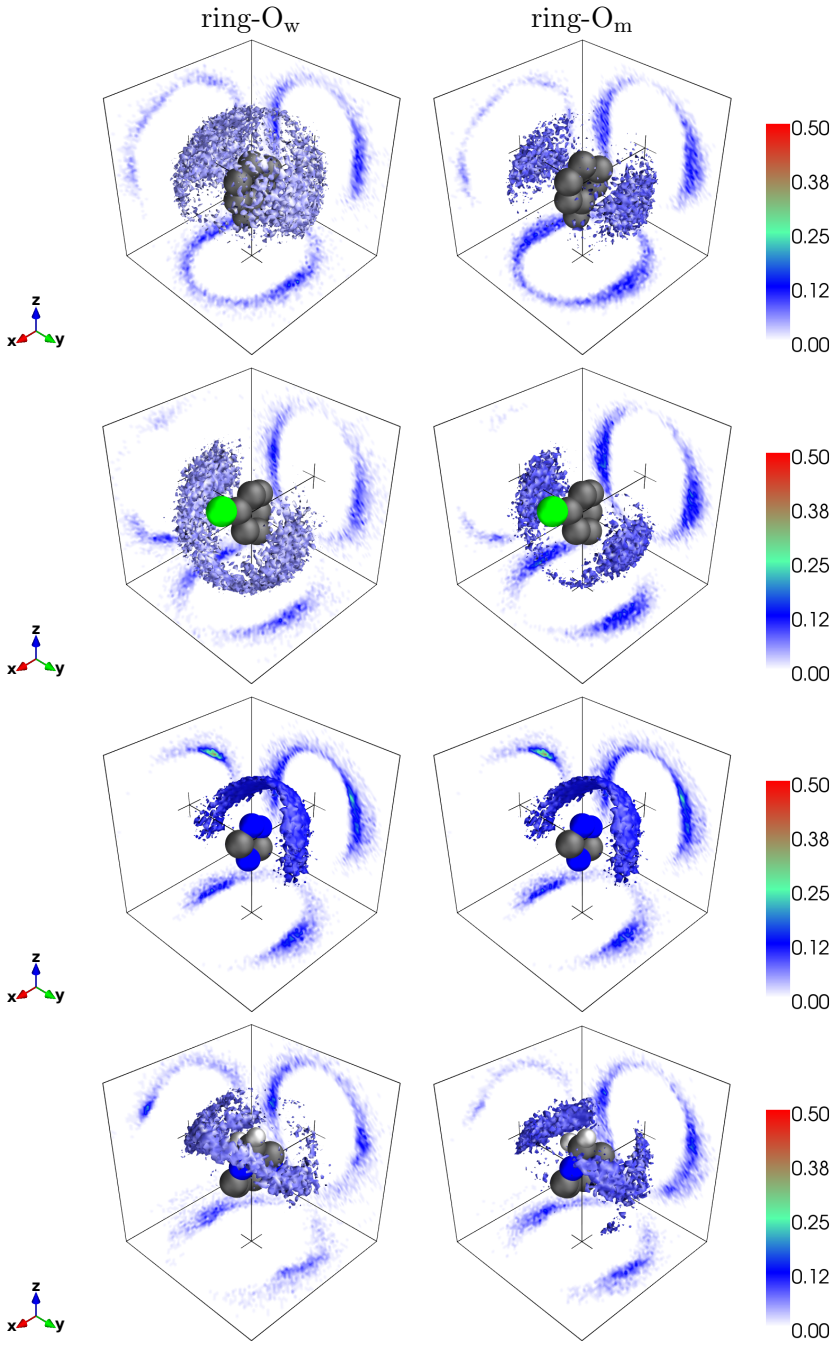


Figure 5.13: SDMs showing the distribution of nearest neighbour water molecules (left) and methanol molecules (right) relative to the benzene ring (1st row), the chlorobenzene ring (2nd row), the triazole ring (3rd row) and the diazepine ring (4th row) generated from EPSR. The isocontour surface enloses the top 30% of solvent molecules.

methanol (0.69 Å^{-3}).

Consistency in the distribution of solvent molecules between EPSR and MD is also seen when comparing the SDMs showing the distribution of solvent molecules around the diazepine ring of alprazolam generated from EPSR and MD. Like the triazole ring,

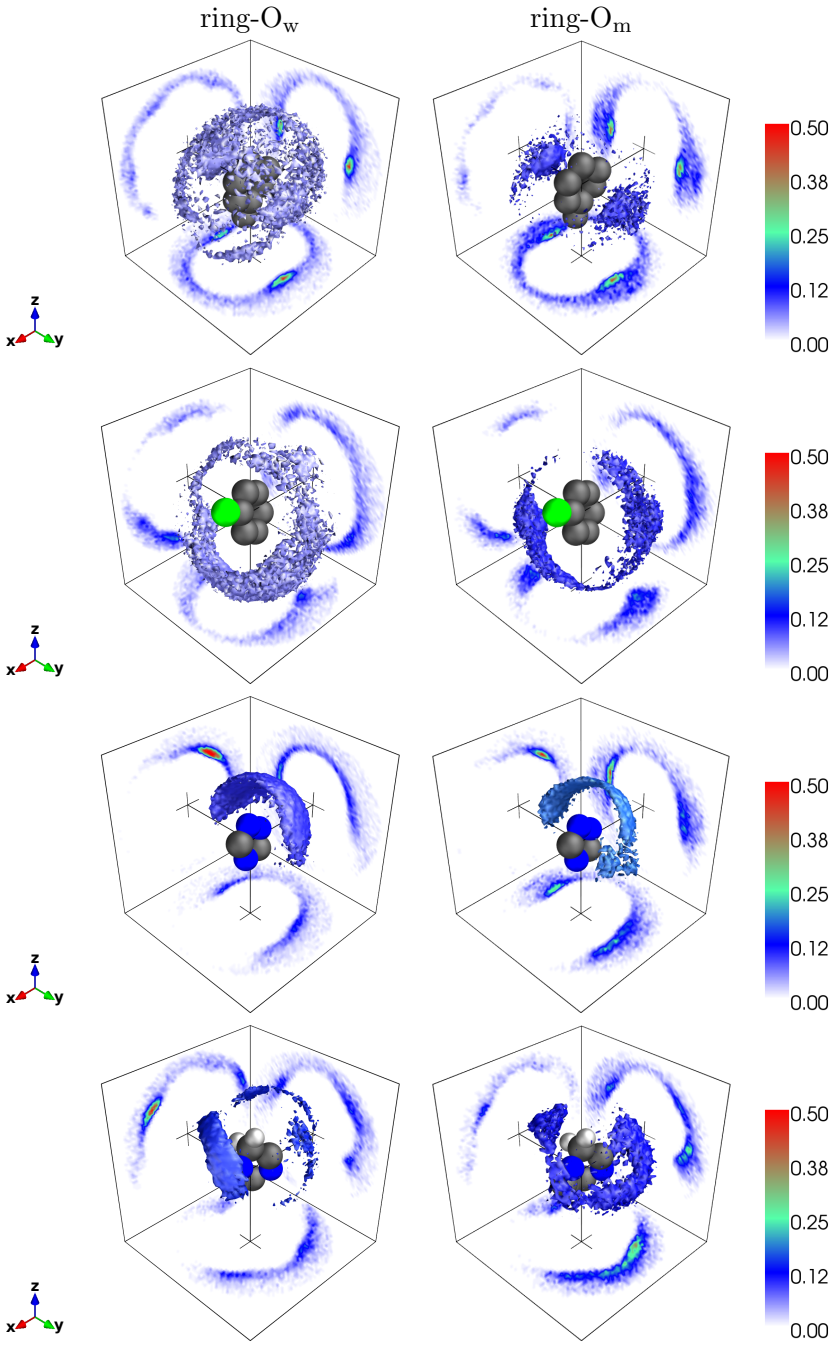


Figure 5.14: SDMs showing the distribution of nearest neighbour water molecules (top) and methanol molecules (bottom) relative to the benzene ring (1st row), the chlorobenzene ring (2nd row), the triazole ring (3rd row) and the diazepine ring (4th row) generated from MD. The isocontour surface enloses the top 30% of solvent molecules.

though, the density of solvent oxygens is greater in MD than EPSR for both water ($0.42\,\text{Å}^{-3}$) and methanol ($0.30\,\text{Å}^{-3}$).

5.3.4.3 Whole Molecule Analysis

WMA for water and methanol molecules within 4 Å of alprazolam is shown in Fig. 5.15 for MD. Also included in this figure are red spheres representing the crystallographic bridging water molecule sites (PDB ID: 3U5J) from the alprazolam-BRD4 complex. The BET protein BRD4 is essential for cellular growth and has been implicated in cell cycle control and DNA replication.[308] Recently a number inhibitors have been developed against BET proteins,[190] which have shown efficacy in several models of cancer.[250] These inhibitors, which include benzodiazepines,[68, 69] displace BET proteins, such as BRD4, from chromatin by competing with their acetyl-lysine recognition modules, leading to inhibition of oncogenic transcriptional programs.[250]

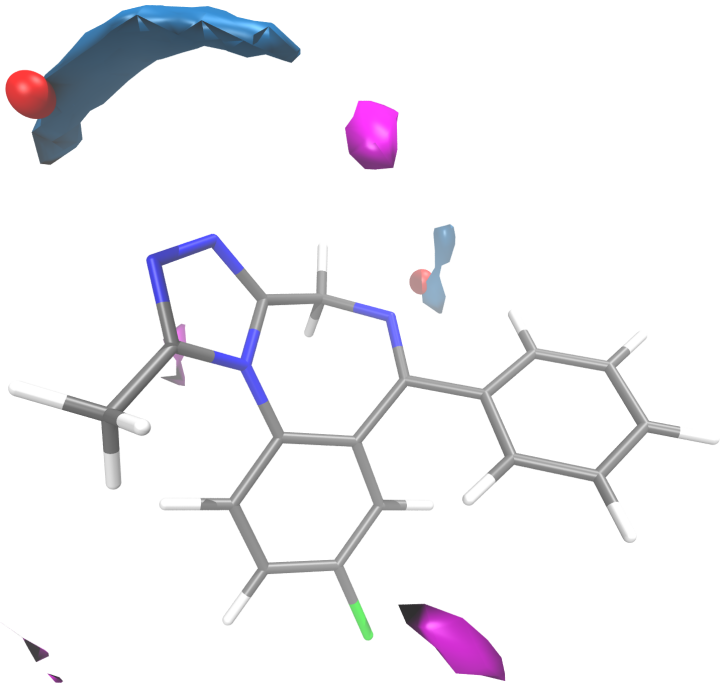


Figure 5.15: WMA for water (blue) and methanol (magenta) within 4 Å of alprazolam for MD. The isocontour surface is cut to include the top 40% of solvent molecules. The crystallographic bridging water molecule sites (PDB ID: 3U5J) are shown as red spheres.

Fig. 5.15 shows that the regions of alprazolam which are preferentially hydrated correspond to the locations of the crystallographic bridging water molecule sites from the alprazolam-BRD4 complex. The two bridging water molecules play an important role in the binding of the drug to BRD4 by mediating N_5-Tyr97 and N_2-Asn140 interactions.[4, 254] Water molecules seem to preferentially hydrate N_5 and N_2 in Fig. 5.15, while methanol interacts with the remaining hydrophobic motifs of alprazolam. Preferen-

tial hydration of N_5 is anticipated given that the region of the alprazolam molecule that has the highest density of nearest neighbour water molecules is N_5 in Fig. 5.13 and Fig. 5.14, and the RDFs with the clearest peaks correspond to triazole-solvent interactions (Fig. 5.10).

5.3.4.4 The Effect of Temperature

RDFs extracted from the MD simulation run at 300 K are shown in Fig. 5.16 along with the RDFs generated from the MD simulation run at 323 K, the RDFs shown up to this point, for the interactions of each of the rings of alprazolam with solvent oxygens. This

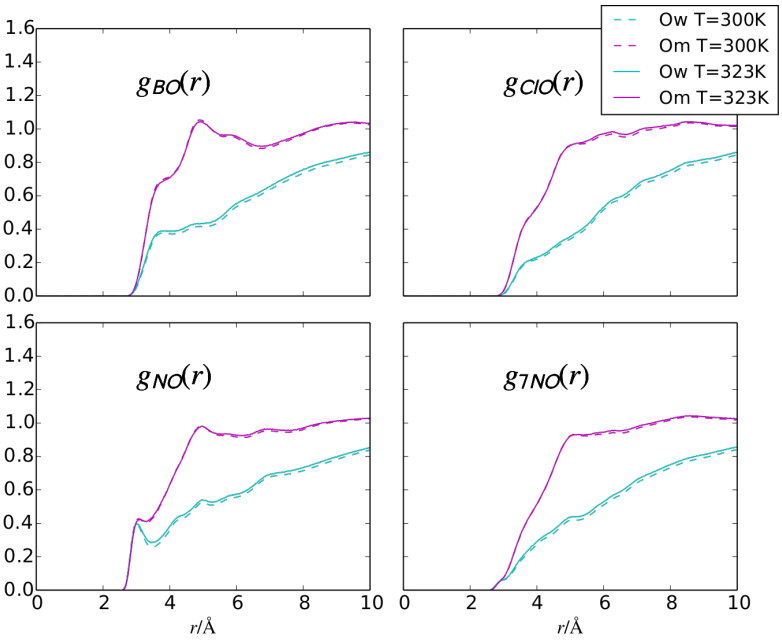


Figure 5.16: RDFs for the interactions of the benzene (B), chlorobenzene (Cl), triazole (N) and diazepam (7N) rings of alprazolam with water (cyan lines) and methanol (magenta lines) oxygens extracted from MD simulations run at 300 K (dashed lines) and 323 K (unbroken lines).

Figure suggests that the solvation of each of the rings of alprazolam is unaffected by measurement of alprazolam in methanol/water solutions at an elevated temperature, as the shape and the position of the peaks in Fig. 5.16 is almost identical for each interaction. Therefore, it is highly likely that over this temperature range (300 K-323 K), which encompasses physiologically relevant temperatures, there is no change in the solvation structure of alprazolam. A small reduction in the number of alprazolam-alprazolam interactions was observed at 323 K compared with 300 K.

5.4 Conclusions

EPSR and MD simulations of alprazolam dissolved in methanol/water solutions at 323 K revealed an absence of highly ordered water molecules around the benzene ring, which, in agreement with previous studies,[68] is rotated ∼30° relative to the plane of the rest of the molecule. The reduction in directed waters around the benzene ring of alprazolam compared with the benzene rings of indole (Section 3.3) and cocaine (Section 4.3) may be due to a greater number of hydrophilic groups with which water can interact compared with indole and a lower concentration of solute molecules compared with cocaine. Notwithstanding this difference, more water molecules were seen at a given distance from the benzene ring in EPSR than in MD. Methanol, in contrast, has a preferred location relative to the benzene ring of alprazolam. However, unlike in the indole/methanol/water system, the -CH_3 group of methanol has a preferred location that is closer to the benzene ring than the -OH group of methanol and water in both EPSR (∼1 Å) and MD (∼1.5 Å). Similar results were seen for the chlorobenzene and diazepine rings, showing that alprazolam motifs generally have a preference for interaction with methanol over water.

The constituent ring of alprazolam that experiences the greatest solvation is the triazole ring, principally through solvation of N_5. This solvation is evident through the triazole-solvent RDFs, where there is a well defined peak indicative of hydrogen bonding for both water and methanol, and the SDMs, which show a high density of solvent molecules around N_5 and a greater maximum density of water and methanol molecules around the triazole ring than for any of the other 3 rings of alprazolam. The importance of the triazole ring in determining the interaction of alprazolam with BRD4 has been highlighted previously, as replacement of the triazole methyl group with an ethyl or NH_2 group has been shown to result in a loss of BET bromodomain binding activity.[80] Triazole rings of BET inhibitors are thought to mimic the acetyl group of the typical biological ligand.[254] Additionally, the triazole ring of alprazolam has been shown to mediate strong hydrogen bonding interactions with a range of acids.[291]

WMA revealed that the preferred location of water molecules relative to the whole alprazolam molecule matched the locations of bridging water molecules in the crystal

structure of the alprazolam-BRD4 complex. Bridging water molecules are known to play a key role in mediating protein-ligand interactions.[160, 280] They have been described as acting as a 'glue' between ligand and protein, stabilising the protein-ligand interaction, and can also dictate specificity.[81, 170] Despite their importance in ligand binding, drug design projects routinely remove bridging water molecules from a biomolecule binding site because of the added complexity they represent.[81] However, bridging water molecules can be involved in ligand design, either by conservation of their hydration sites or by targeting them with ligand functional groups.[81] Multiple empirical and knowledge based protocols have been developed with the aim of predicting the location of bridging water molecules.[82, 108, 191] Many of these algorithms, such as Consolv[82] and WaterDock,[234] explicitly predict limited or no role for the ligand in determining the positions of bridging waters.

The results presented in this Chapter contradict the predictions of these computational protocols and suggest that hydration studies offer a possible pathway to determining the location of bridging water molecule sites. Although waters in crystallographic structures are difficult to locate due to being too mobile or due to low resolution,[39] meaning the results presented here require further validation, investigation of the solvation structure of the important class of pharmaceuticals to which alprazolam belongs, benzodiazepines, propounds a method by which a better understanding of how these molecules might bind to proteins of the BET family, in particular, could be gained.

Chapter 6

Clonidine in Methanol/water Solutions

6.1 Introduction

Adrenoceptors are membrane receptors responsible for mediating responses to the physiological agonists noradrenaline (norepinephrine) and adrenaline (epinephrine),[37, 57] catecholamines consisting of a benzene ring with two hydroxyl side groups and a side-chain amine. Consequently, adrenoceptors are responsible for controlling many important functions and processes (e.g. blood pressure, blood flow, neural modulation, digestion, pupil dilation, endocrine and metabolic processes).[37] These receptors are differentiated into α_1-, α_2- and β- adrenoceptors.[37] Generally α-adrenoceptors are considered excitatory and β-adrenoceptors are considered inhibitory,[57] with α_1- and α_2- adrenoceptors typically classed as postsynaptic and presynaptic, respectively.[238] The role of adrenoceptors in controlling vital bodily functions and processes means that they are of great therapeutic importance in the treatment of a range of diseases,[37] such as angina, asthma, glaucoma and hypertension.

An imidazole compound first marketed as an antihypertensive drug during the 1960s,[122] clonidine belongs to a class drugs thought to induce peripheral symphathoinhibition and reduction in elevated blood pressure as a result of the stimulation of α_2-

adrenoceptors.[227] Clonidine selectively stimulates α_2-adrenoceptors in the medulla oblongata[31, 107] and so inhibits the nerve-mediated release of noradrenaline,[122] the main neurotransmitter of sympathetic nerves in the cardiovascular system. It has been suggested that it is the protonated form of clonidine which interacts with biological receptors rather than the freebase form.[26, 227] Conditions treated with clonidine include high blood pressure, attention deficit hyperactivity disorder (ADHD), anxiety disorders, and certain pain conditions.[122, 159, 176, 276] A side effect associated with the use of clonidine as an antihypertensive drug is sedation.[122]

The action of clonidine has been related to its imidazole structure,[83, 92, 215] having been shown to possess only a slightly lower affinity for non-adrenergic imidazoline receptors compared with α_2-adrenoceptors.[216] In fact, there is evidence, particularly from binding studies, suggesting that the non-adrenergic imidazoline receptor action of clonidine is more important than its α_2-adrenoceptor action.[83, 102] However, it is difficult to separate the functional effects of interaction with α_2-adrenoceptors from that of imidazoline receptors,[102] and it may be possible that some of the effects of α_2-adrenoceptor agonists are mediated by imidazoline receptors.[139] Prior to the discovery of imidazoline binding sites, the mode of action of clonidine was attributed solely to the activation of α_2-adrenoceptors.[146]

In this Chapter, the atomic structure of clonidine hydrochloride is investigated in methanol/water solutions. Detailed analysis of the solvation of the hydrophilic and hydrophobic chemical groups of clonidine by water and methanol molecules means that, similar to preceding Chapters, the preferential solvation of clonidine in solution is determined. The preferential solvation of clonidine and its molecular conformation are used to draw conclusions regarding its interaction with α_2-adrenoceptors and imidazoline receptors.

6.2 Methods

6.2.1 Neutron Diffraction Sample Preparation

Clonidine hydrochloride was purchased from Sigma Aldrich (CAS 4205-91-8) and used without further purification. The molecular structure of the protonated form of clonidine is shown in Fig. 6.1, which is distinguished from the freebase form of clonidine by a proton (H_N) attached to the nitrogen atom (N) joining the the benzene and imidazole rings.[a] Following the procedure outlined in Section 3.2.1, 3g of clonidine·HCl was deuterated. Deuteration of H_N, H_{N1} and H_{N2} was verified by measuring clonidine·HCl dissolved in $CDCl_3$ using ^1H NMR. The NMR sample was prepared and measured as outlined in Section 3.2.6. Integration indicated that each amine hydrogen was ∼91% deuterated. Anhydrous methanol and degassed water were prepared as described in Section 3.2.1. Samples for neutron diffraction were prepared by mass under an N_2 atmosphere, as in Section 3.2.1.

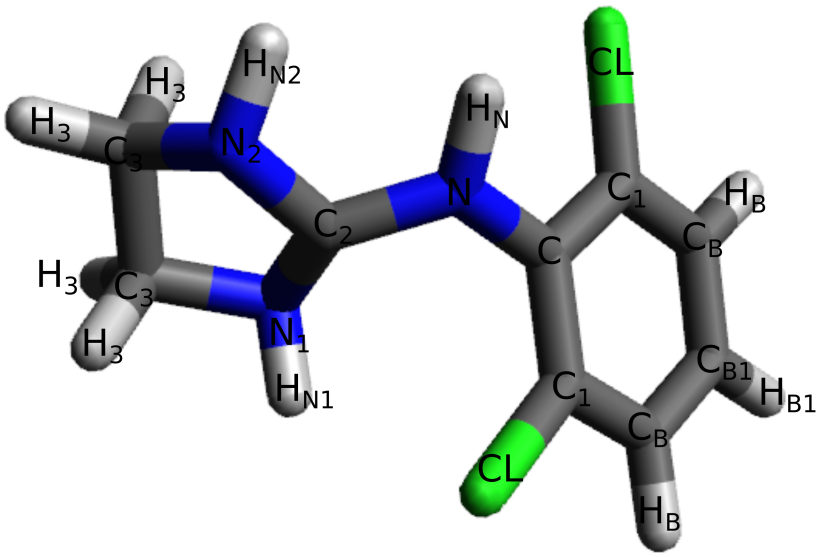


Figure 6.1: The molecular structure and labelling scheme of the protonated form of clonidine.

Eight different isotopomers of clonidine·HCl dissolved in methanol/water solutions were prepared at a relative molar ratio of 1 clonidine hydrochloride: 100 methanol: 100 water (∼0.17 M). The samples measured by neutron diffraction are listed in Table 6.1,

[a]The protonation site which distinguishes the protonated form of clonidine from the freebase form was not actually clear from the literature and was determined by NMR.

where h-clonidine and d-clonidine refer to clonidine·HCl with an H atom and a D atom attached to each of N, N_1 and N_2, respectively.

Sample Name	**Clonidine**	**Methanol**	**Water**
h-clonidine/CH_3OH/H_2O	h-clonidine	CH_3OH	H_2O
h-clonidine/CHD_3OH/H_2O	h-clonidine	$CH_{1.5}D_{1.5}OH$	H_2O
d-clonidine/CH_3OD/D_2O	d-clonidine	CH_3OD	D_2O
d-clonidine/CD_3OD/D_2O	d-clonidine	CD_3OD	D_2O
d-clonidine/CHD_3OD/D_2O	d-clonidine	CH_1D_2OD	D_2O
hd-clonidine/CH_3OHD/HDO	$h_{0.5}d_{0.5}$-clonidine	$CH_3OH_{0.5}D_{0.5}$	HDO
hd-clonidine/CD_3OHD/HDO	$h_{0.5}d_{0.5}$-clonidine	$CD_3OH_{0.5}D_{0.5}$	HDO
hd-clonidine/CHD_3OHD/HDO	$h_{0.5}d_{0.5}$-clonidine	$CH_{1.5}D_{1.5}OH_{0.5}D_{0.5}$	HDO

Table 6.1: The components of the eight samples studied. The molar ratio used was 1 clonidine hydrochloride: 100 methanol: 100 water. Subscripts indicate isotopic proportions.

6.2.2 Neutron Diffraction Measurements

Neutron diffraction data were collected for each of the samples listed in Table 6.1 using the same experimental setup as described for cocaine in aqueous solution in Section 4.2.2.

Data were collected for the empty cells, the empty instrument and a vanadium standard. The μA hrs each sample was measured for and the corresponding DCS levels are shown in Table 6.2. The empty cells were each measured for 1130-1415 μA hrs. The empty instrument and vanadium standard were measured for 501 μA hrs and 564 μA hrs, respectively. Each of the samples had a DCS level within $\sim$14% of their theoretical values, and the mean DCS level of the samples is $\sim$91% of the theoretical value. The vanadium standard and each of the empty SiO_2 cells had DCS levels 99%-100%. and 98%-100% of their theoretical values, respectively.

6.2.3 X-ray Diffraction Measurements

Sample h-clonidine/CH_3OH/H_2O was also measured by X-ray diffraction for 24 hrs, using the same experimental setup and procedures as described for cocaine in aqueous solution in Section 4.2.3.

Sample Name	Measurement Time (μA hrs)	DCS Level
h-clonidine/CH_3OH/H_2O	1176	102.3
h-clonidine/CHD_3OH/H_2O	1229	100.5
d-clonidine/CH_3OD/D_2O	1060	103.2
d-clonidine/CD_3OD/D_2O	1416	111.3
d-clonidine/CHD_3OD/D_2O	990	96.8
hd-clonidine/CH_3OHD/HDO	1185	100.4
hd-clonidine/CD_3OHD/HDO	1152	104.1
hd-clonidine/CHD_3OHD/HDO	1258	102.3

Table 6.2: The measurement time and DCS level for each sample. The average current was $\sim$170μA.

6.2.4 Empirical Potential Structure Refinement

The 61.8 Å cubic EPSR simulation box contained 25 protonated clonidine molecules, 25 Cl^- ions, 2500 methanol molecules and 2500 water molecules at the measured solution density ($\rho = 0.098$ atoms Å^{-3}) at 298 K.

The SPC/E water model was used.[18] The molecular structure of methanol was generated and energetically optimised using Avogadro.[98] Methanol and Cl^- parameters were those used in Section 3.2.3 and Section 4.2.4 for indole in methanol/water solutions and for cocaine in aqueous solution, respectively. The labelling schemes of methanol and water are shown in Fig. 3.1.

The molecular structure of clonidine was generated and energetically optimised using Avogadro,[98] and clonidine parameters were generated with ANTECHAMBER.[299, 300] Subsequently, collections of atoms were grouped together depending on the similarity of their charges and the mean charge of these groups taken. A full list of the parameters used in EPSR is provided in Table 6.3, with the corresponding labelling scheme in Fig. 6.1. This simulation was accummulated for 43021 steps.

Atom Label	ϵ (KJ mol^{-1})	σ(Å)	Atomic Mass (u)†	$\mathbf{q}_{\mathrm{EPSR}}$(e)	$\mathbf{q}_{\mathrm{MD}}$(e)
CL	1.1090	3.470	35.453	-0.0494	-0.4240
CL	1.1090	3.470	35.453	-0.0494	-0.0564
C_{B1}	0.3600	3.400	12.011	-0.0933	-0.1020
C_B	0.3600	3.400	12.011	-0.0933	-0.0740

C_B	0.3600	3.400	12.011	-0.0933	-0.1040
C	0.3600	3.400	12.011	-0.0742	-0.0740
C_1	0.3600	3.400	12.011	0.0239	0.0314
C_1	0.3600	3.400	12.011	0.0239	0.0164
C_2	0.3600	3.400	12.011	0.4820	0.4833
C_3	0.4580	3.400	12.011	0.0023	0.0033
C_3	0.4580	3.400	12.011	0.0023	0.0013
H_{B1}	0.0630	2.600	2.000	0.1780	0.1740
H_B	0.0630	2.600	2.000	0.1780	0.1810
H_B	0.0630	2.600	2.000	0.1780	0.1790
H_3	0.0660	2.470	2.000	0.1000	0.0937
H_3	0.0660	2.470	2.000	0.1000	0.1057
H_3	0.0660	2.470	2.000	0.1000	0.1057
H_3	0.0660	2.470	2.000	0.1000	0.0937
N	0.7110	3.250	14.007	-0.2902	-0.2902
H_N	0.0660	1.070	2.000	0.3407	0.3407
N_1	0.7110	3.250	14.007	-0.3612	-0.3612
H_{N1}	0.0660	1.070	2.000	0.3387	0.3387
N_2	0.7110	3.250	14.007	-0.3772	-0.3762
H_{N2}	0.0660	1.070	2.000	0.3337	0.3337
Cl	0.4184	4.401	35.453	-1.0000	-1.0000
C_m	0.27614	3.5000	12.011	0.1450	0.1450
H_m	0.12552	2.5000	2.000	0.0400	0.0400
O_m	0.71128	3.1200	16.000	-0.6830	-0.6830
H_{mo}	0.00000	0.0000	2.000	0.4180	0.4180
O_w	0.65000	3.1660	16.000	-0.8476	-0.8476
H_w	0.00000	0.0000	2.000	0.4238	0.4238

Table 6.3: The parameters used in the EPSR and MD simulations of clonidine hydrochloride/methanol/water. † The atomic mass of hydrogen used in MD was 1.000 u. The corresponding labelling scheme for clonidine is shown in Fig. 6.1.

6.2.5 Molecular Dynamics

MD simulations were run using Gromacs 5.0.2,[272] with the same number of each molecular species being used as in EPSR. Clonidine parameters (Table 6.3) were generated using AMBER and were converted to Gromacs format using the amb2gmx.pl script.[186] The SPC/E water model was used.[18, 165] Methanol parameters and Cl^- parameters (Table 6.3) were generated as described in Section 3.2.4 and Section 4.2.5, respectively. Initially, energy minimisation corrected atomic overlaps until the maximum force on any atoms was below $1000\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$. Energy minimisation was followed by solute restrained equilibration runs of 1 ns each in the NVT and NPT ensembles. The Berendsen thermostat and barostat were used to maintain temperature and pressure, respectively, in the equilibration runs.[18] The production run was carried out in the NPT ensemble for 80 ns with a timestep of 2 fs. Temperature was maintained at 298 K using the V-rescale thermostat, while pressure was controlled using the Parrinello-Rahman barostat.[206] A cut-off of 14 Å was used for the van der Waals interactions, and long range interactions were calculated using the Particle Mesh Ewald algorithm.[50, 166] Coordinates were saved every 0.2 ns, resulting in 100000 frames.

6.2.6 ANGULA

The orthogonal coordinate systems assigned to clonidine in ANGULA are shown in Fig. 6.2. Coordinate systems were assigned to each of the nitrogen atoms and to the benzene ring of clonidine. Coordinate systems were assigned to methanol and water as described in Section 3.2.5. WMA was performed for a distance range of 0-4 Å. Analysis using ANGULA was performed for ∼3500 snapshots of each of the EPSR and MD simulation boxes.

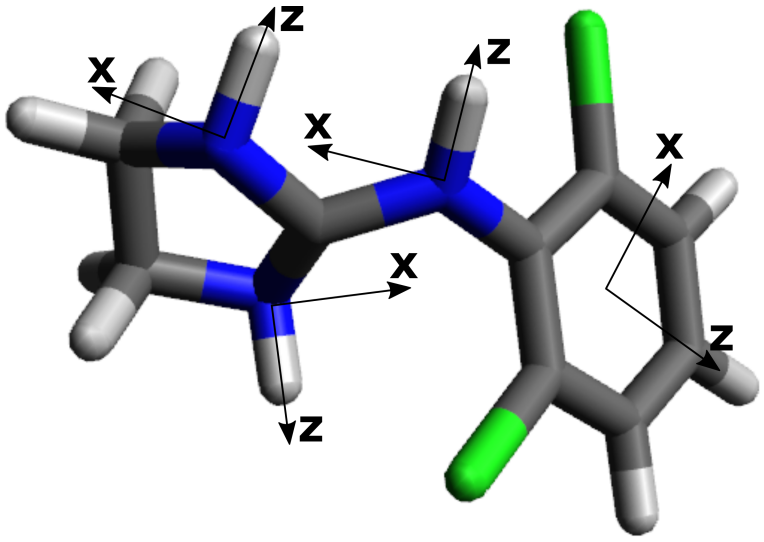


Figure 6.2: The coordinate systems assigned to clonidine using ANGULA. The **y** axes are coming out of the plane of the page.

6.2.7 NMR Sample Preparation and Measurement

NMR samples were prepared and measured[b] as described in Section 3.2.6. Clonidine·HCl and the freebase form of clonidine (Sigma Aldrich, CAS 4205-90-7) were assigned using a combination of ^{1}H and ^{1}H-^{1}H 2D NMR spectra. Clonidine·HCl was measured in H_2O, with HCl being added to the sample to enable exchangeable protons to be discerned. The sample measured by NMR had a pH of 3. The freebase form of clonidine was measured in deuterated tetrahydrofuran, THF (C_4H_8O), which was purchased from Sigma Aldrich (99.5%D) and dried over potassium. Both samples were measured at a concentration of 1 clonidine: 200 solvent. Measurements were also made of clonidine·HCl in methanol/water solutions at different temperatures to verify that the structure measured by diffraction was not different to the structure at physiologically relevant temperatures, using TMS as a standard. The labelling scheme used in NMR spectra of clonidine in solution corresponds to Fig. 6.1.

[b]NMR samples were prepared and measured by exchange student Medeea Popescu.

6.3 Results

6.3.1 Diffraction Data

Measured neutron and X-ray diffraction data and EPSR fits to these data are shown in reciprocal space in Fig. 6.3 (a). The EPSR fits to the neutron data are generally good, with some small differences at low Q values (Q <5 Å^{-1}) due to difficulty correcting for inelastic scattering this region. At larger Q values (Q >5 Å^{-1}) the difference curves are flat. The X-ray data fits are also good, again with some differences at low Q values. This difference is likely because polarizability effects are not taken into account, as the weighted sum of atomic form factors is currently used to obtain the static structure factor.[261] The Fourier transformations of these data and EPSR fits in real space are shown in Fig. 6.3 (b), where a similar level of agreement between the experimental and simulated data is observed as in reciprocal space.

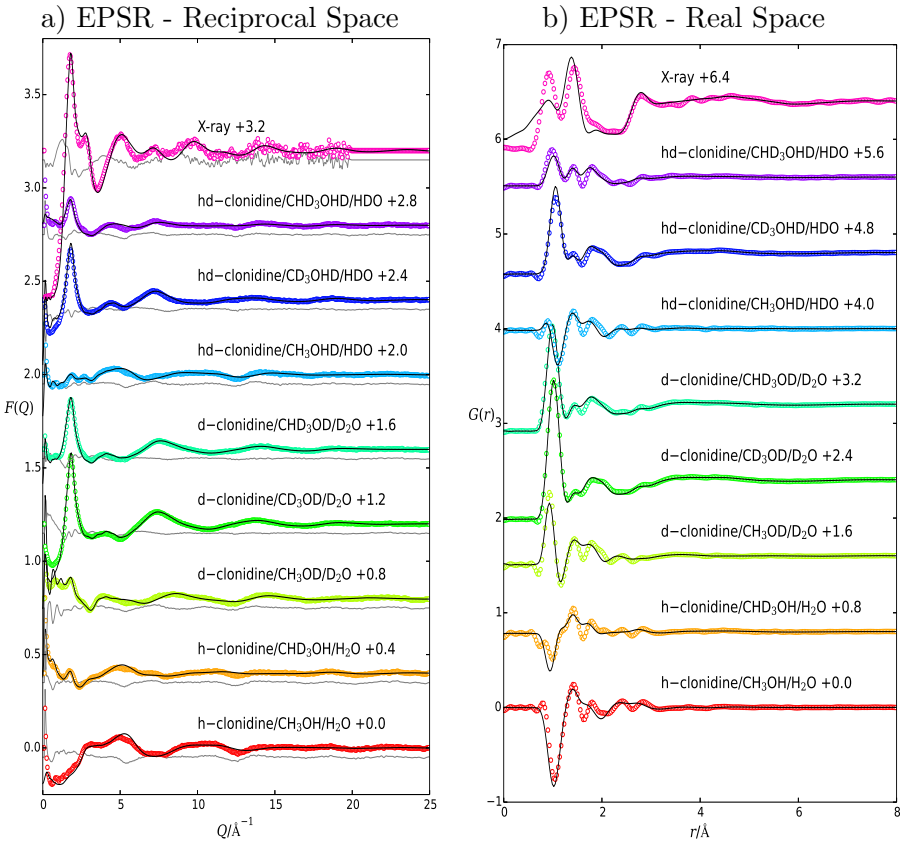


Figure 6.3: Measured neutron and X-ray diffraction data (coloured lines) and EPSR fits to the data (black lines) in reciprocal space (a) and real space (b). The difference curves (grey lines) have been shifted by 0.05 for clarity.

The same neutron and X-ray diffraction data shown in Fig. 6.3 are shown in Fig. 6.4 along with the $F(Q)$s and $G(r)$s generated from the MD trajectory in reciprocal space and real space in Fig. 6.4 (a) and Fig. 6.4 (b), respectively. The difference curves in Fig. 6.4 (a) are not as flat as those in Fig. 6.3 (a), and the MD $F(Q)$s match the experimental data less closely than the EPSR fits. Similarly, the agreement between the experimental and the simulated data in Fig. 6.4 (b) is worse than the agreement in Fig. 6.3 (b). An example of this worse agreement is seen for the first peak/trough in Fig. 6.4 (b) for samples h-clonidine/CH_3OH/H_2O and hd-clonidine/CD_3OHD/D_2O. As inspection of the simulation box revealed no aggregation of clonidine molecules, the worse agreement of the MD $F(Q)$s with the experimental data compared with the EPSR fits may be attributed to the larger distribution of intramolecular distances afforded by EPSR, which is more consistent with the measured diffraction data.

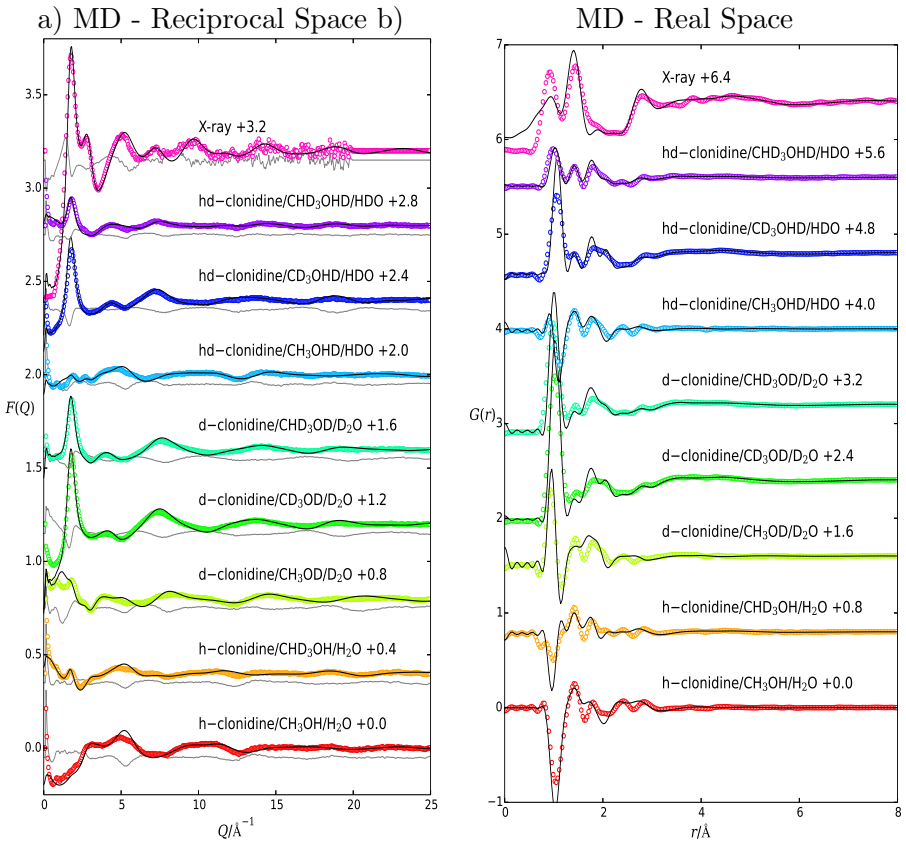


Figure 6.4: Measured neutron and X-ray diffraction data (coloured lines) and the $F(Q)$s and $G(r)$s generated from MD (black lines) in reciprocal space (a) and real space (b), respectively. The difference curves (grey lines) have been shifted by 0.05 for clarity.

6.3.2 Clonidine Conformation

Due to the importance of the conformation of molecules in understanding the behaviour of pharmaceuticals in solution, the conformation of clonidine in EPSR and MD was analysed using the axes assigned to clonidine in Section 6.2.6. As the benzene ring and imidazole ring of clonidine are rigid, the conformation of clonidine is effectively defined by the relative orientation of these two rings.

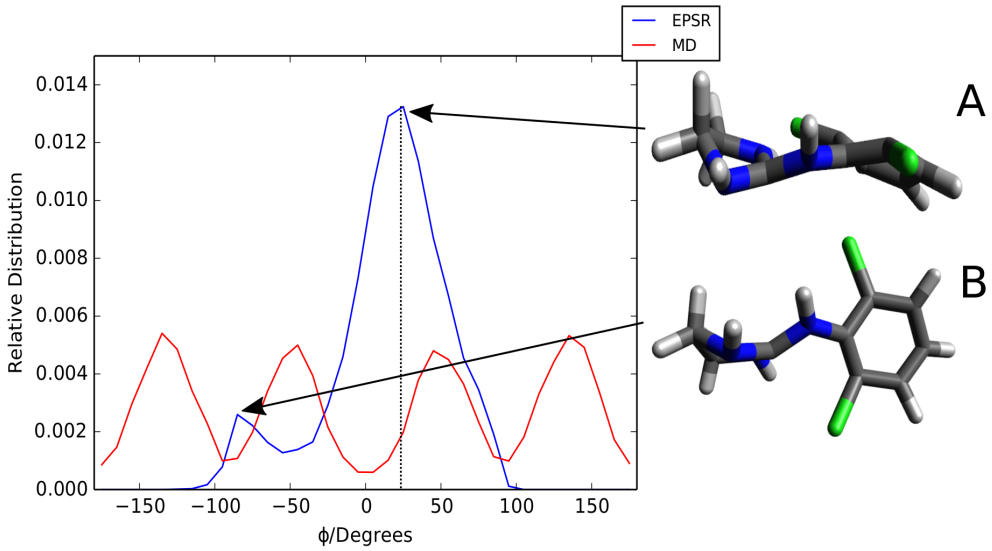


Figure 6.5: The angle of the benzene ring relative to the imidazole ring of clonidine in EPSR (blue line) and MD (red line). An angle of 0° corresponds to the rings being exactly coplanar. Clonidine molecules taken from the EPSR simulation box are shown which correspond to the two identified populations in EPSR, which exist with their rings approximately coplanar (A) and perpendicular (B).

Fig. 6.5 shows that in EPSR clonidine exists in two populations. The smaller of these populations comprises ∼10% of the clonidine molecules. These clonidine molecules have their benzene and imidazole rings orientated almost perpendicularly (∼85°). The remaining clonidine molecules, which constitute ∼90% of the solute molecules in the system, exist with their rings at a relative orientation of ∼25°.

In MD the clonidine molecules do not exist in two populations roughly corresponding to their benzene and imidazole rings being either perpendicular or coplanar. Instead, the benzene and imidazole rings of the clonidine molecules in MD are orientated ±40° relative to each other. This relative orientation manifests itself as 4 peaks in Fig. 6.5 at ±40° and ±140°, corresponding to 4 equal populations of clonidine molecules.

Therefore, although there is a small population of clonidine molecules in EPSR whose benzene and imidazole rings are almost perpendicular, the majority of clonidine molecules in both EPSR and MD exist with their ring structures close to be being in the same plane (ϕ is $\sim$25° or $\sim$40°). The results presented here mirror those presented for alprazolam dissolved in methanol/water solutions in Section 5.3.2, where the benzene ring was slightly tilted from the plane of the rest of the molecule in both EPSR and MD. Studies of the conformation of the freeebase form of clonidine using computational techniques indicate that the benzene and imidazole rings of clonidine are not coplanar.[227]

6.3.3 Solvent Structure

Fig. 6.6 shows water-water and methanol-methanol RDFs extracted from the EPSR and MD simulations, with corresponding CNs computed at the peak minima being shown in Table 6.4. RDFs extracted from EPSR simulations of pure water[266] and pure methanol[56] are included in Fig. 6.6 for comparison.

The positions of the peaks in the water-water RDFs extracted from both EPSR and MD in Fig. 6.6 match those shown in Section 3.3 for indole in methanol/water solutions and in Section 5.3 for alprazolam in methanol/water solutions and, therefore, also the peaks in the RDFs extracted from the EPSR simulations of pure water. As discussed in detail in Section 3.3 and Section 5.3, these findings indicate that there is no enhancement of the water structure in methanol/water solutions compared with pure water.[268] Similarly, the positions of the peaks in the methanol-methanol RDFs agree with the peak positions previously reported for pure methanol and methanol/water solutions.[6, 56] Accordingly, the water-water and methanol-methanol CNs in Table 6.4 differ by no more than 10% with respect to the corresponding CNs in Section 3.3 (indole in methanol/water solutions). The trend seen in Section 3.3 for the water-water CNs to increase from EPSR1 to EPSR2 to MD was attributed to differences in the propensity of water to interact with indole in each of the simulations due to differences in the polarity of the charges of some atoms between the three simulations. The CNs shown in Table 6.4 in this Section support this conclusion, as the water-water CNs here, where

the variations in charge distribution described in Section 3.3 do not exist, are slighly greater in EPSR than MD, meaning that the water-water CN trend is not a result of the simulation technique.

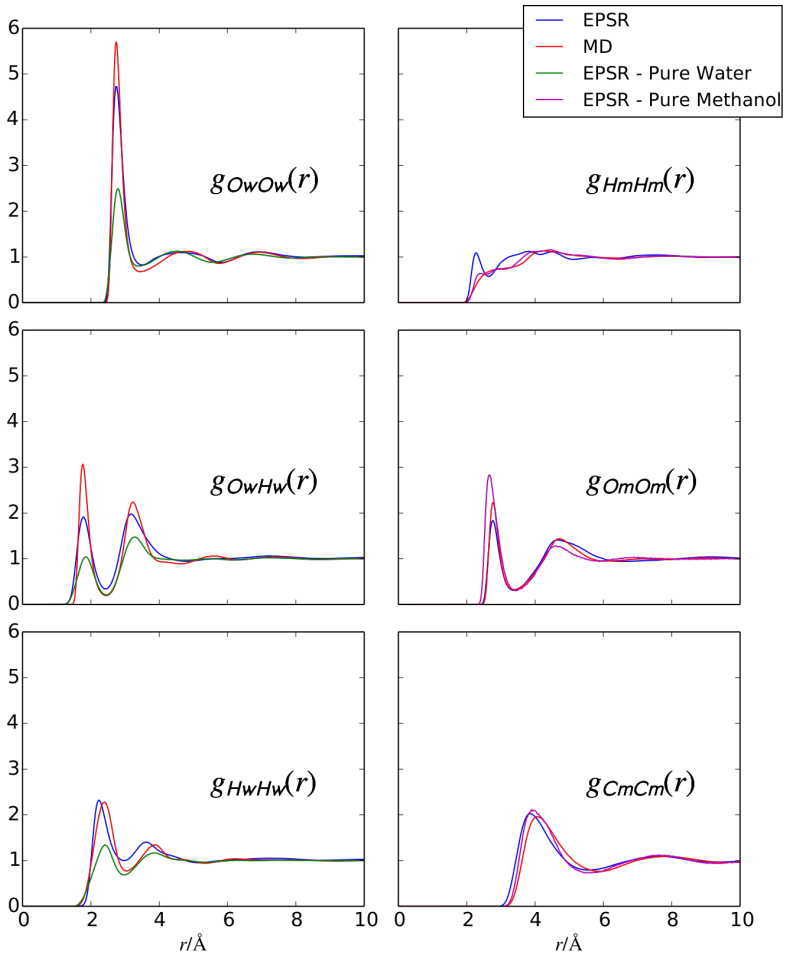


Figure 6.6: RDFs for water-water interactions extracted from EPSR (blue line) and MD (red line) simulations of clonidine hydrochloride/methanol/water. RDFs extracted from EPSR simulations of pure water[266] (green line) and pure methanol[56] (magenta line) are also included.

n_{α}^{β}	**EPSR**	**MD**	r/Å
$n_{O_w}^{O_w}$	2.48	2.41	3.48
$n_{O_w}^{H_w}$	1.01	1.00	2.43
$n_{H_w}^{H_w}$	2.61	2.58	3.00
$n_{H_m}^{H_m}$	1.03	0.82	2.64
$n_{O_m}^{O_m}$	0.89	0.93	3.39
$n_{C_m}^{C_m}$	7.61	7.34	5.58

Table 6.4: CNs of the water-water and methanol-methanol RDFs shown in Fig. 6.6.

Fig. 6.7 shows the RDFs for the solvent-solvent hydrogen bonding interactions be-

tween methanol and water molecules in the present solution. As in Section 3.3 (indole in methanol/water solutions) and Section 5.3 (alprazolam in methanol/water solutions), the first peak positions in all of the RDFs in this figure show that, on average, H_{mo} and H_w form bonds at ∼1.8 Å to O_m and O_w. Coupled with the CNs shown in Table 6.5, this indicates that both the solvents have a similar propensity to interact with each other as with themselves.

For EPSR, $n_{H_w}^{O_w}+n_{H_w}^{O_m} = 0.50 + 0.44 = 0.94$ and $n_{H_{mo}}^{O_w}+n_{H_{mo}}^{O_m} = 0.50 + 0.41 = 0.91$. Performing the same analysis for MD, $n_{H_w}^{O_w}+n_{H_w}^{O_m} = 0.50+0.43 = 0.93$ and $n_{H_{mo}}^{O_w}+n_{H_{mo}}^{O_m} = 0.52 + 0.44 = 0.96$. Therefore, the hydrogen bonding of the system is marginally undersaturated with respect to H_w and H_{mo} hydrogen bonding compared with the theoretical value of 1 in the absence of any solute, with no clear differences associated with the simulation technique. This slight undersaturation is likely due to hydrogen bonding between solvent and solute. The greater saturation of the system with respect to H_w and H_{mo} hydrogen bonding compared to Section 3.3 may be a consequence of the reduced concentration of the clonidine/methanol/water system compared to the concentration of indole in methanol/water solutions.

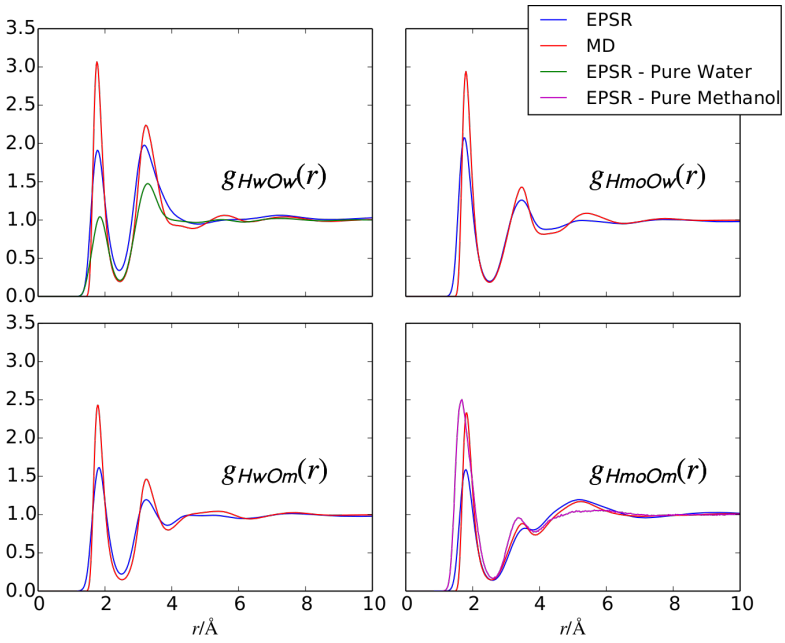


Figure 6.7: RDFs for hydrogen bonding of water (O_w and H_w) and methanol (O_m and H_{mo}) atoms extracted from EPSR (black line) and MD (red line). Corresponding RDFs extracted from EPSR simulations of pure water[266] (green line) and pure methanol[56] (magenta line) are also included.

n_{α}^{β}	**EPSR**	**MD**	r**/Å**
$n_{H_w}^{O_w}$	0.50	0.50	2.43
$n_{H_w}^{O_m}$	0.44	0.43	2.52
$n_{H_{mo}}^{O_w}$	0.50	0.52	2.52
$n_{H_{mo}}^{O_m}$	0.41	0.44	2.58

Table 6.5: CNs of the hydrogen bonding RDFs shown in Fig. 6.7.

6.3.4 Clonidine Solvation

6.3.4.1 Radial Distribution Functions

Fig. 6.8 shows RDFs for the interactions of the amine nitrogens of clonidine with solvent oxygens. For both EPSR and MD the sharp peaks in these RDFs indicate that both water and methanol oxygens have a well defined preferred distance of ∼2.9 Å from the clonidine nitrogens. A second, broader peak is also seen at greater distances in each of these RDFs. For both water and methanol, across EPSR and MD, this peak is located at ∼6.8 Å, ∼4.7 Å and ∼4.7 Å for N, N_1 and N_2, respectively. The second peak in each of these RDFs is likely to be associated with interactions between neighbouring atoms of clonidine and solvent molecules which necessarily manifest themselves in the RDFs. For example, the second peak in $g_{N_1O_w}(r)$ is most likely a result of the interactions between the amine nitrogen on the opposite side of the imidazole ring (N_2) and water oxygens, the waters around N_2 necessarily contributing density to $g_{N_1O_w}(r)$. Hence, the symmetry of the imidazole ring accounts for why the $g_{N_1O_w}(r)$ and $g_{N_2O_w}(r)$, and the corresponding nitrogen-methanol RDFs, have a second peak at the same distance (∼4.7 Å).

Corresponding CNs shown in Table 6.6 reveal differences in the average number of solvent molecules around each nitrogen in EPSR and MD. For example, N and N_1 are more hydrated in MD than in EPSR, while the average number of waters around N_2 is greater for EPSR. Differences in the hydration of these nitrogens between EPSR and MD is most apparent for N_1, where the CN is more than 50% larger in MD compared with EPSR. The CNs for the methanol molecules around the amine nitrogens show that N and N_2 have more methanol molecules located at the distances at which the CNs were

computed in Table 6.6 for EPSR compared with MD. The amine nitrogen for which there is the largest difference in methanol CN between EPSR and MD is N_2, which has a CN 46% larger for EPSR compared with MD.

There is no clear pattern of increasing CN associated with the simulation technique for either solvent. There is also no evidence of any of the amine nitrogens being preferentially more solvated in either EPSR or MD when comparing between the two solvents. For instance, $n_N^{O_w}$ is larger for MD but $n_N^{O_m}$ is larger for EPSR. The total CNs for the interactions between amine nitrogens and water and methanol oxygens at the distances listed in Table 6.6 are 5.65 and 5.00 for EPSR and MD, respectively. That the total solvent oxygen CNs for EPSR and MD only differ by 13% highlights that discrepancies in the solvation of the amine groups between the two simulation techniques are minimal. Moreover, the discussed similarities between the peak positions in Fig. 6.8 indicate that the nature of any such solvation is preserved between EPSR and MD.

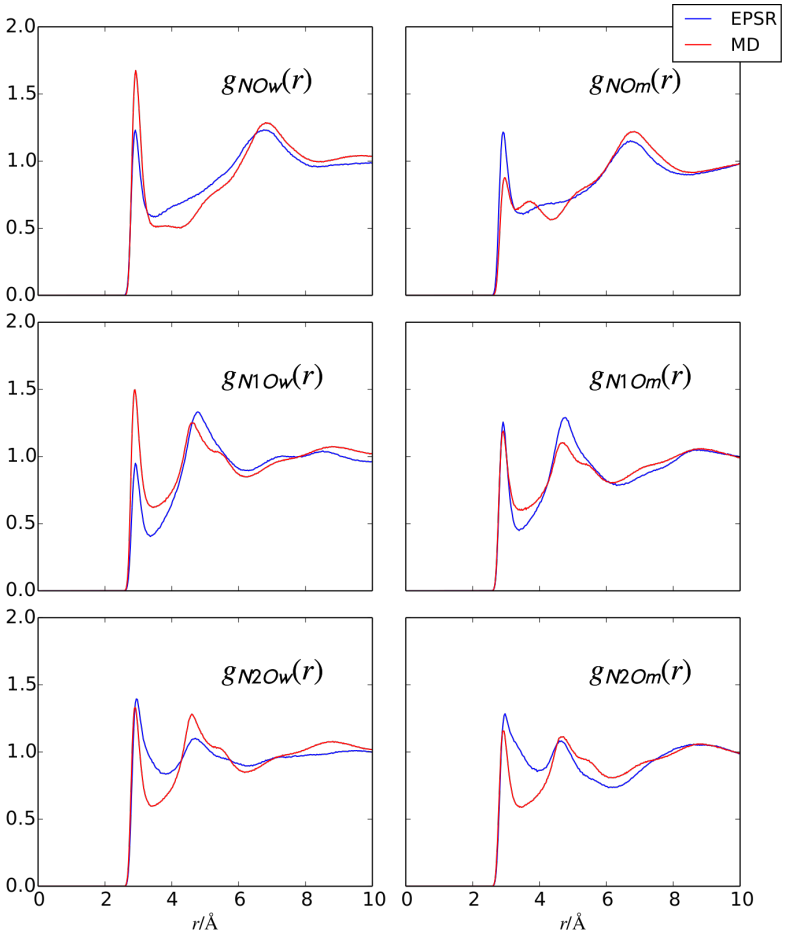


Figure 6.8: RDFs for the interactions of the nitrogen atoms of clonidine (N, N_1 and N_2) with solvent oxygens (O_w and O_m) extracted from EPSR and MD.

n_{α}^{β}	**EPSR**	**MD**	r/Å
$n_{N}^{O_w}$	0.72	0.82	3.42
$n_{N_1}^{O_w}$	0.48	0.74	3.36
$n_{N_2}^{O_w}$	1.40	1.04	3.72
$n_{N}^{O_m}$	0.73	0.58	3.42
$n_{N_1}^{O_m}$	0.63	0.66	3.39
$n_{N_2}^{O_m}$	1.69	1.16	3.87

Table 6.6: CNs of the clonidine/methanol/water RDFs shown in Fig. 6.8.

RDFs analogous to those in Fig. 6.8 are shown in Fig. 6.9 for the interactions between amine hydrogens and solvent oxygens in EPSR and MD. CNs corresponding to Fig. 6.9 are shown in Table 6.7. The RDFs in Fig. 6.9 each show sharp, well defined peaks at ∼1.95 Å, with no difference in the position of this peak between EPSR and MD. This is consistent with a moderate, mostly electrostatic hydrogen bond between the amine hydrogens and then solvent oxygens.[274] Second peaks are visible in these RDFs, which can be explained by interactions between the solvent and other atoms of clonidine, as was discussed in relation to Fig. 6.8.

As might be anticipated, the CNs in Table 6.7 are of similar relative values as those shown for the amine nitrogens in Table 6.6. For instance, the average number of water oxygens around the amine hydrogens is larger for MD compared with EPSR in the cases of H_N and H_{N1}, as was seen in Table 6.6 for the corresponding amine nitrogens. Moreover, there continues to be no obvious correlation between the value of the CNs and either the simulation technique or the solvent, although there does seem to be a slight trend for greater interaction between the amine hydrogens and the methanol oxygens in EPSR compared with MD. The CNs in Table 6.7, however, emphasise that H_{N2} is the most solvated amine hydrogen in EPSR and H_{N1} is the most solvated amine hydrogen in MD: $n_{H_{N2}}^{O_w}+n_{H_{N2}}^{O_m} = 0.72 + 0.63 = 1.35$ and $n_{H_{N1}}^{O_w}+n_{H_{N1}}^{O_m} = 0.66 + 0.54 = 1.20$ for EPSR and MD, respectively. The same analysis for the remaining amine hydrogens yields 1.19 and 1.06 for H_N and H_{N1}, respectively, in EPSR and 1.14 and 1.16 for H_N and H_{N2}, respectively, in MD. Summing these numbers and averaging over the number of interactions reveals that in EPSR, on average, 0.60 solvent oxygens are bound to any given amine hydrogen. The equivalent figure in MD is 0.58. Minor differences in the degree to which each amine hydrogen is solvated in EPSR and MD

aside, again, a consistent picture of the nature of this solvation is maintained between the two simulations, as evidenced by the RDFs in Fig. 6.9.

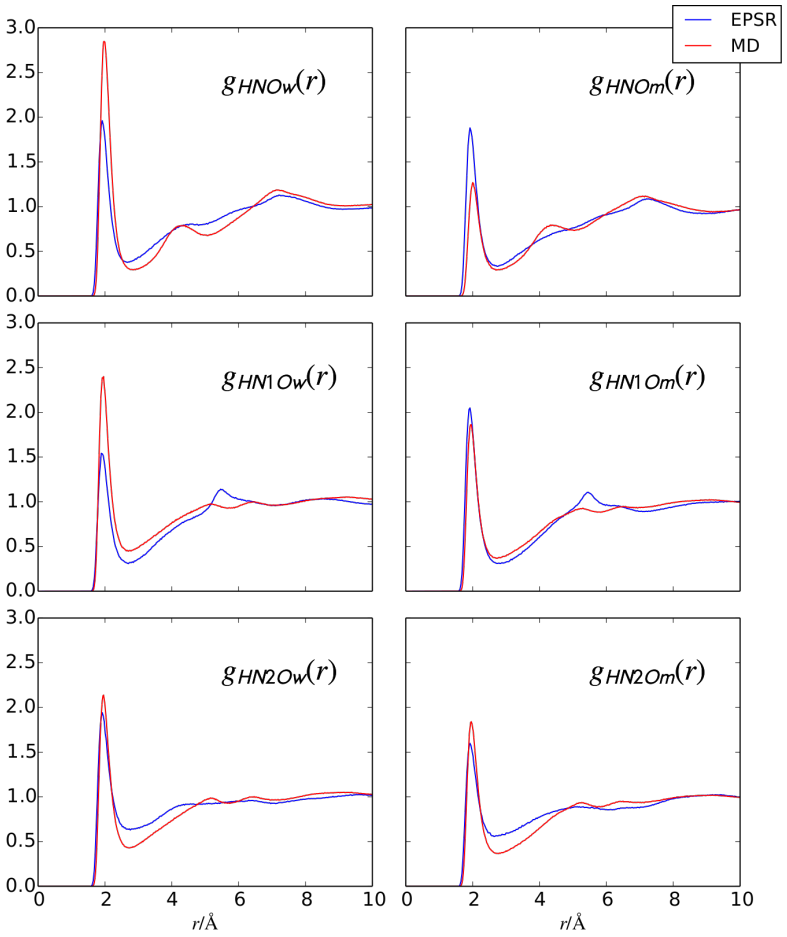


Figure 6.9: RDFs for the interactions of the amine hydrogens of clonidine (H_N, H_{N1} and H_{N2}) with solvent oxygens (O_w and O_m) extracted from EPSR (blue line) and MD (red line).

n_α^β	**EPSR**	**MD**	r**/Å**
$n_{H_N}^{O_w}$	0.63	0.74	2.88
$n_{H_{N1}}^{O_w}$	0.47	0.66	2.76
$n_{H_{N2}}^{O_w}$	0.72	0.61	2.76
$n_{H_N}^{O_m}$	0.56	0.40	2.79
$n_{H_{N1}}^{O_m}$	0.59	0.54	2.79
$n_{H_{N2}}^{O_m}$	0.63	0.55	2.79

Table 6.7: CNs of the clonidine/methanol/water RDFs shown in Fig. 6.9.

Fig. 6.10 shows the RDFs for interactions of the benzene ring of clonidine (B) with water and methanol atoms extracted from EPSR and MD. CNs corresponding to Fig. 6.10 are shown in Table 6.8. What is immediately clear from Fig. 6.10 is that, like for alprazo-

lam in methanol/water solutions in Section 5.3, water molecules are not highly ordered around the benzene ring of clonidine, as they were seen to be in Section 3.3 (indole) and to a lesser extent in Section 4.3 (cocaine). The absence of highly ordered waters around the benzene ring is particularly obvious in $g_{\mathrm{BH_w}}(r)$, where the RDF extracted from EPSR especially is almost flat. Similarly, although peaks in the RDFs corresponding to interactions between the benzene ring of clonidine and methanol are more discernible, particularly the benzene-CH_3 interaction, than interactions of the benzene ring with water, they are at distances greater than ∼4 Å. The position and breadth of these peaks also indicates that methanol is not highly ordered relative to the benzene ring, despite broad peaks, such as in $g_{\mathrm{BH_{mo}}}(r)$, suggesting a preferred distance of methanol from the benzene ring. A broad peak can be seen in $g_{\mathrm{BO_m}}(r)$ at distances of ∼4.6 Å and ∼5.0 Å for EPSR and MD, respectively. There is also a peak in $g_{\mathrm{BC_m}}(r)$ at ∼4.9 Å and ∼4.3 Å for EPSR and MD, respectively. The peak in $g_{\mathrm{BC_m}}(r)$ is particularly clear in MD, again, pointing towards a preferred distance of methanol around the benzene ring of clonidine. Opposed to what was seen for indole in methanol/water solutions in Section 3.3, the -CH_3 group is found closer to the benzene ring (∼4.3 Å) than the hydroxyl group (∼5.0 Å) in MD. In EPSR the reverse is true, with hydroxyl groups being found closer to the benzene ring (∼4.6 Å) than the -CH_3 group (∼4.9 Å), as was the case in Section 3.3.

Differences in the interactions of the benzene ring of clonidine and the benzene ring of indole (Section 3.3) and cocaine (Section 4.3) with solvent molecules may be due to the number of hydrophilic groups of the solute and the concentration of the solute in the samples measured by diffraction. The fewer highly localised solvent molecules around the benzene ring of clonidine compared with the benzene ring of indole may be because clonidine has more direct electrostatic hydrogen bonding opportunities than indole, which only has a single amine group. Although clonidine has a similar number of hydrophilic groups as cocaine, which comprises an amine group and two carbonyl groups, clonidine was measured at a lower solute concentration (1:200) than cocaine (1:65). Measurement at a lower concentration might mean that waters do not have to 'resort' to binding to the benzene ring of clonidine as much as they did for cocaine in aqueous solution.

Following the results presented for indole in methanol/water solutions in Section 3.3, Table 6.8 indicates that water has a greater propensity than methanol to interact with the benzene ring in EPSR compared with MD. The CNs for the interactions between the benzene ring of clonidine and methanol atoms illustrate that, although the average number of O_m and H_{mo} atoms around the benzene ring are very similar (within 2%) in EPSR and MD, there is a noticeable increase in the number of methyl groups around the benzene ring. Explicitly, $n_B^{C_m}$ increases by ∼11% in MD compared with EPSR.

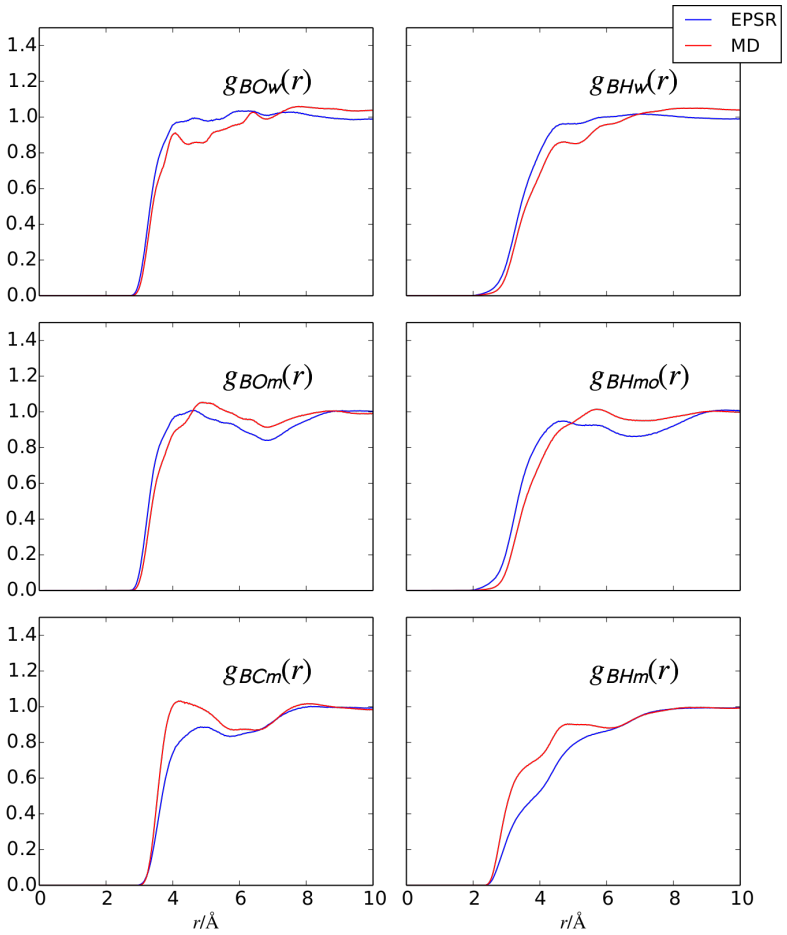


Figure 6.10: RDFs for the interactions of the benzene ring of clonidine (B - the weighted average of C, C_1, C_B and C_{B1}) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR (blue line) and MD (red line).

RDFs analogous to those shown in Fig. 6.10 are shown in Fig. 6.11 for the interactions of the imidazole ring of clonidine (P) with water and methanol atoms. Unlike in Fig. 6.10, RDFs showing the interactions between the imidazole ring and water atoms are not flat. $g_{PO_w}(r)$ shows peaks located at ∼3.6 Å and ∼4.8 Å for both EPSR and MD. These peaks are likely to be a result of the previously discussed interactions of N_1 and N_2 with water oxygens. Similarly, peaks at ∼3.5 Å and ∼4.7 Å in $g_{PO_m}(r)$ for EPSR

n_{α}^{β}	**EPSR**	**MD**	r/Å
$n_B^{O_w}$	3.17	2.67	4.80
$n_B^{H_w}$	8.12	6.75	5.16
$n_B^{O_m}$	11.69	11.62	6.84
$n_B^{H_{mo}}$	11.76	11.57	6.90
$n_B^{C_m}$	5.56	6.17	5.80
$n_B^{H_m}$	17.35	19.76	5.94

Table 6.8: CNs of the clonidine/methanol/water RDFs shown in Fig. 6.10.

and MD, respectively, are most likely a result of the interactions between the amine groups that form the imidazole ring and methanol oxygens. The absence of any sharp peaks in $g_{PH_w}(r)$ and $g_{PH_{mo}}(r)$, which show broad peaks a distances greater than $\sim$4 Å, suggests a lack of any highly ordered interactions between the imidazole ring and solvent molecules in either simulation.

CNs corresponding to the RDFs shown in Fig. 6.11 are included in Table 6.9. Coupled with the RDFs having the same peak positions for EPSR and MD in Fig. 6.11, these CNs underline the similar interactions between the imidazole ring and solvent atoms in EPSR and MD. However, as with the solvation of the benzene ring, there appears to be a slight trend for a greater number of imidazole-hydroxyl group interactions in EPSR compared with MD and a greater number of imidazole-methyl group interactions in MD compared with EPSR.

n_{α}^{β}	**EPSR**	**MD**	r/Å
$n_P^{O_w}$	1.95	2.06	4.14
$n_P^{H_w}$	21.50	20.72	6.54
$n_P^{O_m}$	2.07	1.82	4.14
$n_P^{H_{mo}}$	9.34	8.77	6.39
$n_P^{C_m}$	3.84	4.06	5.25
$n_P^{H_m}$	15.16	16.21	5.70

Table 6.9: CNs of the clonidine/methanol/water RDFs shown in Fig. 6.10.

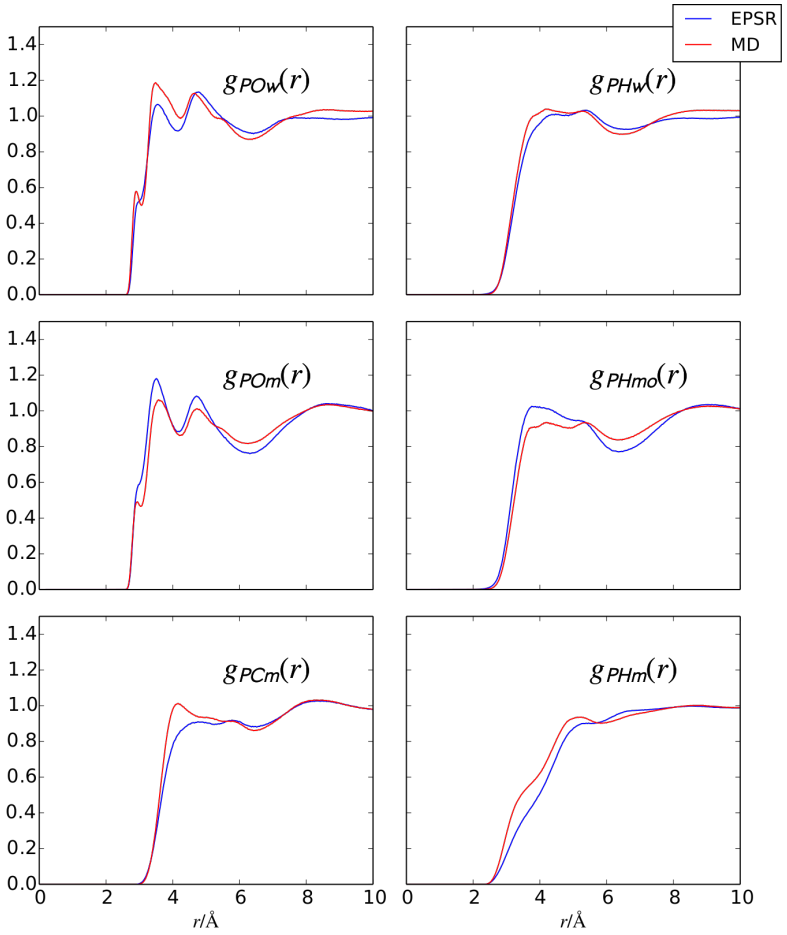


Figure 6.11: RDFs for the interactions of the imidazole ring of clonidine (P - the weighted average of C_2, C_3, N_1 and N_2) with water (H_w and O_w) and methanol atoms (H_m, O_m, C_m and H_{mo}) extracted from EPSR (blue line) and MD (red line).

6.3.4.2 Spatial Density Maps

SDMs generated from EPSR which show the distribution of nearest neighbour water and methanol oxygens relative to the amine groups of clonidine are shown in Fig. 6.12. In Fig. 6.12, regions of high water and methanol density are apparent above ($\pm\mathbf{z}$ direction) each of the amine groups of clonidine. Considering the hydration of these amine groups, N appears to be the most densely hydrated. The density of water around N_1 and N_2 is similar, but the density of water around each of these amine groups is less than the density of water around N. In contrast, the amine groups around which the highest density of methanol is seen in Fig. 6.12 are N and N_1, which have a similar density of methanol molecules around them as water molecules around N. N seems to be the most densely solvated amine group when considering solvation by either by water or methanol. Overall, however, the solvation of these groups is similar, with the position

of the regions of highest solvent density in Fig. 6.12 consistent and the range of the maximum solvent density narrow (2.5-3.3 $Å^{-3}$). That these groups have areas of a high density of solvent molecules in the SDMs in Fig. 6.12 is consistent with Fig. 6.8 and Fig. 6.9, where sharp, well defined peaks are seen in the RDFs for amine-solvent interactions. The amine groups that have the highest density of solvent molecules around them in the SDMs in Fig. 6.12 do not correspond to the amine groups in Table 6.6 or Table 6.7 that have the largest CNs, as there is not necessarily a link between the average number of molecules solvating a group and the density of these molecules in the SDMs. However, the similarity of the density around each of the amine groups shown by the SDMs in Fig. 6.12 is consistent with the small degree of variation between the CNs in Table 6.6 or Table 6.7 for the solvation of amine nitrogens and hydrogens, respectively, by water and methanol molecules.

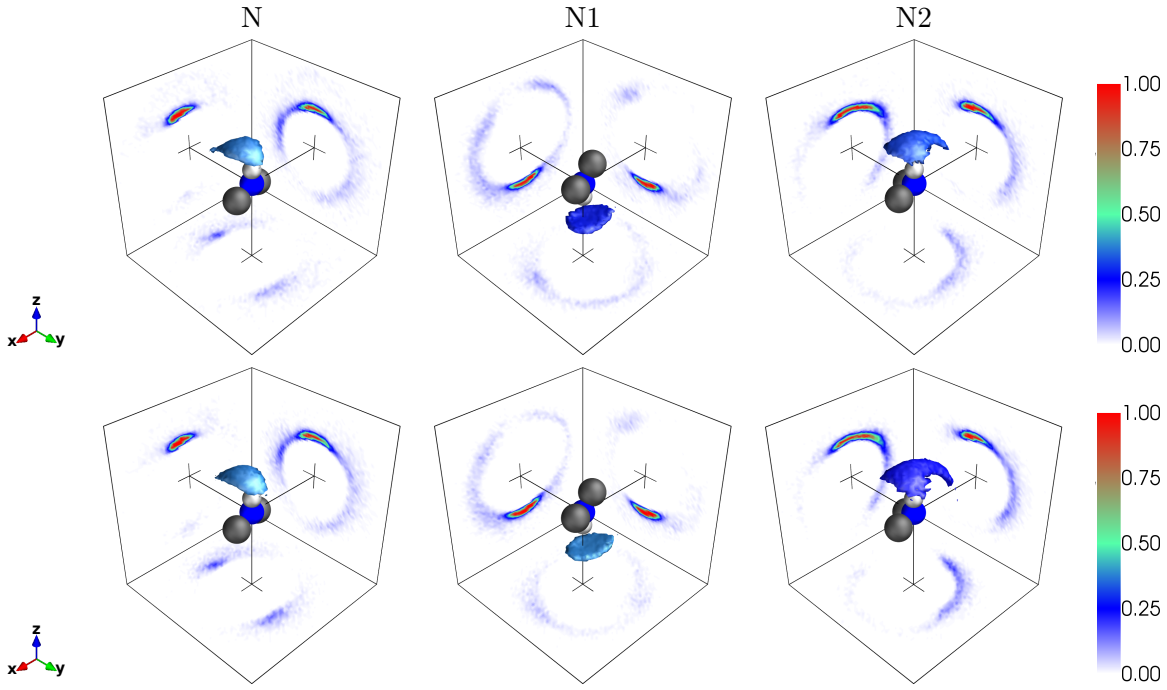


Figure 6.12: SDMs showing the distribution of nearest neighbour water (top) and methanol molecules (bottom) relative to the amine groups of clonidine generated from EPSR.

Analogous SDMs generated from the MD trajectory are displayed in Fig. 6.13, where the location of the solvation of the amine groups follows the pattern seen in Fig. 6.12 for the solvation of the amine groups in EPSR. The similarity in the location of the solvation of the amine groups in the SDMs generated from the EPSR and MD simulations is particularly highlighted by the cut-throughs. The maximum density of solvent oxygens

around the amine groups for the SDMs generated from the MD trajectory falls in the same range as the density of solvent oxygens around the amine groups in the SDMs generated from EPSR (2.5-3.3 Å^{-3}). The exception is the hydration of N, where the maximum density of water oxygens is 8.5 Å^{-3}, more the double the solvation density for any other SDM shown in either Fig. 6.12 or Fig. 6.13. N_1 is the second most densely hydrated amine group, having a slightly higher density of water oxygens around it in MD than in EPSR. The density around the nitrogens in the SDMs generated from the MD trajectory corresponding to nitrogen-methanol interactions is similar to the density around these groups in Fig. 6.12 for EPSR. The cut-throughs of the SDM showing the distribution of methanol oxygens relative to N show regions of density, additonal to the density in the **z** direction, in the **yz** plane. As was the case in Fig. 6.12, N seems to be the group that has the greatest propensity to interact with solvent molecules.

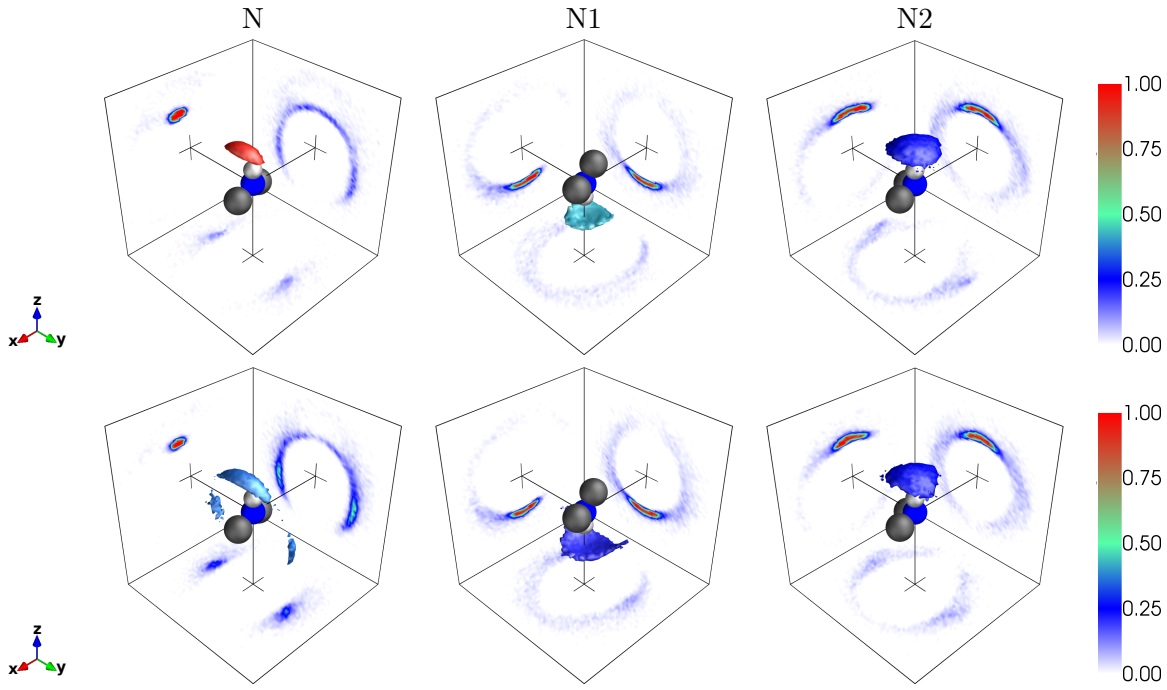


Figure 6.13: SDMs showing the distribution of nearest neighbour water (top) and methanol molecules (bottom) relative to the amine groups of clonidine generated from MD.

A more thorough understanding of the solvation of the amine groups of clonidine can be accrued by computation of the percentage of nearest neighbour solvent molecules that are either water or methanol. This information is shown in Table 6.10 for both EPSR and MD. For EPSR Table 6.10 indicates no clear preference for interaction with nearest neighbour waters by either N or N_2, as may be anticipated from Fig. 6.12. N_1

shows a preference for interacting with nearest neighbour methanols, as ∼13% more methanols comprise the nearest neighbour solvent molecules of this group than waters. This is consistent with Fig. 6.12, where a higher density of methanol is seen around N_1 compared with water. For MD, like for EPSR, the percentage of nearest neighbour solvent molecules around N_2 that are either methanol or water is very similar (within 4%). In contrast to EPSR, there is a greater percentage of nearest neighbour solvent molecules of N_1 that are water rather than methanol for MD. The biggest difference in the percentage of nearest neighbour solvent molecules is around N for which, though the percentage of water molecules compared with methanol molecules is within 1% for EPSR, in MD ∼75% more of the nearest neighbour solvent molecules are water than methanol. This finding tallies with Fig. 6.13, where, as mentioned, the SDM corresponding to the distribution of nearest neighbour methanols around N shows a greater density of solvent molecules than any of the other SDMs shown in either Fig. 6.12 or Fig. 6.13.

Simulation →	**EPSR**		**MD**	
Amine Group ↓	**Water**	**Methanol**	**Water**	**Methanol**
N	49.7%	50.3%	63.7%	36.3%
N1	43.5%	56.5%	54.1%	45.9%
N2	52.1%	47.9%	51.7%	48.3%

Table 6.10: The percentage of nearest neighbour solvent molecules relative to the amine groups of clonidine which are either water or methanol.

Given the findings of Section 3.3, the distribution of solvent molecules relative to the benzene ring of clonidine was also analysed *via* SDMs for EPSR and MD (Fig. 6.14). Each of the SDMs in Fig. 6.14 shows a reduced density of solvent molecules around the benzene ring compared with indole in methanol/water solutions (Section 3.3) and cocaine in aqueous solution (Section 4.3). This is to be expected given the absence of any clear peaks in the corresponding RDFs shown in Fig. 6.10 and Fig. 6.11. There are no clear differences in the SDMs related to either the different solvents or the simulation techniques. Despite the low density of solvent around the benzene ring in Fig. 6.14, the preferred location of solvation above and below the plane of the ring that was seen in Section 3.3, Section 4.3 and Section 5.3 for indole, cocaine and alparazolam, respectively is also seen here, although, as discussed, this density is not a consequence of hydrogen bonding between the benzene ring and solvent molecules.

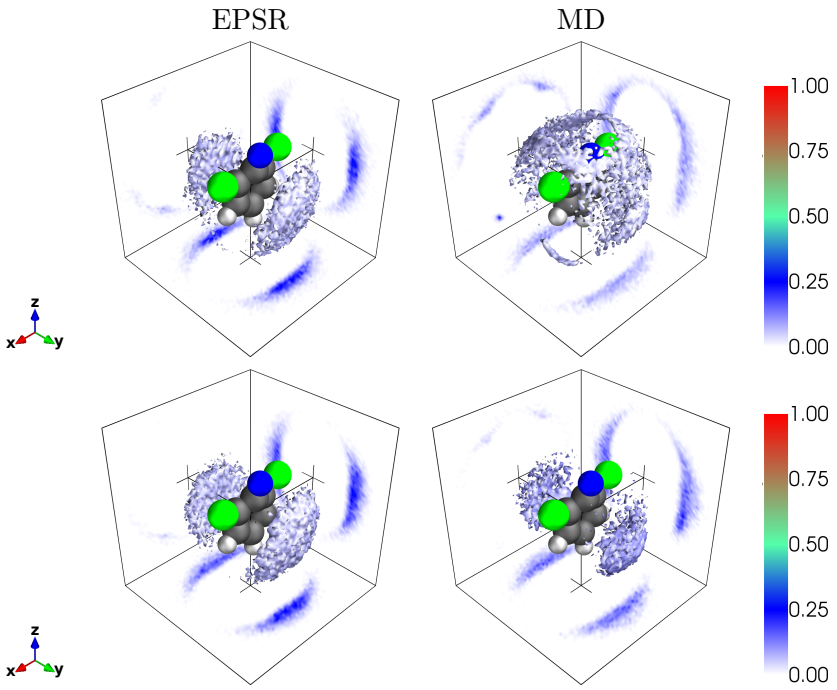


Figure 6.14: SDMs showing the distribution of nearest neighbour water (top) and methanol molecules (bottom) relative to the benzene ring of clonidine generated from EPSR (left) and MD (right).

6.3.4.3 Whole Molecule Analysis

Fig. 6.15 shows the results of WMA for both water and methanol within 4 Å of clonidine for EPSR (Fig. 6.15 (a)) and MD (Fig. 6.15 (b)). A number of clonidine molecules taken from each simulation have been included in Fig. 6.15 to underline the range of conformations adopted by the solute in EPSR and MD.

WMA immediately highlights that the part of the clonidine molecule preferentially solvated by both water and methanol is the nitrogen between the benzene and imidazole rings (N). In Fig. 6.15 (a), a distinct cap of density (water and methanol) around N is visible. That this is the preferred hydration site in EPSR agrees with the SDMs presented in Fig. 6.12, where the highest density of nearest neighbour water molecules was seen around N. Similarly, the highest density of nearest neighbour methanol molecules was seen around N and N_1 for EPSR in Fig. 6.12. Turning to Fig. 6.15 (b), a cap of density around N is also apparent for WMA of the MD trajectory. This finding is in agreement with the SDMs in Fig. 6.13, where this amine group showed the highest density of nearest neighbour water and methanol molecules in the SDMs. Fig. 6.15 (b)

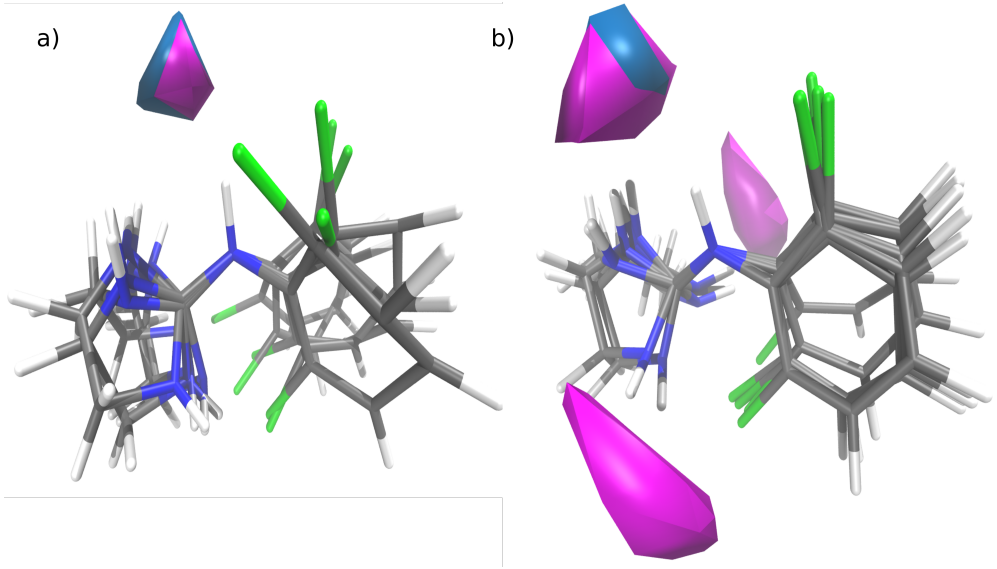


Figure 6.15: WMA for water (blue) and methanol (magenta) within 4 Å of clonidine for EPSR (a) and MD (b). The isocontour surface is cut to include the top 30% of solvent molecules. Different clonidine molecules taken from each of the simulations are shown at the origin to display the range of conformations adopted by clonidine.

shows regions of a high density of methanol molecules either side of the amine group. Inspection of Fig. 6.13 again reveals the density in Fig. 6.15 (b) to be consistent with SDMs, as areas of high density of nearest neighbour methanol molecules relative to N are seen in the cut-through of the **yz** plane that are not seen in the corresponding plot for nearest neighbour waters.

As mentioned, Fig. 6.15 also allows a qualitative assessment of the conformation adopted by the clonidine molecules in solution. It appears that the molecules adopt a wider range of conformations in EPSR than in MD. This is particularly true of the benzene ring of clonidine, in agreement with the results presented in Fig. 6.5 showing that the benzene and imidazole rings of around 10% of clonidine molecules in the EPSR simulation box are mutually perpendicular.

6.3.5 NMR

6.3.5.1 Assignment of Hydrochloride and Freebase

Fig. 6.16 shows NMR spectra of clonidine hydrochloride and the freebase form of clonidine dissolved in H_2O and deuterated THF, respectively, at a concentration of 1 cloni-

dine: 200 solvent. H_3, H_{B1} and H_B were identified by their intensity and peak splitting. The remaining peak(s) in each spectra, which correspond to the amine protons, were assigned using 2D NMR spectra (Appendix B.3). Coupling to H_3 was used to assign the peak in each spectra corresponding to H_{N1} and H_{N2}, which have the same chemical shift. Therefore, the most downfield peak in the the spectrum of clonidine hydrochloride dissolved in H_2O corresponds to H_N. Comparison between the spectra shown in Fig. 6.16 suggests that it is H_N, the amine proton between the imidazole and benzene rings, which distinguishes the protonated form of clonidine from its deprotonated form.

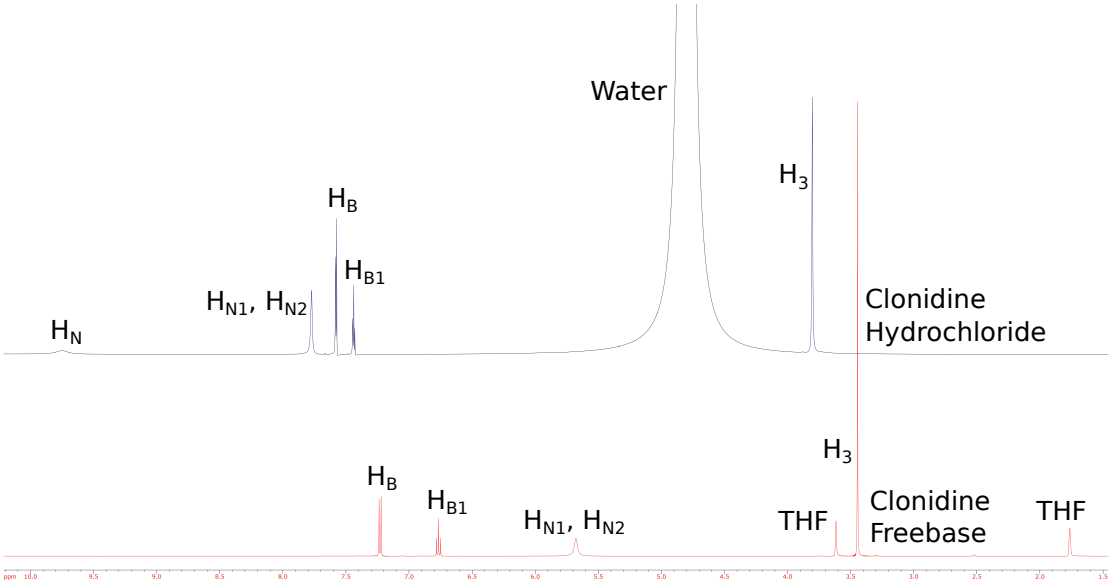


Figure 6.16: ^{1}H NMR spectra of clonidine hydrochloride (blue) and the freebase form of clonidine (red) dissolved at at concentration of 1 clonidine: 200 solvent in H_2O and deuterated THF, respectively.

6.3.5.2 The Effect of Temperature

NMR spectra of clonidine hydrochloride dissolved in methanol/water solutions recorded at 298 K, the temperature at which diffraction data were collected, and 310 K (Fig. 6.17) show no change in either the chemical shifts of the peaks or in the splitting of the peaks over this temperature range. Fig. 6.17 suggests that the structure of clonidine measured at 298 K is no different to the structure of clonidine at physiologically relevant temperatures.

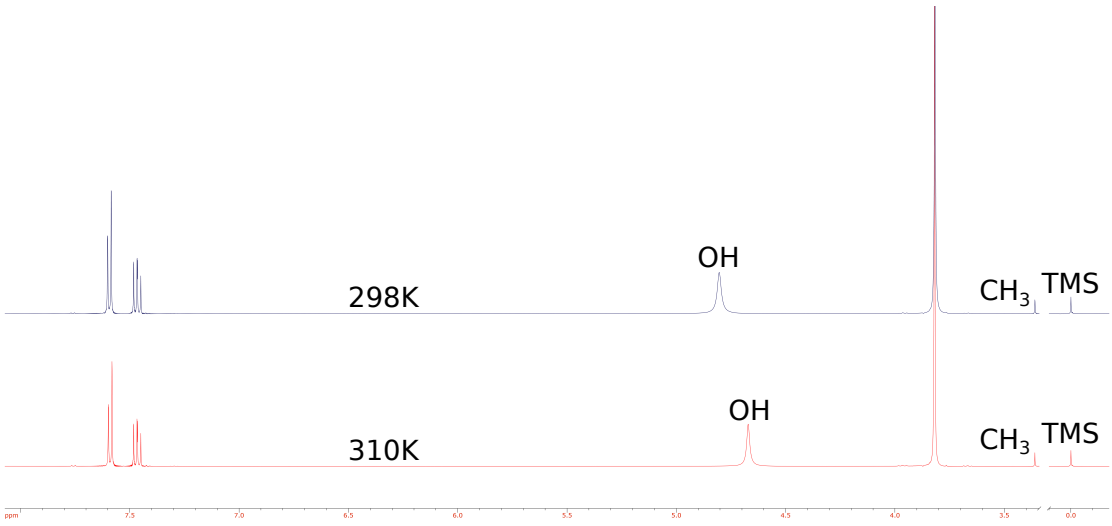


Figure 6.17: ^{1}H NMR spectra of clonidine hydrochloride dissolved in CD_3OD/D_2O at a concentration of 1 clonidine :200 water at 298 K (blue) and 310 K (red).

6.4 Conclusions

As for indole and alprazolam in methanol/water solutions in Section 3.3 and Section 5.3, respectively, there appears to be be no enhancement of either the methanol or water structure on mixing the two liquids. RDFs show that the majority of the interactions between clonidine molecules and the solvent occur *via* hydrogen bonding with the NH groups of the molecule. There are some minor differences between the solvation of the amine groups shown by the RDFs and CNs extracted from EPSR and MD, but peaks in the RDFs are located at the same position for amine-water and amine-methanol interactions for both EPSR and MD, indicating that the nature of the solvation of the amine groups is preserved in both simulations. The imidazole ring does not interact with the solvent atoms other than *via* its amine groups. The pattern of solvation seen in the RDFs and the SDMs for the imidazole ring is consistent with studies of its hydration by *ab initio* molecular orbital calculations, where water molecules primarily interact with the nitrogen atoms of the ring.[3] SDMs show that it is around the amine group located between the benzene and imidazole rings of clonidine where the greatest density of nearest neighbour solvent molecules are found in both EPSR and MD. This conclusion is supported by WMA, which shows that the top 30% of solvent molecules are located around this amine group for both EPSR and MD.

Although studies of the freebase form of clonidine have shown the imidazole and

benzene ring of clonidine not to be mutually perpendicular,[289] the results presented here indicate that around 10% of protonated clonidine molecules in solution exist with benzene and imidazole rings perpendicular relative to each other. However, the majority of clonidine molecules in both simulations have their rings only tilted (<40°) relative to each other. This finding agrees with the results of more recent computational approaches that show that the ortho chlorine group prevents the two rings of clonidine hydrochloride from adopting a coplanar conformation.[9, 226, 227] This is likely the conformation of clonidine *in vivo*, as its pK_a (8.1) indicates that it is protonated.[115, 184] The conformation adopted by clonidine hydrochloride is relevant to its solvation, as a conformational energy minimum in which the two rings are coplanar allows the imidazole ring to be conjugated with the phenyl moiety,[47] influencing the solvent accessibility of the molecule. If the rings were completely coplanar, not slightly tilted relative to each other, the solvation of the amine group joining the rings might be reduced compared with the results presented here.

A basic nitrogen is generally accepted to be essential for bonding to a carboxylic group at adrenoceptors,[177] a result confirmed by genetically engineered receptor proteins. [147] Comparison between the mechanism by which agonists bind to adrenoceptors, however, is not straightforward. For phenylethylamine and imidazoline-type ligands, for example, there are various modes of interaction with the carboxylate that are possible due to the presence one, two or three nitrogen atoms.[177] Computational studies have proposed that clonidine binds to adrenoceptors *via* an imidazole nitrogen.[43] However, the results presented in this Chapter, where the amine group joining the rings of clonidine experiences the greatest solvation, may indicate that it is actually this nitrogen where clonidine binds. The nitrogen joining the benzene and imidazole rings of clonidine has also been shown to be involved in the binding of clonidine in the active site of bovine plasma amine oxidase, an enzyme which catalyzes the oxidation of primary amines.[106] Furthermore, despite confirming the existence of imidazoline receptive sites, studies of the clonidine binding sites in the intestinal membranes of seawater eels suggested that clonidine binding sites do not recognise the imidazole ring itself.[141]

Results suggesting that the nitrogen joining the benzene and imidazole rings is more influential in the mode of action of clonidine than the nitrogen atoms of the imidazole

ring are particularly pertinent in light of Section 5.3, where it was found that the position of bridging water molecules in the crystal structure of the alprazolam-BRD4 complex were in good agreement with the most densely solvated region seen by WMA. The results of this investigation support observations that arguments against the imidazoline hypothesis outweigh supporting evidence and that the sympathoinhibitory effects of clonidine-like drugs are best explained by activation of α_2-adrenoceptors.[279]

Chapter 7

Conclusions and Further Work

7.1 Conclusions

By using neutron diffraction with isotopic substitution in conjunction with Empirical Potential Structure Refinement (EPSR), a Monte Carlo-based technique which uses diffraction data as a direct constraint, along with Molecular Dynamics (MD) and nuclear magnetic resonance (NMR) the structures of indole, cocaine hydrochloride, alprazolam and clonidine hydrochloride have been determined in solution on the atomic scale and related to their biological function and behaviour *in vivo*. Besides their individually biologically interesting functions, these molecules were chosen for investigation because of shared structural features, indole acting as a simple starting point. Cocaine, alprazolam and clonidine were chosen, in particular, because each is able to cross the blood-brain barrier (BBB).

For each system investigated, there was no perturbation to the solvent structure compared with water or methanol. For indole, alprazolam and clonidine, each of which was measured by diffraction in methanol/water solutions, this finding agrees with previous studies showing that the underlying water structure of methanol/water mixtures is not greatly affected by the presence of methanol molecules, and methanol and water molecules in methanol/water solutions persist in structures similar to the ones found in the pure liquids.[6, 56, 59, 91, 94, 286] Solvent-solvent hydrogen bonding varied $<10\%$

across the different systems. This small variation can be explained by differences in the concentration at which each molecule was dissolved. Indole, which was measured at a greater concentration (1 indole: 59 solvent) than either alprazolam (1 alprazolam: 250 solvent) or clonidine (1 clonidine: 200 solvent) in methanol/water solutions, showed reduced solvent-solvent hydrogen bonding as a result of more frequent solute-solvent interactions compared with alprazolam and clonidine. The inherently low flux of neutron sources necessitates measuring as highly concentrated samples as possible, meaning that there is always a compromise between consistency of concentration and the quality of the diffraction data when measuring a series of molecules in solution.

Differences between the solvation of each of the molecules was seen when comparing the results from EPSR and MD in terms of the degree to which different molecular motifs were solvated, such as the amine groups of clonidine. However, the nature of solvation, for instance, the region which had the greatest density of nearest neighbour solvent molecules, was maintained across both simulation techniques for each molecule. This conclusion applies to cocaine in aqueous solution, where in MD, despite ∼800 times more cocaine molecules, on average, associating with each other *via* interactions between the amine group and the carbomethoxy carbonyl oxygen, the preferential hydration of the molecule was maintained compared with EPSR.

In addition to ubiquitous tools, such as radial distribution functions (RDFs) and coordination numbers (CNs), standardly used to characterise the structure of amorphous materials, novel analysis methods were successfully developed and implemented to characterise the solvation of molecules in solution. These methods include Spatial Density Maps (SDMs) and Whole Molecule Analysis (WMA). The advantage of SDMs compared with Spatial Density Functions (SDFs),[181, 262] which are spherical harmonic expansions of the RDFs, is that SDFs require an axis of symmetry to be defined by the user,[268] which can potentially result in loss of spatial distribution information. An instance where use of SDFs may have been disadvantageous is in the analysis of the hydration of the carbomethoxy oxygen and the amine hydrogen of cocaine, where the offset position of the water density in the SDMs contributed to the hypothesis of internal hydrogen bond formation. Although more computationally expensive because of the requirement of many coordinate files, SDMs can facilitate a more nuanced un-

derstanding of molecular solvation. Reliable quantification any error in RDFs, CNs or SDMs hinges on the development of better understood methods to correct for inelastic neutron scattering, as it is difficult to characterise the error introduced by the current method. This aspect of the analysis presented in this thesis is an important limitation of the field of neutron diffraction.

Investigation of the structure of indole, the functional group of the amino acid tryptophan, in methanol/water solutions showed that the solution structure of indole is dominated by electrostatic interactions as opposed to hydrophobic-hydrophobic interactions between the methyl group of methanol and indole. The results of this investigation imply that electrostatic forces are more powerful in dictating amphiphilic solvation of a mostly hydrophobic molecule, suggesting that it is possible that tryptophan is stabilised at the membrane bilayer interface *via* electrostatic interactions between the benzene ring of indole and its amphiphilic environment. Understanding the solvation of structural features of indole, such as its benzene ring, also acted as a starting point for understanding the solvation of more complex molecules in solution, as the benzene ring is a common drug motif.

Measurement of cocaine hydrochloride dissolved in aqueous solution at a concentration of 1 cocaine: 65 water led to the hypothesis that cocaine forms a water mediated internal hydrogen bond in solution. The combination of intramolecular hydrogen bonding and the apparent affinity of water molecules to bind to the atoms associated with this bond offer an explanation for the mechanism by which cocaine is able to cross the BBB that does not require the molecule to be deprotonated, as is sometimes thought.[316] The internal hydrogen bond of cocaine results in hydrophilic groups of the molecule effectively being shielded from the surrounding environment, meaning that cocaine becomes sufficiently lipophilic to cross the BBB. Internal hydrogen bonding has been recognised as important in increasing the lipophilicity of a drug by shielding hydrogen bonding motifs.[153]

The solvation of alprazolam, a benzodiazepine, in methanol/water solutions showed that the part of the molecule around which the greatest density of water molecules was located, the nitrogen atoms of the triazole ring, coincided with the location of bridging

water molecules in the crystal structure of the BRD4-alprazolam complex. This finding contradicts the predictions of computational protocols developed to predict the location of bridging water molecules,[82, 234] which are known to play a key role in protein-ligand interactions.[160, 280] These computational protocols predict a limited role for the ligand in determining the positions of bridging water molecules. However, the results presented in this thesis suggest that hydration studies offer a pathway to determining how benzodiazepines might bind to BET proteins, in particular, which perform transcription regulatory function and regulate transcription of several oncogenes.[78, 189, 242].

Investigation of the structure of clonidine hydrochloride in methanol/water solutions indicated that the majority of clonidine molecules (∼90% in EPSR) exist with their ring structures slightly tilted with respect to each other but close to being coplanar, in agreement with studies showing that the ortho chlorine group of the benzene ring of clonidine prevents the benzene ring from being in the same plane as the imidazole ring.[9, 226] WMA showed that the part of clonidine which experienced the densest hydration was the amine group linking the benzene and imidazole rings. Although some studies have attributed the sympathoinhibitory effects of clonidine to imidazoline receptor action,[83, 92, 215] the results of this investigation suggest that clonidine interacts with receptors *via* the nitrogen atom between the benzene and imidazole rings. This preferential hydration is consistent with the mode of action of clonidine being due to activation of adrenoceptors, excitatory receptors involved in the control of vital bodily functions.[37, 122, 159, 176, 276].

Considering each of the systems together hints at a relationship between the number of hydrophilic groups and the number of hydrogen bonds formed between the benzene ring of each molecule and the solvent. The total number of hydrogen bonds a molecule forms in solution has been reported as a predictor of BBB penetration, with increasing hydrogen bonding decreasing BBB penetration.[198, 201, 205] Indole, a largely hydrophobic molecule, formed hydrogen bonds between the benzene ring and both water molecules and methanol -OH groups. Cocaine, which has a greater number of hydrophilic groups compared with indole (i.e. 2 carbonyl groups and an amine group) showed a distinct reduction in the number of ordered waters around the benzene ring compared with indole, despite being dissolved at approximately the same concentration as indole in

solution (1 indole: 59 solvent vs. 1 cocaine: 65 solvent). Alprazolam has 4 nitrogen atoms, and almost no ordered water molecules were found around the benzene ring. In fact, methanol -CH_3 groups were more highly localised around the benzene ring of alprazolam -OH groups, the converse of what was seen for indole in methanol/water solutions. Similarly, clonidine, which has 3 amine groups, showed no ordered waters around the benzene ring. Although the reduced concentration of alprazolam and clonidine in solution compared with indole could provide an explanation for the reduction in hydrogen bonding, a complete absence is indicative of more fundamental molecular differences. Differences in the interactions between the benzene ring of each of the molecules investigated and water molecules are encapsulated by Fig. 7.1.

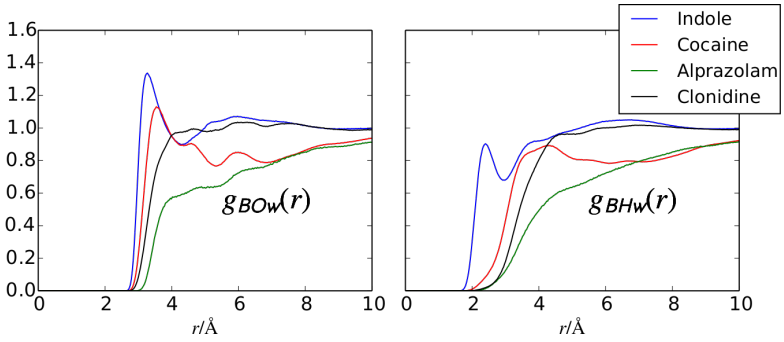


Figure 7.1: RDFs for the interactions between the benzene (B) ring of indole (blue line), cocaine·HCl (red line), alprazolam (green line) and clonidine·HCl (black line) with water atoms (H_w and O_w) extracted from EPSR.

In addition to the factors thought to influence the ability of a molecule to cross the BBB discussed in Section 1.3, the polar surface area (PSA), which is defined as the sum of surfaces of polar atoms in a molecule, has been successfully applied to the prediction of BBB penetration.[64] Various protocols have been reported to calculate the PSA, differing in their definition of polar atoms. Typically only N and O atoms are regarded as polar.[198] Perhaps hydrogen bonding to the benzene ring by solvent molecules is related to the partial charge distribution across a molecule resulting from the presence of polar atoms.[162] Hence, an increase in the number of polar atoms results in a decrease in hydrogen bonds between the benzene ring and solvent molecules, as there are more polar atoms to which -OH motifs can bind. The sum of the number of O and N atoms in a molecule has been recognised as being highly correlated with PSA, though the total hydrogen bonding area has been suggested as a better measure of BBB penetration.[201] The importance of electronegative atoms, such as O and N, in medicinal chemistry is

underlined by the use of fluorine to increase the lipophilicity of molecules.[221, 253] Regardless of the relationship between the number of polar atoms and the ability of the benzene ring of a molecule to hydrogen bond to solvent molecules, O and N atoms are clearly important in determining the ability of a molecule to cross the BBB and in determining interactions between a molecule and its solvent environment.

The work presented in this thesis underlines the importance of determining the structure of pharmaceuticals, and biomolecules in general, in physiologically relevant environments (i.e. in solution). By doing so conformational and structural information, such as the location of water molecules either involved in or indicative of biological function, may be obtained that is not available from solely crystallographic studies. This thesis provides insight into how water molecules may influence the ability of molecules to cross the BBB and has shown how hydration in solution can reveal the location of bridging water molecules and shed light on the interactions between ligands and receptors. Although much more work is needed to develop a consistent methodology to predict BBB penetration, determining the solution structure of pharmaceuticals is requisite in order to develop a comprehensive understanding of their behaviour *in vivo*.

7.2 Further Work

To further understand how the conformation adopted by cocaine in solution is influenced by the solvent in which it is dissolved, it would be useful to measure cocaine hydrochloride dissolved in methanol using neutron diffraction. In this way, the role played by water in determining the conformation adopted by the molecule could be further elucidated by comparison between measurements made of cocaine dissolved in water and cocaine dissolved in methanol. Results of diffraction measurements of cocaine dissolved in methanol would complement NMR measurements of cocaine in solution presented in this thesis, as these NMR spectra implied a concentration related chemical shift for cocaine dissolved in water that was not apparent over the same concentration range (55 mM-275 mM) for cocaine dissolved in methanol.

Measurement of the freebase form of cocaine in solution would aid understanding of

the role of the amine proton in determining the conformation of cocaine in solution, as the hypothesis presented here relies on internal hydrogen bond formation between this proton and the the carbomethoxy carbonyl oxygen. Unfortunately, the low solubility of the freebase form of cocaine in solution precludes neutron diffraction as a suitable technique for measurement of its structure in aqueous solution. However, the structure of the freebase form of cocaine in methanol/water solutions could be determined by neutron diffraction or MD. Measurement of the structure of the freebase form of cocaine in methanol/water solutions, however, would require complementary measurements of the hydrochloride in methanol/water solutions for direct comparison.

Similarly, measurement of the freebase form of clonidine in solution could offer insight into the preferential solvation of the molecule in the absence of the proton attached to the nitrogen atom which joins the benzene and imidazole rings. The most densely solvated region of clonidine in its deprotonated form could influence conclusions regarding whether the mechanism of action of clonidine is related to its interaction with adrenoceptors or imidazole receptors.

A more detailed understanding of the behaviour of drugs able to cross the BBB may be gained by investigating the structure of cocaine, alprazolam and clonidine in solution with a lipid. Although the methanol/water solutions in which indole, alprazolam and clonidine were measured provides a simple amphiphilic environment, the amphiphilic environment of of biological membranes is clearly more complex than this perfunctory binary system. A suitable lipid for such an investigation is C_3-PC, because of its short, simple, previously determined structure.[76] The simple structure of C_3-PC would facilitate extraction of interactomic distances using neutron diffraction with isotopic substitution. Furthermore, phosphatidylcholine lipids like C_3-PC are physiologically relevant as they are common in most mammalian cells, making up 28-33% of human brain endothelial cells.[38] It would be necessary to undertake any NMR measurements of pharmaceuticals in solution with C_3-PC at a concentration below the critical micelle concentration, which is the concentration at which the lipid molecules begin to form micelles, or else any behaviour witnessed could not be confidently attributed to the presence of C_3-PC molecules but rather may be a consequence of micelle formation. The critical micelle concentration of a lipid can be verified using dynamic light scatter-

ing, which measures time dependent fluctuations in scattered light incident on a sample. These fluctuations are directly related to the rate of diffusion of the molecule through the solvent, which is in turn related to the radii of the particles (individual molecules or aggregated groups of them). In a similar vein, more ambitious plans include performing analogous measurements of drugs able to cross the BBB in solution with ceramide. Ceramide is a lipid found in the brain and provides a more realistic model environment than C_3-PC. It is a particularly appropriate choice as its structure has already been measured using neutron diffraction.[90]

The interaction of drugs with biological membranes could be investigated by means of supported lipid bilayers. Supported lipid bilayers are lipid bilayers formed on atomically smooth solid supports such as silicon or mica by thermally fusing lipid vesicles onto these solid supports.[40, 150, 208] If a supported lipid bilayer were exposed to an aqueous solution of a drug, interaction between this model lipid bilayer and the drug could be measured. Consequently, understanding of the interplay between molecular motifs of the bilayer and the drug could be enhanced. Although a variety of techniques are used to study supported lipid bilayers, such as scanning electron microscopy, atomic force microscopy and Fourier transform infrared resonance,[208] the most appropriate technique to investigate the proposed system would be neutron reflectivity because of the advantages of H/D contrast, lack physical damage to biological materials and measurement on an absolute scale associated with neutron scattering.[296]

To expand the conclusions of this thesis the solution structure of other BBB penetrating molecules, for example diazepam, oxazepam and prilocaine (Fig. 7.1) could be determined.[a] Selection of additional molecules for investigation would be made based on shared structural features, (e.g. benzene ring) and if each of these molecules were measured at the same concentration in the same solvent more reliable conclusions could be deduced regarding the influence of the total number of O and N atoms on benzene ring solvation, expanding on ongoing work relating the PSA to BBB penetration. Both diazepam and oxazepam are, like alprazolam, benzodiazepines. Thus investigation of their

[a]Diazepam was measured by myself in methanol/water solutions at a concentration of 1 diazepam: 100 methanol: 100 water during the course of this thesis, but analysis of the data has not yet been undertaken. I have also helped to measure the structure of prilocaine hydrochloride and the freebase form of prilocaine in water and methanol/water solutions, respectively.

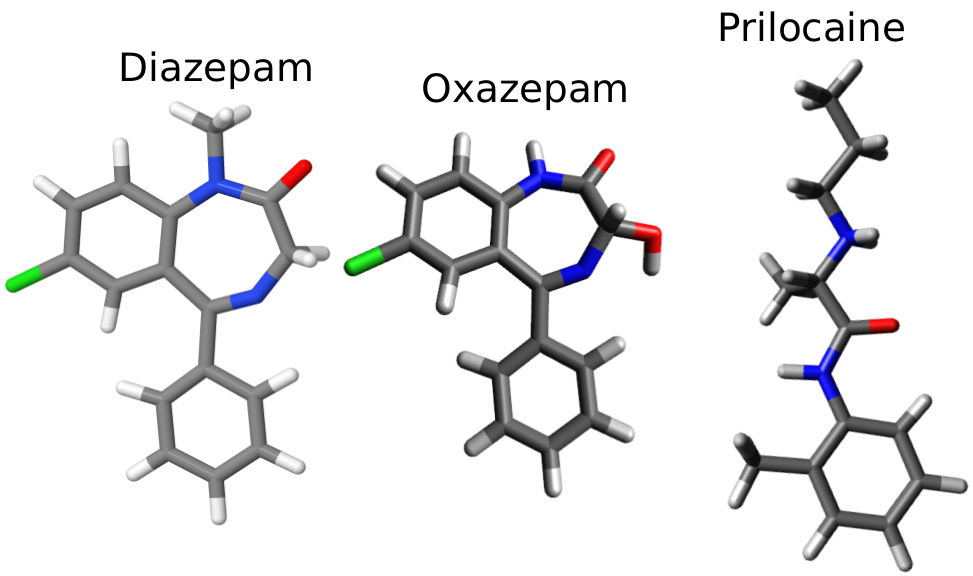


Figure 7.2: The molecular structures of diazepam, oxazepam and prilocaine.

structure in solution would also allow further study of the relationship between the location of bridging water molecules in the crytsal structures of BET protein-benzodiazepine complexes and the preferential solvation of benzodiazepines in methanol/water solutions. Prilocaine is a local anesthetic widely used in dentistry which binds to the voltage-gated sodium channels of axons, blocking the propagation of the nerve impulse.[35]

The thesis could also be extended by determining the location of bound water molecules for some of the molecules studied here in solution. This work would enable comparisions to be made between the location of these bound waters and the regions of the molecule shown to be preferentially hydrated. By supersaturating an aqueous solution of a given compound and subsequently allowing molecules to recrystallise, crystals incorporating bound waters could be produced. These crystals could then be measured to determine the crystal structure of the molecule and the location of bound water molecules. Neutron diffraction would again be the ideal technique to measure the structure of these crystals because of its sensitivity to hydrogen.

Appendix A

Self and Distinct Scattering Terms

h The auto correlation function (ACF) [a] of $\beta(r)$ is the given by

$$\mathrm{ACF}[\beta(r)] = \iiint_V \sum_{\mathrm{j}=1}^{N} \beta_{\mathrm{j}}(\mathbf{r}' - \mathbf{r}_{\mathrm{j}})^* \sum_{\mathrm{k}=1}^{N} \beta_{\mathrm{k}}(\mathbf{r}' - \mathbf{r}_{\mathrm{k}} + \mathbf{r}) \mathrm{d}^3\mathbf{r}'. \tag{A.1}$$

This expression can then be decomposed into two terms:

$$\mathrm{ACF}[\beta(r)] = g'(\mathbf{r}) + g(\mathbf{r}), \tag{A.2}$$

where

$$g'(\mathbf{r}) = \sum_{\mathrm{j}=1}^{N} \iiint_V \beta_{\mathrm{j}}(\mathbf{r}' - \mathbf{r}_{\mathrm{j}})^* \beta_{\mathrm{j}}(\mathbf{r}' - \mathbf{r}_{\mathrm{j}} + \mathbf{r}) \mathrm{d}^3\mathbf{r}' \tag{A.3}$$

and

$$g(\mathbf{r}) = \sum_{\mathrm{j}\neq\mathrm{k}}^{N} \iiint_V \beta_{\mathrm{j}}(\mathbf{r}' - \mathbf{r}_{\mathrm{j}})^* \beta_{\mathrm{k}}(\mathbf{r}' - \mathbf{r}_{\mathrm{k}} + \mathbf{r}) \mathrm{d}^3\mathbf{r}'. \tag{A.4}$$

[a] The ACF provides information on repeating patterns in a set of data. Mathematically it is defined as $\mathrm{ACF}(x) = \int_{-\infty}^{\infty} f(t)^* f(x+t)\mathrm{d}t$. Largest at the origin, the ACF of a function has a linear relationship to the modulus of its FT: $\mathrm{ACF}(x) = \int_{-\infty}^{\infty} |\mathscr{F}[f(x)]|^2 e^{\mathrm{i}kx}]\mathrm{d}k = \int_{-\infty}^{\infty} |F(k)|^2 e^{\mathrm{i}kx}]\mathrm{d}k$. In this instance, by inspection of Eq. 2.28, $\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\mathrm{el}} \propto \iiint_V \mathrm{ACF}[\beta(\mathbf{r})] e^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r}$.

$g'(\mathbf{r})$ has a maximum at the origin ($\mathrm{r_j}$=0) and results from the correlation of the $\mathrm{j^{th}}$ atom with itself. The decay of $g'(\mathbf{r})$, in the case of neutrons, is determined by the thermal motion of the atoms. $g(\mathbf{r})$ conveys information about the sample structure, depending on the type and spatial distribution of the atoms in the sample. Effectively, $g(\mathbf{r})$ is the sum of the integrals representing the correlation between the $\mathrm{j^{th}}$ and $\mathrm{k^{th}}$ atoms. It has a maximum value of

$$\iiint \beta_{\mathrm{j}}(r')^{*}\beta_{\mathrm{k}}(r') \tag{A.5}$$

when $\mathbf{r} = \mathbf{r}_{\mathrm{k}} - \mathbf{r}_{\mathrm{j}}$. $g(\mathbf{r})$ is symmetric, as an equivalent contribution will be made when the subscripts j and k are exchanged, such that

$$g(-\mathbf{r}) = g(\mathbf{r}). \tag{A.6}$$

As $g_2(\mathbf{r})$ has to be spherically symmetric for amorphous materials, since constituent molecules are found in all orientations:

$$g(\mathbf{r}) = g(r). \tag{A.7}$$

Peaks in $g(\mathbf{r})$ will be more visible closer to $r = 0$, because of fixed intramolecular and nearest neighbour intermolecular distances. As r increases, lack of long-range order leads to a uniform continuum of interatomic distances. It is usually helpful to use a normalised version of $g(\mathbf{r})$, which tends to unity for large r:

$$G(r) = \frac{g(r)}{\langle\beta\rangle^2 V}. \tag{A.8}$$

This is called a pair distribution function, where $\langle\beta\rangle$ is the average value of the SLD and V is the illuminated volume.

The linear relationship between the differential cross section and the ACF of the SLD function means that the decomposition of the SLD into $g'(\mathbf{r})$ and $g(\mathbf{r})$ must be mirrored in the scattering data. Consider the FT of Eq. 2.29:

$$\iiint_{V} \beta(\mathbf{r}) e^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}} \mathrm{d}^3\mathbf{r} = \sum_{\mathrm{j}=1}^{N} f_{\mathrm{j}}(\mathbf{Q}) e^{\mathrm{i}\mathbf{Q}\cdot\mathbf{r}_{\mathrm{j}}}, \tag{A.9}$$

where the j^{th} atom is located at $\mathbf{r}_j$ and, by comparison with Eq. 2.30 and Eq. 2.31, has a form factor

$$f_j(\mathbf{Q}) = \iiint \beta_j(\mathbf{r}) e^{i\mathbf{Q}\cdot\mathbf{r}} d^3\mathbf{r}. \tag{A.10}$$

As the differential cross section is given by the modulus squared of the FT of $\beta(\mathbf{r})$ (see Eq. 2.28),[271]

$$\left(\frac{d\sigma}{d\Omega}\right)_{el} \propto \left| \sum_{j=1}^{N} f_j(\mathbf{Q}) e^{i\mathbf{Q}\cdot\mathbf{r_j}} \right|^2 . \tag{A.11}$$

When j=k (i.e. the atom interacts with itself), the product of the sum of the right hand side of Eq. A.9 with its complex conjugate leads to

$$\left| \sum_{j=1}^{N} f_j(\mathbf{Q}) \right|^2 . \tag{A.12}$$

When j>k, the product of the sum of the right hand side of Eq. A.9 with its complex conjugate leads to

$$2 \sum_{j>k} Re\{ f_j(\mathbf{Q}) f_k(\mathbf{Q})^* e^{i\mathbf{Q}\cdot(\mathbf{r_j}-\mathbf{r_k})} \}. \tag{A.13}$$

Hence, Eq. A.11 can be expressed as[266]

$$\left(\frac{d\sigma}{d\Omega}\right)_{el} \propto \overbrace{F'(Q)}^{\text{Self Scattering}} + \underbrace{F(Q)}_{\text{Distinct Scattering}} , \tag{A.14}$$

where

$$F'(Q) = \sum_{j=1}^{N} |f_j(Q)|^2 \qquad F(Q) = 2 \sum_{j>k} Re\{ f_j(\mathbf{Q}) f_k(\mathbf{Q})^* e^{i\mathbf{Q}\cdot(\mathbf{r_j}-\mathbf{r_k})} \}. \tag{A.15}$$

Appendix B

2D NMR Spectra

B.1 Indole

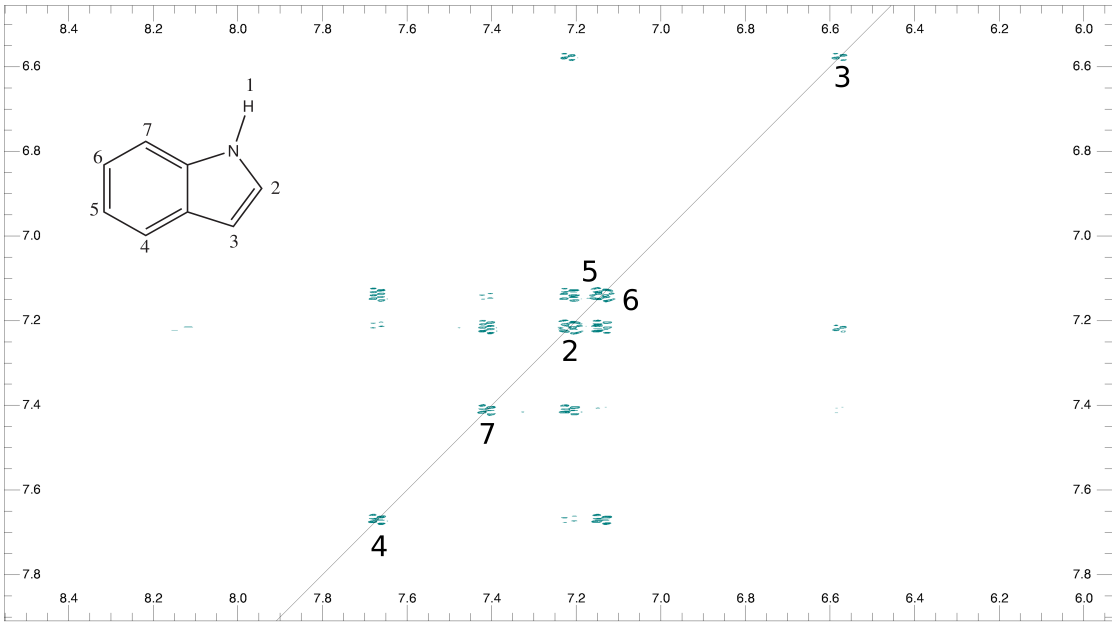


Figure B.1: A ^{1}H-^{1}H COSY spectrum of indole dissolved in $CDCl_3$ at a concentration of 1 indole: 150 $CDCl_3$. The corresponding 1D spectrum is shown in Fig. 3.25.

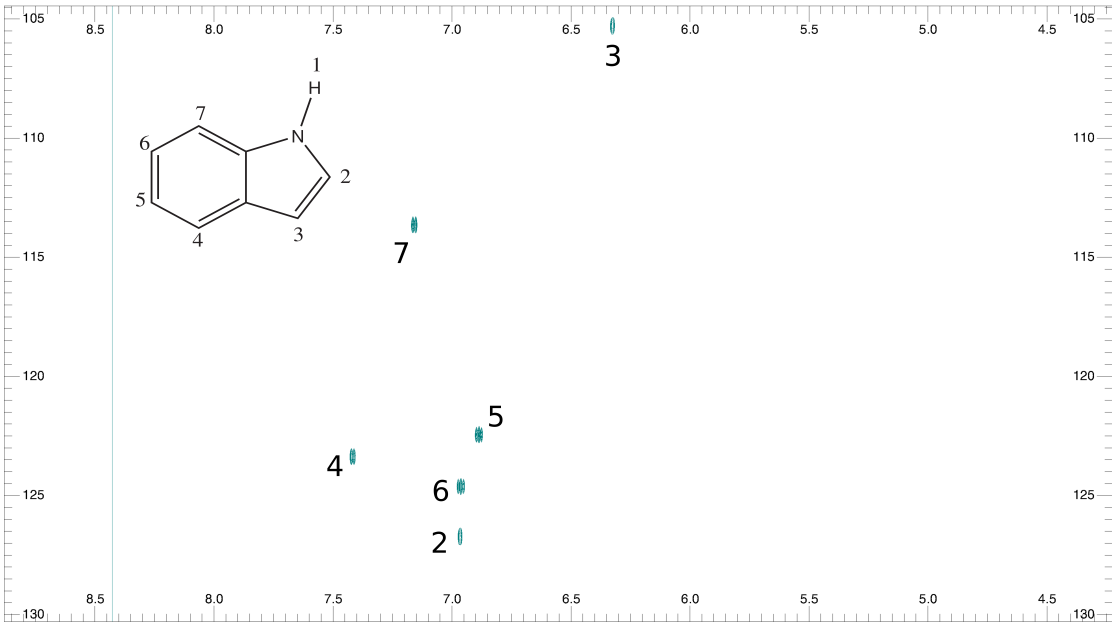


Figure B.2: A ^{13}C HSQC spectrum of indole dissolved in $CDCl_3$ at a concentration of 1 indole: 150 $CDCl_3$. The corresponding 1D spectrum is shown in Fig. 3.25.

B.2 Cocaine

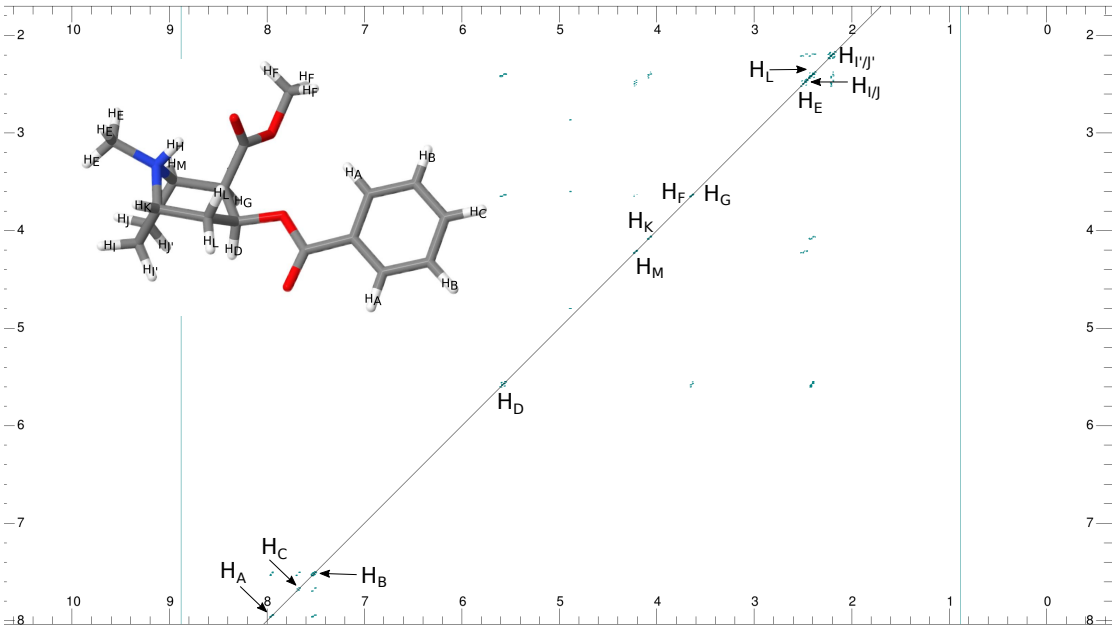


Figure B.3: A ^{1}H-^{1}H COSY spectrum of cocaine hydrochloride dissolved in D_2O at a concentration of 1 cocaine: 150 D_2O. The corresponding 1D spectrum is shown in Fig. 4.24.

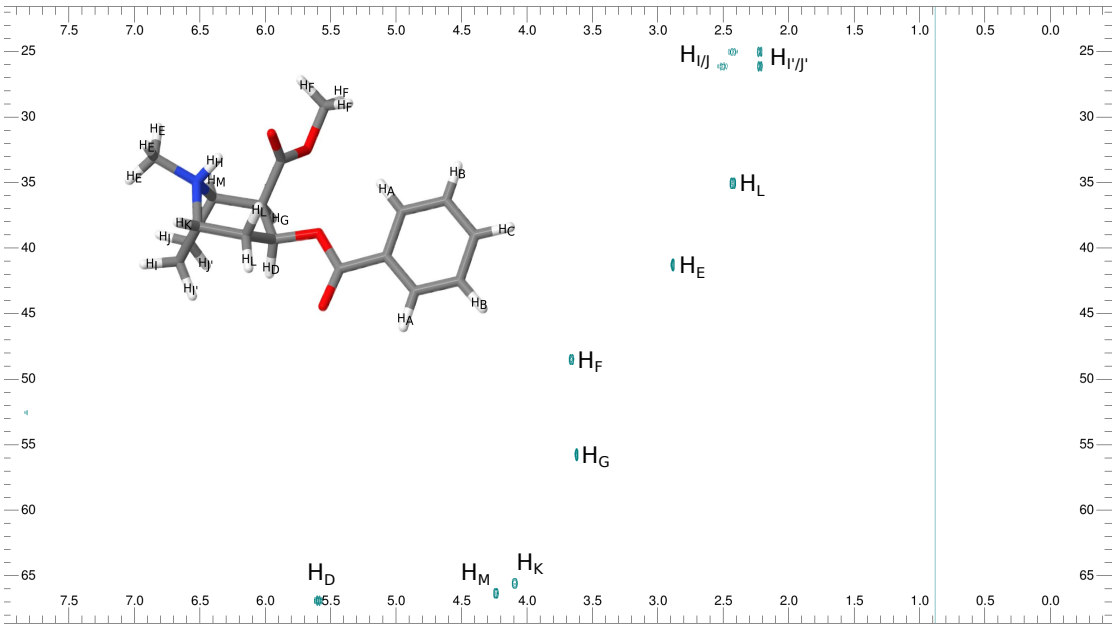


Figure B.4: A ^{13}C HSQC spectrum of cocaine hydrochloride dissolved in D_2O at a concentration of 1 cocaine: 150 D_2O. The corresponding 1D spectrum is shown in Fig. 4.24. The carbon frequency is centred on 40ppm.

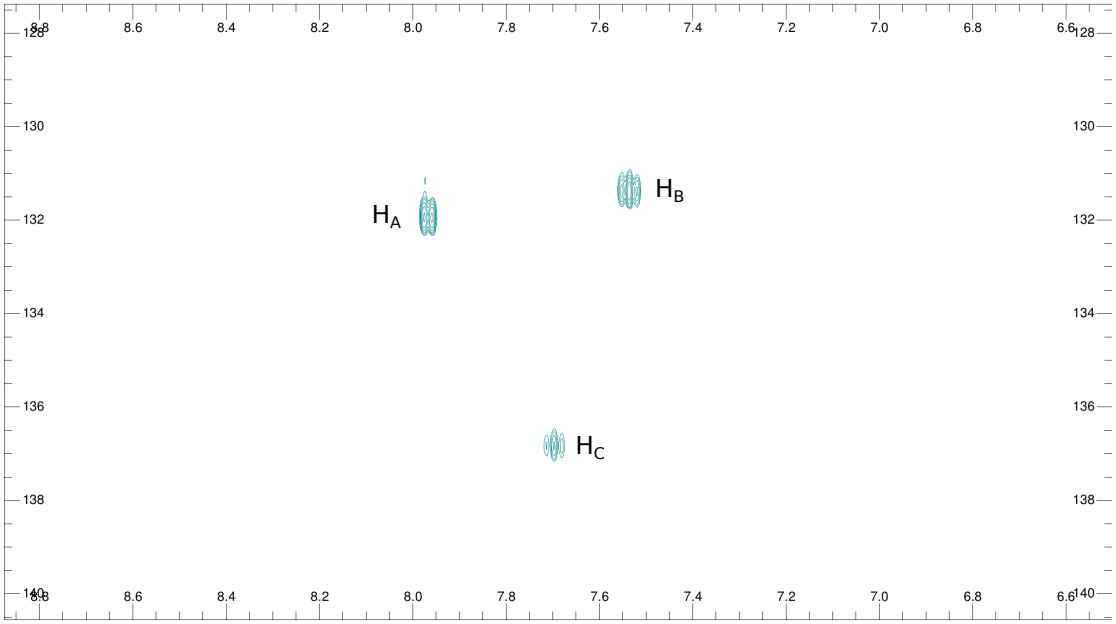


Figure B.5: A ^{13}C HSQC spectrum of cocaine hydrochloride dissolved in D_2O at a concentration of 1 cocaine: 150 D_2O. The corresponding 1D spectrum is shown in Fig. 4.24. The carbon frequency is centred on 120ppm.

B.3 Clonidine

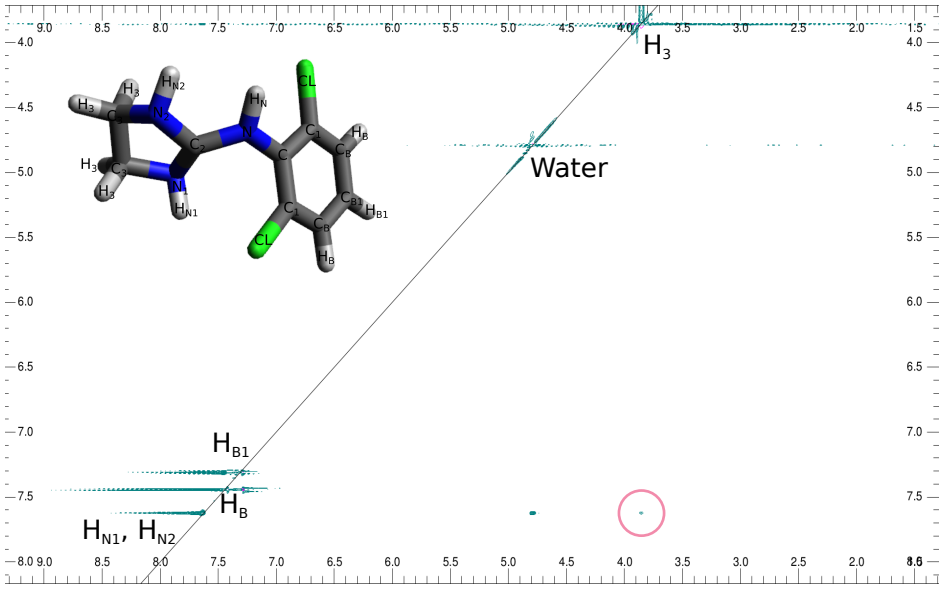


Figure B.6: A ^{1}H-^{1}H TOCSY spectrum of clonidine hydrochloride dissolved in H_2O at a concentration of 1 clonidine: 200 H_2O. The corresponding 1D spectrum is shown in Fig. 6.16. Ringed in pink is the coupling between H_{N1}/H_{N2} and H_3.

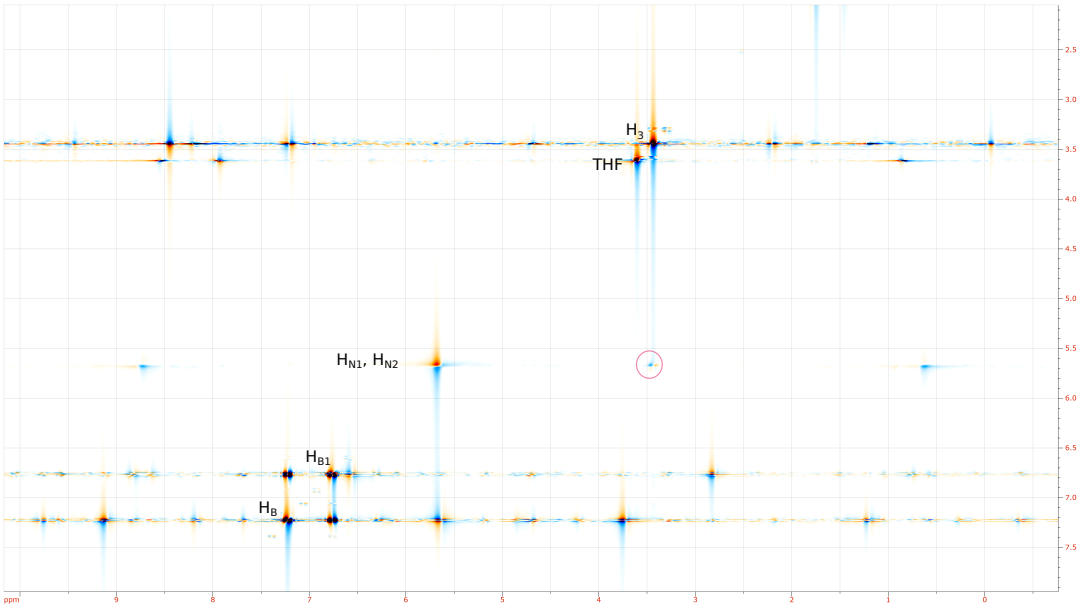


Figure B.7: A ^{1}H-^{1}H COSY spectrum of the freebase form of clonidine dissolved in deuterated THF at a concentration of 1 clonidine: 200 THF. The corresponding 1D spectrum is shown in Fig. 6.16. Ringed in pink is the coupling between H_{N1}/H_{N2} and H_3.

References

[1] M. H. Abraham, K. Takács-Novák, and R. C. Mitchell. On the Partition of Ampholytes: Application to Blood-brain Distribution. *J. Pharm. Sci.*, 86(3):310–315, 1997. 11

[2] H. Adams, F. J. Carver, C. A. Hunter, and N. J. Osborne. Amide-aromatic Hydrogen Bonds in Host-guest Recognition. *Chem. Commun.*, pages 2529–2530, 1996. 100

[3] G. Alagona, C. Ghio, P. Nagy, K. Simon, and Gábor G. Náray-Szabo. Comparative Study of Imidazole Hydration: *Ab initio* and Electrostatic Calculations vs. Cambridge Structural Database Analysis. *J. Comp. Chem.*, 11(9):1038–1046, 1990. 199

[4] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.*, 7:207–218, 2016. 167

[5] M. P. Allen. *Introduction to Molecular Dynamics Simulation.* John von Neumann Institute for Computing, Warwick, UK, 2004. 5, 13

[6] S. Allison, J. Fox, R. Hargreaves, and S. Bates. Clustering and Microimmiscibility in Alcohol-Water Mixtures: Evidence from Molecular Dynamics Simulations. *Phys. Rev. B*, 71, 2005. 74, 89, 152, 182, 203

[7] L. M. Amzel. Structure-based Drug Design. *Current Opinion in Biotech.*, 9(4):366–369, 1998. 2

[8] A. C. Anderson. The Process of Structure-based Drug Design. *Chem. & Bio.*, 10(9):787–797, 2003. 1

[9] F. Avbelji and D. Hadži. Potential Energy Functions and the Role of the Conformational Entropy of Clonidine-like Imidazolidines in Determining their Affinity for α-adrenergic Receptors. *Mol. Pharm.*, 27(4):466–470, 1985. 200, 206

[10] S. Baer. Structural Correlations and the 'Structural Diffusion' Model for Liquids. *Physica A: Stat. Mech. and its Applications*, 87(3):569–583, 1977. 4

[11] P. Ballabh, A. Braun, and M. Nedergaard. The Blood-brain Barrier: An Overview Structure, Regulation, and Clinical Implications. *Neurobio. of Disease*, 16, 2004. 9

[12] W. A. Banks. Characteristics of Compounds that Cross the Blood-brain Barrier. *BMC Neurology*, 9(1):1–5, 2009. 10

[13] J. A. Barker and D. Henderson. Perturbation Theory and Equation of State for Fluids: The Square-Well Potential. *J. Chem. Phys.*, 47(8):2856–2861, 1967. 4

[14] M. G. J. Baud, E. Lin-Shiao, M. Zengerle, C. Tallant, and A. Ciulli. New Synthetic Routes to Triazolo-benzodiazepine Analogues: Expanding the Scope of the Bump-and-Hole Approach for Selective Bromo and Extra-terminal (BET) Bromodomain Inhibition. *J. Med. Chem.*, 59(4):1492–1500, 2016. 142

[15] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A Well-behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.*, 97:10269–10280, 1993. 67

[16] C.J. Benmore and A.K. Soper. The SANDALS Manual: A Guide to Performing Experiments on the Small Angle Neutron Diffractometer for Amorphous and Liquid Samples at ISIS. Version 1.0. *RAL Technical Reports*, 1998. 28, 30, 31

[17] S. M. Bennington, A. C. Hannon, and S. E. Rogers. *Neutron Training Course*. Science and Technology Facilities Council, 2015. 21, 29

[18] H. J. C. Berendsen, J. R. Grigera, and T. P Straatsma. The Missing Term in Effective Pair Potentials. *J. Chem. Phys*, 91:6269–6271, 1987. 64, 67, 107, 109, 145, 147, 175, 177

[19] H. J. C. Berendsen, J. P. M. Postma, W. F. von Gunsteren, A. DiNola, and J. R. Haak. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.*, 81:3684–3690, 1984. 68, 110

[20] C. A. S. Bergström, U. Norinder, K. Luthman, and P. Artursson. Experimental and Computational Screening Models for Prediction of Aqueous Drug Solubility. *Pharm. Research*, 19(2):182–188, 2002. 2

[21] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bournel. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 2000. 142

[22] J. D. Bernal. An Attempt at a Molecular Theory of Liquid Structure. *Trans. Faraday Soc.*, 33:27–40, 1937. 4

[23] J. D. Bernal. The Bakerian Lecture, 1962. The Structure of Liquids. *Proc. of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 280(1382):299–322, 1964. 4

[24] S. Biswal, U. Sahoo, S. Sethy, H. K. S. Kumar, and M. Banjeree. Indole: The Molecule of Diverse Biological Activities. *Asian J. Pharm. and Clin. Research*, 5:1–6, 2012. 59

[25] G. Blaser, J. M. Sanderson, and M. R. Wilson. Free-energy Relationships for the Interactions of Tryptophan with Phosphocholines. *Organic & Biomol. Chem.*, 7(24):5119–5128, 2009. 60

[26] R. Bloch, P. Bousquet, J. Feldman, and J. Schwartz. The Mechanism of Action of Clonidine. *Biochem. Pharm.*, 23:707–711, 1974. 172

[27] T. L. Blundell and S. Patel. High-throughput X-ray Crystallography for Drug Discovery. *Current Opinion in Pharm.*, 4(5):490–496, 2004. 2

[28] D. Boda and D. Henderson. The Effects of Deviations from Lorentz-Berthelot Rules on the Properties of a Simple Mixture. *Mol. Phys.*, 106(20):2367–2370, 2008. 41

[29] M. Born and H. S. Green. A General Kinetic Theory of Liquids. I. The Molecular Distribution Functions. *Proc. of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 188(1012):10–18, 1946. 3

[30] D. T. Bowron and J. L. Finney. Association and Dissociation of an Aqueous Amphiphile at Elevated Temperatures. *J. Phys. Chem. B*, 111(33):9838–9852, 2007. 100

[31] M. A. Braga, M. F. Martini, M. Pickholz, F. Yokaichiya, M. K. D. Franco, L. F. Cabeça, V. A. Guilherme, C. M. G. Silva, C. E. G. Limia, and E. de Paula. Clonidine Complexation with Hydroxypropyl-beta-cyclodextrin: From Physicochemical Characterization to *in vivo* Adjuvant Effect in Local Anesthesia. *J. Pharm. and Biomed. Analysis*, 119:27–36, 2016. 172

[32] S. Busch, C. D. Bruce, C. Redfield, C. D. Lorenz, and S. E. McLain. Water mediation Essential to Nucleation of β-Turn Formation in Peptide Folding Motifs. *Angew. Chem. - Int. Ed.*, 49:13091–13095, 2013. 77, 122

[33] S. Busch, C. D. Lorenz, J. Taylor, L. C. Pardo, and S. E. McLain. Short-range Interactions of Concentrated Proline in Aqueous Solution. *J. Phys. Chem. B*, 118(49):14267–14277, 2014. 50, 55, 117

[34] S. Busch, L. C. Pardo, W. B. O'Dell, C. D. Bruce, C. D. Lorenz, and S. E. McLain. On the Structure of Water and Chloride Ion Interactions with a Peptide Backbone in Solution. *Phys. Chem. Chem. Phys.*, 15:21023–21033, 2013. 47, 50, 55, 121

[35] L. F. Cabeça, M. Pickholz, E. de Paula, and A. J. Marsaioli. Liposome-Prilocaine Interaction Mapping Evaluated through STD NMR and Molecular Dynamics Simulations. *J. Phys. Chem. B*, 113(8):2365–2370, 2009. 211

[36] S. K. Callear, A. J. Johnston, S.E. McLain, and S. Imberti. Conformation and Interactions of Dopamine Hydrochloride in Solution. *J. Chem. Phys.*, 142(1), 2015. 77

[37] B. C. Calzada and A. Aleixandre de Artiñano. Alpha-adrenoceptor Subtypes. *Pharm. Research*, 44(3):195–208, 2001. 171, 206

[38] S. D. Campbell, K. J. Regina, and E. D. Kharasch. Significance of Lipid Composition in a Blood-brain Barrier-Mimetic PAMPA Assay. *J. Biomol. Screen.*, 19:437–444, 2013. 9, 209

[39] O. Carugo and D. Bordo. How Many Water Molecules Can be Detected by Protein Crystallography? *Acta Crystallogr. Sect. D Biol. Crystallogr*, 55:479–483, 1999. 170

[40] E. T. Castellana and P. S. Cremer. Solid Supported Lipid Bilayers: From Biophysical Studies to Sensor Design. *Surface Sci. Reports*, 61(10):429–444, 2006. 210

[41] R. Cecchelli, V. Berezowski, S. Lundquist, M. Culot, M. Renftel, M. Dehouck, and L. Fenart. Modelling of the Blood-brain Barrier in Drug Discovery and Development. *Nature Rev. Drug Discovery*, 6:650–661, 2007. 10

[42] D. Chandler. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature*, 437(7059):640–647, 2005. 99

[43] B. V. Cheney and J. Kalantar. Computer-aided Structural Comparisons of Clonidine and Guanfacine with Cyclazocine. *J. Mol. Graphics*, 4(1):21–27, 1986. 200

[44] M. Chesler. Regulation and Modulation of pH in the Brain. *Physio. Rev.*, 83(4):1183–1221, 2003. 124

[45] L .K. Chico, L. J. Van Eldik, and D. M. Watterson. Targeting Protein Kinases in Central Nervous System Disorders. *Nature Rev. Drug Discovery*, 8(11):892–909, 2009. 142

[46] G. Cobourne, G. Mountjoy, J. D. Rodriguez-Blanco, L. G. Benning, A. C. Hannon, and J. R. Plaisier. Neutron and X-ray Diffraction and Empirical Potential Structure Refinement Modelling of Magnesium Stabilised Amorphous Calcium Carbonate. *J. Non-crystalline Solids*, 401:154–158, 2014. 40

[47] V. Cody and G. T. DeTitta. The Molecular Conformation of Clonidine Hydrochloride, an α-andrenergic Agonist. *J. Crystal and Mol. Structure*, 9(1):33–43, 1979. 200

[48] U. Dahlborg and B. Kunsch. On the Placzek Correction for Simple Liquids and Gases. *Phys. and Chem. of Liquids*, 12(3):237–254, 1983. 39

[49] F. Dangkoob, M. R. Housaindokht, A. Asoodeh, O. Rajabi, Z. R. Zaeri, and A. V. Doghaei. Spectroscopic and Molecular Modeling Study on the Separate and Simultaneous Bindings of Alprazolam and Fluoxetine Hydrochloride to Human Serum Albumin (HSA): With the Aim of the Drug Interactions Probing. *Spectrochimica Acta Part A: Mol. and Biomol. Spec.*, 137:1106–1119, 2015. 142, 151

[50] T. Darden, D. York, and L. Pedersen. Particle Mesh Ewald: An N.log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.*, 98, 1993. 68, 110, 147, 177

[51] S. Datta and D. J. W. Grant. Crystal Structures of Drugs: Advances in Determination, Prediction and Engineering. *Nature*, 3:42–57, 2004. 1

[52] A. M. Davis, S. A. St-Gallay, and G. J. Kleywegt. Limitations and Lessons in the Use of X-ray Structural Information in Drug Design. *Drug Discovery Today*, 13(19-20):831–841, 2008. 1

[53] A. J. de Jesus and T. W. Allen. The Role of Tryptophan Side Chains in Membrane Protein Anchoring and Hydrophobic Mismatch. *Biochimica et Biophysica Acta*, 1828(2):864–876, 2012. 60

[54] M. R. de Planque, J. A. Kruijtzer, R. M. Liskamp, D. Marsh, D. V. Greathouse, R. E. Koeppe, B. de Kruijff, and J. A. Killian. Different Membrane Anchoring Positions of Tryptophan and Lysine in Synthetic Transmembrane Alpha-helical Peptides. *J. Bio. Chem.*, 274(30):20839–20846, 1999. 59, 60

[55] J. S. Delaney. Predicting Aqueous Solubility from Structure. *Drug Discovery Today*, 10(4):289–295, 2005. 2

[56] S. Dixit, J. Crain, W. C. K. Poon, J. L. Finney, and A. K. Soper. Molecular Segregation Observed in a Concentrated Alcohol-Water Solution. *Nature*, 416:42–57, 2002. 72, 73, 74, 76, 89, 98, 152, 153, 154, 182, 183, 184, 203

[57] J. R. Docherty. Subtypes of Functional α1- and α2-adrenoceptors. *Euro. J. Pharm.*, 361(1):1–15, 1998. 171

[58] A. Doran, R. S. Obacht, B. J. Smith, N. A. Hosea, S. Becker, E. Callegari, C. Chen, X. Chen, E. Choo, J. Cianfrogna, L. M. Cox, J. P. Gibbs, M. A. Gibbs, H. Hatch, C. E. C. A. Hop, I. N. Kasman, J. LaPerle, J. Liu, X. Liu, M. Logman, D. Maclin, F. M. Nedza, F. Nelson, E. Olson, S. Rahematpura, D. Raunig, S. Rogers, K. Schmidt, D. K. Spracklin, M. Szewc, M. Troutman, E. Tseng, M. Tu, J. W. Van Deusen, K. Venkatakrishnan, G. Walens, E. Q. Wang, D. Wong, A. S. Yasgar, and C. Zhang. The Impact of the P-glycoprotein on the Disposition of Drugs Targeted for Indications of the Central Nervous System: Evaluation Using the mdr1a/1b Knockout Mouse Model. *Drug Metabolism and Disposition*, 33(1):165–174, 2005. 11

[59] L. Dougan, S. P. Bates, R. Hargreaves, J. P. Fox, J. Crain, J. L. Finney, V. Reat, and A. K. Soper. Methanol-Water Solutions: A Bi-percolating Liquid Mixture. *J. Chem. Phys.*, 121, 2004. 73, 203

[60] L. Dougan, R. Hargreaves, S. P. Bates, J. L. Finney, V. Réat, A. K. Soper, and J. Crain. Segregation in Aqueous Methanol Enhanced by Cooling and Compression. *J. Chem. Phys.*, 122(17), 2005. 74

[61] D. Dougherty. Cation-π Interactions in Chemistry and Biology : A New View of Benzene, Phe, Tyr, and Trp. *Science*, 271:163–168, 1996. 60

[62] D. J. Earl and M. W. Deem. *Molecular Modeling of Proteins*, chapter Monte Carlo Simulations, pages 25–36. Humana Press, Totowa, NJ, 2008. 6

[63] P. A. Egelstaff and A. K. Soper. The Scattering of Slow Neutrons by Diatomic Molecular Fluids. *Mol. Phys.*, 40(3):569–584, 1980. 16

[64] P. Ertl, B. Rohde, and P. Selzer. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.*, 43(20):3714–3717, 2000. 207

[65] E. Esbjörner, C. E. B. Caesar, B. Albinsson, P. Lincoln, and B. Nordeń. Tryptophan Orientation in Model Lipid Membranes. *Biochem. and Biophys. Research Commun.*, 361:645–650, 2007. 60

[66] H. Eyring and R. P. March. Significant Structure Theory of Liquids. *J. Chem. Edu.*, 40(11):562, 1963. 3

[67] C. J. Fecko, J. D. Eaves, J. J. Loparo, A. Tokmakoff, and P. L. Geissler. Ultrafast Hydrogen Bond Dynamics in the Infrared Spectroscopy of Water. *Science*, 301(5640):1698–1702, 2003. 7

[68] P. Filippakopoulos, S. Picaud, O. Fedorov, M. Keller, M. Wrobel, O. Morgenstern, F. Bracher, and S. Knapp. Benzodiazepines and Benzotriazepines as Protein Interaction Inhibitors Targeting Bromodomains of the BET Family. *Bioorg. & Med. Chem.*, 20(6):1878–1886, 2012. 142, 151, 167, 169

[69] P. Filippakopoulos, J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. La Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp, and J. E. Bradner. Selective Inhibition of BET Bromodomains. *Nature*, 468:1067–1073, 2010. 142, 167

[70] J. L. Finney, B. J. Gellatly, I. C. Golton, and J. Goodfellow. Solvent Effects and Polar Interactions in the Structural Stability and Dynamics of Globular Proteins. *Biophys. J.*, 32:17–33, 1980. 100

[71] J. L. Finney and A. K. Soper. Solvent Structure and Perturbations in Solutions of Chemical and Biological Importance. *Chem. Soc. Rev.*, 23:1–10, 1994. 20

[72] H. Fischer, R. Gottschlich, and A. Seelig. Blood-brain Barrier Permeation: Molecular Parameters Governing Passive Diffusion. *J. Membrane Bio.*, 165, 1998. 10

[73] H. E. Fischer, A. C. Barnes, and P. S. Salmon. Neutron and X-ray Diffraction Studies of Liquids and Glasses. *Reports on Progress in Phys.*, 69(1):233, 2006. 16, 39

[74] J. A. Fleming, R. Byck, and P. G. Barash. Pharmacology and Therapeutic Applications of Cocaine. *Anaesthesiology*, 73:518–513, 1990. 10

[75] M. A. Floriano, E. Whalley, E. C. Svensson, and V. F. Sears. Structure of High-density Amorphous Ice by Neutron Diffraction. *Phys. Rev. Lett.*, 57:3062–3064, 1986. 26

[76] F. Foglia, J. M. Lawrence, C. D. Lorenz, and S. E. McLain. On the Hydration of the Phosphocholine Headgroup in Aqueous Solution. *J. Chem. Phys.*, 133(14), 2010. 122, 209

[77] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, . Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009. 67

[78] L. Fu, M. Tian, X. Li, J. Li, J. Huang, L. Ouyang, Y. Zhang, and B. Liu. Inhibition of BET Bromodomains as a Therapeutic Strategy for Cancer Drug Discovery. *Oncotarget*, 6(8):5501–5516, 2015. 142, 206

[79] E. J. Gabe and W. H. Barnes. The Crystal and Molecular Structure of *l*-cocaine Hydrochloride. *Acta Cryst.*, 16(8):796–801, 1963. 138

[80] D. Gallenkamp, K. A. Gelato, B. Haendler, and H. Weinmann. Bromodomains and Their Pharmacological Inhibitors. *Chem. Med. Chem*, 9(3):438–464, 2014. 169

[81] A. T. García-Sosa. Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-bound, Bridging Water Molecules and Their Effects and Conse-

quences on Molecular Design Strategies. *J. Chem. Info. and Modeling*, 53(6):1388–1405, 2013. 170

[82] A. T. García-Sosa, R. Mancera, and P. M. Dean. WaterScore: A Novel Method for Distinguishing Between Bound and Displaceable Water Molecules in the Crystal Structure of the Binding Site of Protein-ligand Complexes. *J. Mol. Modeling*, 9(3):172–182, 2003. 170, 206

[83] P. Gard. Imidazole Receptors - An Enlightening Experience. *Drug Discovery Today*, 3(10):445–446, 1998. 172, 206

[84] gcm.upc.edu/en/members/luis carlos/angula/ANGULA. June 2016. 47

[85] W. J. Geldenhuys, D. D. Allen, and J. R. Bloomquist. Novel Models for Assessing Blood-brain Barrier Drug Permeation. *Expert Opin. Drug Metab. Toxicol.*, 8:647–653, 2012. 9

[86] R. J. Gillams, J. V. Busto, S. Busch, F. M. Go ni, C. D. Lorenz, and S. E. McLain. Solvation and Hydration of the Ceramide Headgroup in a Non-polar Solution. *J. Phys. Chem. B*, 119(1):128–139, 2015. 47, 50

[87] glenergy.biz/res_XvsN.php. August 2016. 21

[88] J. T. Goodwin and D. E. Clark. *In Silico* Predictions of Blood-brain Barrier Penetration: Considerations to 'Keep in Mind'. *J. Pharm. and Exp. Therapeutics*, 315(2):477–483, 2005. 11

[89] J. Greer, J. W. Erickson, J. J. Baldwin, and M. D. Varney. Application of the Three-dimensional Structures of Protein Target Molecules in Structure-based Drug Design. *J. Med. Chem.*, 37(8):1035–1054, 1994. 1, 2

[90] D. Groen, G. S. Gooris, D. J. Barlow, M. J. Lawrence, J. B. van Mechelen, B. Deme, and J. A. Bouwstra. Disposition of Ceramide in Model Lipid Membranes Determined by Neutron Diffraction. *Biophys. J.*, 100:1481–1489, 2011. 210

[91] J. Guo, Y. Luo, A. Augustsson, S. Kashtanov, J. Rubensson, D. K. Shuh, H. Agren, and J. Nordgren. Molecular Structure of Alcohol-Water Mixtures. *Phys. Rev.*, 91, 2003. 74, 89, 203

[92] P. G. Guyenet. Is the Hypotensive Effect of Clonidine and Related Drugs due to Imidazoline Binding Sites? *Am. J. Physio. - Regulatory, Integrative and Comparative Physio.*, 273(5):1580–1584, 1997. 172, 206

[93] A. B. Haberman. Strategies to Overcome Blood-brain Barrier. *Genetic Eng. and Biotech. News*, 29, 2009. 10

[94] J. Handgraaf, T. S. van Erp, and E. J. Meijer. Ab Initio Molecular Dynamics Study of Liquid Methanol. *Chem. Phys. Lett.*, 367(56):617–624, 2003. 74, 203

[95] A. C. Hannon, W. S. Howells, and A. K. Soper. ATLAS: A Suite of Programs for the Analysis of Time-of-flight Neutron Diffraction Data from Liquid and Amorphous Samples. *Inst. Phys. Conf. Ser.*, (107):193–211, 1990. 30

[96] J. P. Hansen and I.R. McDonald. Academic Press, Burlington, 3 edition, 2006. 6, 7

[97] S. B. Hansen and P. Taylor. Galanthamine and Non-competitive Inhibitor Binding to ACh-binding Protein: Evidence for a Binding Site on Non-α-subunit Interfaces of Heteromeric Neuronal Nicotinic Receptors. *J. Mol. Bio.*, 369(4):895–901, 2007. 103, 104

[98] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchinson. Avogadro: An Advanced Semantic Chemical Editor, Visualisation and Analysis Platform. *J. Cheminformatics*, 4, 2012. 64, 107, 145, 175

[99] J. Harmon, C. Coffman, S. Villarrial, S. Chabolla, K. A. Heisel, and V. V. Krishnan. Determination of Molecular Self-diffusion Coefficients Using Pulsed-field-Gradient NMR: An Experiment for Undergraduate Physical Chemistry Laboratory. *J. Chem. Edu.*, 89(6):780–783, 2012. 137

[100] D. K. Hatksukami and M. W. Fischman. Crack Cocaine and Cocaine Hydrochloride: Are the Differences Myth or Reality. *JAMA*, 276(16):1580–1588, 1996. 103

[101] C. Haw and J. Stubbs. Benzodiazepines - A Necessary Evil? A Survey of Prescribing at a Specialist UK Psychiatric Hospital. *J. Psychopharm.*, 21(6):645–649, 2007. 141

[102] G. A. Head, C. K. S. Chan, and S. L. Burke. Relationship Between Imidazoline and α2-adrenoceptors Involved in the Sympatho-inhibitory Actions of Centrally Acting Antihypertensive Agents. *J. Auto. Nerv. Sys.*, 72(2-3):163–169, 1998. 172

[103] E. R. Hernández. Molecular Dynamics: From Basic Techniques to Applications (A Molecular Dynamics Primer). *AIP Conf. Proc.*, 1077(1):95–123, 2008. 5, 13

[104] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comp. Chem.*, 18:1463–1472, 1997. 68, 110

[105] A. M. Holbrook, R. Crowther, A. Lotter, C. Cheng, and D. King. Meta-analysis of Benzodiazepine Use in the Treatment of Insomnia. *Canadian Med. Assoc. J.*, 162(2):225–233, 2000. 141

[106] A. Holt, D. J. Smith, L. Cendron, G. Zanotti, A. Rigo, and M. L. Di Paolo. Multiple Binding Sites for Substrates and Modulators of Semicarbazide-sensitive Amine Oxidases: Kinetic Consequences. *Mol. Pharm.*, 73(2):525–538, 2008. 200

[107] M. C. Houston. Clonidine Hydrochloride: Review of Pharmacologic and Clinical Aspects. *Progress in Cardiovascular Diseases*, 23(5):337350, 1981. 172

[108] B. Hu and M. A. Lill. WATsite: Hydration Site Prediction Program with PyMOL Interface. *J. Comp. Chem*, 35:1255–1260, 2014. 170

[109] W. Hu and T. A. Cross. Tryptophan Hydrogen Bonding and Electric Dipole Moments: Functional Roles in the Gramicidin Channel and Implications for Membrane Proteins. *Biochem.*, 34:14147–14155, 1995. 60, 100

[110] W. Hu, K. C. Lee, and T. A. Cross. Tryptophans in Membrane Proteins: Indole Ring Orientations and Functional Implications in the Gramicidin Channel. *Biochem.*, 32:7035–7047, 1993. 60

[111] J. H. Hubbell, Wm. J. Veigele, E. A. Briggs, R. T. Brown, D. T. Cromer, and R. J. Howerton. Atomic Form Factors, Incoherent Scattering Functions, and Photon Scattering Cross Sections. *J. Phys. and Chem. Reference Data*, 4(3):471–538, 1975. 40

[112] E. C. Hulme, A. K. Soper, S. E. McLain, and J. L Finney. The Hydration of the Neurotransmitter Acetylcholine in Aqueous Solution. *Biophys. J.*, 91(6):2371–2380, 2006. 122

[113] W. Hunkeler, H. Mohler, L. Pieri, P. Polc, E. P. Bonetti, R. Cumin, R. Schaffner, and W. Haefely. Selective Antagonists of Benzodiazepines. *Nature*, 290:514–516, 1981. 141

[114] C. A. Hunter and J. K. M. Sanders. The Nature of π-π Interactions. *J. Am. Chem. Soc.*, 112(14):5525–5534, 1990. 136

[115] J. Huwyler, G. Fricker, M. Török, M. Schneider, and J. Drewe. Transport of Clonidine Across Cultured Brain Microvessel Endothelial Cells. *J. Pharm. and Exp. Therapeutics*, 282:8185, 1997. 200

[116] L. Infantes and W. D. S. Motherwell. Hydrogen Bond Competition Between Chemical Groups: New Methodology and the Cambridge Structural Database. *Z. Kristallogr.*, 220:333–339, 2005. 139

[117] E. P. Irany. A Theory of Liquid Structure. *J. Am. Chem. Soc.*, 61(6):1436–1443, 1939. 3

[118] G. K. Isbister, L. O'Regan, D. Sibbritt, and I. M. Whyte. Alprazolam is Relatively More Toxic than other Benzodiazepines in Overdose. *Brit. J. Clin. Pharm.*, 58(1):88–95, 2004. 142

[119] R. E. Jacobs and S. H. White. The Nature of the Hydrophobic Binding of Small Peptides at the Bilayer Interface: Implications for the Insertion of Transbilayer Helices. *Biochem.*, 28(8):3421–3437, 1989. 59

[120] B. Jacrot. The Study of Biological Structures by Neutron Scattering from Solution. *Reports on Progress in Phys.*, 39(10):911–953, 1976. 21

[121] R. Jakkala-Saibaba, P. G. Morgan, and G. L. Morton. Treatment of Cocaine Overdose with Lipid Emulsion. *Anaesthesia*, 66(12):1168–1170, 2011. 103

[122] B. Jarrott, E. L. Conway, C. Maccarrone, and S. J. Lewis. Clonidine: Understanding its Disposition, Sites and Mechanism of Action. *Clin. and Exp. Pharm. and Phys.*, 14(5):471–479, 1987. 171, 172, 206

[123] G. A. Jeffrey and W. Saenger. Intramolecular Hydrogen Bonds. In *Hydrogen Bonding in Biological Structures*, pages 147–155. Springer Berlin Heidelberg, 1991. 120

[124] M. S. Jhon and H. Eyring. Liquid Theory and the Structure of Water. *Ann. Rev. Phys. Chem.*, 27:45–57, 1976. 4

[125] A. J. Johnston, S. B. Busch, L. C. Pardo, S. K. Callear, P. C. Biggin, and S. E. McLain. On the Atomic Structure of Cocaine in Solution. *Phys. Chem. Chem. Phys.*, 18:991–999, 2016. 47, 50, 52, 55, 56, 83

[126] A. J. Johnston, Y. Zhang, S. Busch, L. C. Pardo, S. Imberti, and S. E. McLain. Amphipathic Solvation of Indole: Implications for the Role of Tryptophan in Membrane Proteins. *J. Phys. Chem. B*, 119(19):5979–5987, 2015. 47, 50, 53, 137

[127] D. R. Jones, S. D. Hall, E. K. Jackson, R. A. Branch, and G. R. Wilkinson. Brain Uptake of Benzodiazepines: Effects of Lipophilicity and Plasma Protein Binding. *J. Pharm. and Exp. Therapeutics*, 245(3):816–822, 1988. 141

[128] M. O. Jones, A. D. Taylor, and S. F. Parker. Neutron Scattering Studies of Catalyst Systems at the ISIS Neutron Spallation Source. *Applied Petrochemical Research*, 2(3):97–104, 2012. 30

[129] 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.*, 79(2):926–935, 1983. 109

[130] W. L. Jorgensen and J. Tirado-Rives. Monte Carlo vs. Molecular Dynamics for Conformational Sampling. *J. Phys. Chem.*, 100(34):14508–14513, 1996. 5, 13

[131] C. Joy, P. P. Boyle, and K. S. Tan. Quasi-Monte Carlo Methods in Numerical Finance. *Management Sci.*, 42(6):926–938, 1996. 6

[132] K. Kachel, E. Asuncion-Punzaland, and E. London. Anchoring of Tryptophan and Tyrosine Analogs at the Hydrocarbon-Polar Boundary in Model Membrane Vesicles: Parallax Analysis of Fluorescence Quenching Induced by Nitroxide-Labeled Phospholipids. *Biochem.*, 34:15475–15479, 1995. 60

[133] G. A. Kaminski and R. A. Friesner. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins *via* Comparison with Accurate Qunatum Chemical Calculations on Peptides. *J. Phys. Chem. B*, 105:6474–6487, 2001. 64, 65, 67, 109

[134] P. A. Karplus and C. Faerman. Ordered Water in Macromolecular Structure. *Current Opinion in Structural Bio.*, 4(5):770–776, 1994. 1, 2

[135] D. A. Keen and R. L. McGreevy. Structural Modelling of Glasses using Reverse Monte Carlo Simulation. *Nature*, 344:423–425, 1990. 6

[136] D. B. Kell, P. D. Dobson, E. Bilsland, and S. G. Oliver. The Promiscuous Binding of Pharmaceutical Drugs and their Transporter-mediated Uptake into Cells: What We (Need to) Know and How We Can Do So. *Drug Discovery Today*, 18:218–239, 2013. 10

[137] D. B. Kell, P. D. Dobson, and S. G. Oliver. Pharmaceutical Drug Transport: The Issues and the Implications that it is Essentially Carrier-mediated Only. *Drug Discovery Today*, 16(15-16):704–714, 2011. 10

[138] M. Khalil, N. Demirdöven, and A. Tokmakoff. Coherent 2D IR Spectroscopy: Molecular Structure and Dynamics in Solution. *J. Phys. Chem. A*, 107(27):5258–5279, 2003. 7

[139] Z. P. Khan, C. N. Ferguson, and R. M. Jones. Alpha-2 and Imidazoline Receptor Agonists: Their Pharmacology and Therapeutic Role. *Anaesthesia*, 54(2):146–165, 1999. 172

[140] J. A. Killian and G. von Heijne. How Proteins Adapt to a Membrane-Water Interface. *Trends in Biochem. Sci.*, 25(9):429–434, 2000. 59, 100

[141] H. T. Kim, T. Sakamoto, and Masaaki Ando. Novel [^{3}H]Clonidine Binding Sites in the Intestine of the Eel Acclimated to Sea Water. *Zoo. Sci.*, 15(2):205–212, 1998. 200

[142] J. Kirkwood, G. John, and F. P. Buff. The Statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.*, 19(6):774–777, 1951. 4

[143] J. G. Kirkwood. Molecular Distribution in Liquids. *The J. Chem. Phys.*, 7(10):919–925, 1939. 4

[144] J. G. Kirkwood and E. M. Boggs. The Radial Distribution Function in Liquids. *J. Chem. Phys.*, 10(6):394–402, 1942. 4

[145] J. G. Kirkwood, V. A. Lewinson, and B. J. Alder. Radial Distribution Functions and the Equation of State of Fluids Composed of Molecules Interacting According to the Lennard-Jones Potential. *J. Chem. Phys.*, 20(6):929–938, 1952. 4

[146] A. E. Knaus, V. Muthig, S. Schickinger, E. Moura, N. Beetz, R. Gilsbach, and L. Hein. α2-adrenoceptor Subtypes - Unexpected Functions for Receptors and Ligands Derived from Gene-targeted Mouse Models. *Neurochem. Int.*, 51(5):277–281, 2007. 172

[147] B. K. Kobilka, H. Matsui, T. S. Kobilka, T. L. Yang-Feng, U. Francke, M. G. Caron, R. J. Lefkowitz, and J. W. Regan. Cloning, Sequencing, and Expression of the Gene Coding for the Human Platelet Alpha 2-adrenergic Receptor. *Science*, 238(4827):650–656, 1987. 200

[148] R. E. Koeppe. Concerning Tryptophan and Protein-Bilayer Interactions. *J. Gen. Physio.*, 130:223–224, 2007. 60

[149] E. M. Komoroski, R. A. Komoroski, J. L. Valentine, J. M. Pearce, and G. L. Kearns. The Use of Nuclear Magnetic Resonance Spectroscopy in the Detection of Drug Intoxication. *J. Analy. Toxicol.*, 24(3):180–187, 2000. 7

[150] S. Krueger, C. W. Meuse, C. F. Majkrzak, J. A. Dura, N. F. Berk, M. Tarek, and A. L. Plant. Investigation of Hybrid Bilayer Membranes with Neutron Reflectometry: Probing the Interactions of Melittin. *Langmuir*, 17(2):511–521, 2001. 210

[151] R. F. Kruh. Diffraction Studies of the Structure of Liquids. *Chem. Rev.*, 62(4):319–346, 1962. 4

[152] T. Kudo, T. Mizoguchi, N. Watanabe, N. Niimura, M. Misawa, and K. Suzuki. Partial Structure Factor of Amorphous $Cu_{0.57}Zr_{0.43}$ Alloy Determined by TOF Pulsed Neutron Diffraction. *J. Phys. Soc. Japan*, 45(5):1773–1774, 1978. 27

[153] B. Kuhn, P. Mohr, and M. Stahl. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.*, 53:2601–2611, 2010. 139, 205

[154] J. E. Ladbury. Just Add Water! The Effect of Water on the Specificity of Protein-ligand Binding Sites and its Potential Application to Drug Design. *Chem. & Bio*, 3(12):973–980, 1996. 1, 2

[155] M. Lader. History of Benzodiazepine Dependence. *J. Substance Abuse Treatment*, 8(1-2):53–59, 1991. 142

[156] P. Y. Lam, P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bacheler, J. L. Meek, M. J. Otto, M. M. Rayner, Y. N. Wong C. H. Chang, P. C. Weber, D. A. Jackson, T. R. Sharpe, and S. Erickson-Viitanen. Rational Design of Potent, Bioavailable, Nonpeptide Cyclic Ureas as HIV Protease Inhibitors. *Science*, 263(5145):380–384, 1994. 2

[157] C. Landron, L. Hennet, T. E. Jenkins, G. N. Greaves, J. P. Coutures, and A. K. Soper. Liquid Alumina: Detailed Atomic Coordination Determined from Neutron Diffraction Data Using Empirical Potential Structure Refinement. *Phys. Rev. Lett.*, 86:4839–4842, 2001. 40

[158] N. A. Larsen, B. Zhou, A. Heine, P. Wirsching, K. D. Janda, and I. A. Wilson. Crystal Structure of a Cocaine-binding Antibody. *J. Mol. Bio.*, 311:9–15, 2001. 139

[159] J. F. Leckman, M. T. Hardin, M. A. Riddle, J. Stevenson, S. I. Ort, and D. J. Cohen. Clonidine Treatment of Gilles de la Tourettes Syndrome. *Archives of Gen. Psychiatry*, 48(4):324–328, 1991. 172, 206

[160] R. U. Lemieux. How Water Provides the Impetus for Molecular Recognition in Aqueous Solution. *Accounts Chem. Res.*, 29(8):373–380, 1996. 170, 206

[161] P. J. Leonard, D. Henderson, and J. A. Barker. Perturbation Theory and Liquid Mixtures. *Trans. Faraday Soc.*, 66:2439–2452, 1970. 4

[162] M. Levitt and M. F. Perutz. Aromatic Rings Act as Hydrogen Bond Acceptors. *J. Mol. Bio.*, 201:751–754, 1988. 100, 207

[163] D. R. Lide, editor. *CRC Handbook of Chemistry and Physics, Internet Version 2005*. CRC Press, Boca Raton, FL, 2005. 60

[164] E. R. Lindahl. *Molecular Modeling of Proteins*, chapter Molecular Dynamics Simulations, pages 25–36. Humana Press, Totowa, NJ, 2008. 5

[165] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved Side-chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010. 147, 177

[166] B. Linse and P. Linse. Tuning the Smooth Particle Mesh Ewald Sum: Application on Ionic Solutions and Dipolar Fluids. *J. Chem. Phys.*, 141(18), 2014. 147, 177

[167] J. E. G. Lipson. Studies on Fluids and their Mixtures Using the Born-Green-Yvon Integral Equation Technique. *Macromol. Theory and Sim.*, 7(2):263–282, 1998. 4

[168] R. Loewenthal, J. Sancho, and A. R. Fersht. Histidine-Aromatic Interactions in Barnase. *J. Mol. Bio.*, 224:759–770, 1991. 100

[169] V. Lounnas, T. Ritschel, J. Kelder, R. McGuire, R. P. Bywater, and N. Foloppe. Current Progress in Structure-based Rational Drug Design Marks a New Mindset in Drug Discovery. *Comp. and Struct. Biotech. J.*, 5(6):1–14, 2013. 2

[170] Y. Lu, R. Wang, C. Y. Yang, and S. Wang. Analysis of Ligand-bound Water Molecules in High-resolution Crystal Structures of Protein-Ligand Complexes. *J. Chem. Info. and Modeling*, 47(2):668–675, 2007. 170

[171] J. M. Luco. Prediction of the Brain-blood Distribution of a Large Set of Drugs from Structurally Derived Descriptors Using Partial Least-squares (PLS) Modeling. *J. Chem. Info. and Comp. Sci.*, 39(2):396–404, 1999. 11

[172] S. Maitra, B. Saha, C. R. Santra, A. Mukherjee, S. Goswami, P. K. Chanda, and P. Karmakar. Alprazolam Induced Conformational Change in Hemoglobin. *Inter. J. Bio. Macromol.*, 41(1):23–29, 2007. 142

[173] J. F. Malone, C. M. Murray, M. H. Charlton, R. Dochertyb, and A.J. Lavery. X–H$\cdots\pi$ (phenyl) Interactions. *J. Chem. Soc., Faraday Trans.*, 437(7059):640–647, 1997. 99, 100

[174] P. Mark and L. Nilsson. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A*, 105(43):9954–9960, 2001. 64, 109

[175] T. J. Marrone, J. M. Briggs, and J. A. McCammon. Structure-based Drug Design: Computational Advances. *Annual Rev. Pharm. and Toxicol.*, 37(1):71–90, 1997. 1

[176] P. R. Martin, M. H. Ebert, E. K. Gordon, M. Linnoila, and I. J. Kopin. Effects of Clonidine on Central and Peripheral Catecholamine Metabolism. *Clin. Pharm. & Therapeutics*, 35(3):322–327, 1984. 172, 206

[177] J. Mavri and D. Hadži. Tautomerism, Protonation, and Interaction with Formiate, of Phenyliminoimidazoline and Benzylimidazolidine: AM1 and *ab initio* 431G Calculations. *J. Mol. Structure*, 224:285–296, 1990. 200

[178] J. Mayers. The Use of Vanadium as a Scattering Standard for Pulsed Source Neutron Spectrometers. *Nuclear Instruments and Methods in Phys. Research*, 221(3):609–618, 1984. 34

[179] R. L. McGreevy. Reverse Monte Carlo Modelling. *J. Phys.: Condensed Matter*, 13(46), 2001. 6, 40

[180] S. E. McLain, S. Imberti, A. K. Soper, A. Botti, F. Bruni, and M. A. Ricci. Structure of 2 Molar NaOH in Aqueous Solution from Neutron Diffraction and Empirical Potential Structure Refinement. *Phys. Rev. B*, 74:094201, 2006. 27

[181] S. E. McLain, A. K. Soper, and A. Luzar. Orientational Correlations in Liquid Acetone and Dimethyl Sulfoxide: A Comparative Study. *J. Chem. Phys.*, 124(7), 2006. 204

[182] S. E. McLain, A. K. Soper, A. E. Terry, and A. Watts. Structure and Hydration of l-Proline in Aqueous Solutions. *J. Phys. Chem.*, 111(17):4568–4580, 2007. 77

[183] J. Meller. *Molecular Dynamics*. John Wiley & Sons, Ltd, 2001. 5, 13

[184] T. J. Mitchell, M. S. Tute, and G. A. Webb. A Theoretical Study of the pKa Values of some Clonidine-like Imidazolidines. *Euro. J. Med. Chem.*, 25(2):117–120, 1990. 200

[185] A. Mitra, P. J. Seaton, R. A. Assarpour, and T. Williamson. Unprecedented Concentration Dependent Chemical Shift Variation in ^{1}H-NMR Studies: A Caveat in the Investigations of Molecular Recognition and Structure Elucidation. *Tetrahedron*, 54:15489–15498, 1998. 135

[186] D. L. Mobley, J. D. Chodera, and K. A. Dill. On the use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. *J. Chem. Phys.*, 125(8), 2006. 67, 109, 147, 177

[187] M. Muehlbacher, G. M. Spitzer, K. R. Liedl, and J. Kornhuber. Qualitative Prediction of Blood-brain Barrier Permeability on a Large and Refined Dataset. *J. Computer-Aided Mol. Design*, 25(12):1095–1106, 2011. 11

[188] S. Mukamel. Multidimensional Femtosecond Correlation Spectroscopies of Electronic and Vibrational Excitations. *Annual Rev. Phys. Chem.*, 51(1):691–729, 2000. 7

[189] S. Muller, P. Filippakopoulos, and S. Knapp. Bromodomains as Therapeutic Targets. *Expert Rev. Mol. Med.*, 13, 2011. 142, 206

[190] S. Müller and S. Knapp. Discovery of BET Bromodomain Inhibitors and their Role in Target Validation. *Med. Chem. Commun.*, 5:288–296, 2014. 167

[191] C. N. Nguyen, T. K. Young, and M. K. Gilson. Grid Inhomogeneous Solvation Theory: Hydration Structure and Thermodynamics of the Miniature Receptor Cucurbit[7]uril. *J. Chem. Phys.*, 137, 2012. 170

[192] B. R. A. Nijboer and L. Van Hove. Radial Distribution Function of a Gas of Hard Spheres and the Superposition Approximation. *Phys. Rev.*, 85:777–783, 1952. 4

[193] U. Norinder and M. Haeberlein. Computational Approaches to the Prediction of the Blood-Brain Distribution. *Advanced Drug Delivery Rev.*, 54:291–313, 2002. 10

[194] K. E. Norman and H. Nymeyer. Indole Localization in Lipid Membranes Revealed by Molecular Simulation. *Biophys. J.*, 91(6):2046–2054, 2006. 66, 79

[195] Y. Nozaki and C. Tanford. The Solubility of Amino Acids and Two Glycine Peptides in Aqueous Ethanol and Dioxane Solutions. *J. Bio. Chem.*, 246:2211–2217, 1971. 98

[196] A. Ntranos and P. Casaccia. Bromodomains: Translating the Words of Lysine Acetylation into Myelin Injury and Repair. *Neurosci. Lett.*, 2015. 142

[197] A. M. O'Connell, R. E. Koeppe, and O. S. Andersen. Kinetics of Gramicidin Channel Formation in Lipid Bilayers: Transmembrane Monomer Association. *Science (New York, N.Y.)*, 250(4985):1256–1259, 1990. 59

[198] T. Österberg and U. Norinder. Prediction of Polar Surface Area and Drug Transport Processes Using Simple Parameters and PLS Statistics. *J. Chem. Info. and Comp. Sci.*, 40(6):1408–1411, 2000. 206, 207

[199] C. N. Pace, S. Trevino, E. Prabhakaran, and J. M. Scholtz. Protein Structure, Stability and Solubility in Water and Other Solvents. *Royal Soc.*, 359:1225–1235, 2011. 100

[200] S. E. Pagnotta, S. E. McLain, A. K. Soper, F. Bruni, and M. A. Ricci. Water and Trehalose: How Much Do They Interact with Each Other? *J. Phys. Chem. B*, 114(14):4904–4908, 2010. 137

[201] H. Pajouhesh and G. R. Lenz. Medicinal Chemical Properties of Successful Central Nervous System Drugs. *J. Am. Soc. Exp. Neurotherapeutics*, 2:541–553, 2005. 11, 206, 207

[202] L. C. Pardo, A. Henao, S. Busch, E. Guardia, and J. L. Tamarit. A Continuous Mixture of Two different Dimers in Liquid Water. *Phys. Chem. Chem. Phys.*, 16:24479–24483, 2014. 47

[203] L. C. Pardo, A. Henao, and A. Vispa. Characterizing Ordering in Liquids: An Information Theoretic Approach. *J. Non-crystalline Solids*, 407:220–227, 2015. 47

[204] L. C. Pardo, M. Rovira-Esteva, J. L. Tamarit, N. Veglio, J. F. Bermejo, and G. J. Cuello. A Procedure to Quantify the Short Range Order of Disordered Phases. In R. Sylwester, D. Aleksandra, and M. Victor, editors, *Metastable Systems under Pressure*, NATO Science for Peace and Security Series A: Chemistry and Biology, pages 79–91. Springer Netherlands, 2010. 47

[205] W. M. Pardridge. Drug Transport Across the Blood-brain Barrier. *J. Cerebral Blood Flow and Metabolism*, 32:1959–1972, 2012. 9, 137, 206

[206] M. Parrinello and A. Rahman. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Applied Phys.*, 52(12):7182–7190, 1981. 147, 177

[207] W. D. M. Paton. A Theory of Drug Action Based on the Rate of Drug-Receptor Combination. *Proc. of the Royal Society of London B: Biological Sciences*, 154(954):21–69, 1961. 1

[208] C. Peetla, A. Stine, and V. Labhasetwar. Biophysical Interactions with Model Lipid Membranes: Applications in Drug Discovery and Drug Delivery. *Mol. Pharm.*, 6(5):1264–1276, 2009. 210

[209] G. Pèpe, G. Guiliani, S. Loustalet, and P. Halfon. Hydration Free Energy a Fragmental Model and Drug Design. *Euro. J. Med. Chem.*, 37(11):865–872, 2002. 1, 2

[210] M. Perez-Reyes, S. Di Guiseppi, G. Ondrusek, A. R. Jeffcoat, and C. E. Cook. Free-base Cocaine Smoking. *Clin. Pharm. & Therapeutics*, 32(4):459–465, 1982. 103

[211] S. Persson, J. A. Killian, and G. Lindblom. Molecular Ordering of Interfacially Localized Tryptophan Analogs in Ester- and Ether-Lipid Bilayers Studied by ^{2}H-NMR. *Biophys. J.*, 75:1365–1371, 1998. 59

[212] M. F. Perutz. The Role of Aromtaic Rings as Hydrogen-bond Acceptors in Molecular Recognition. *Phil. Trans. R. Soc.*, 345:105–112, 1993. 100

[213] F. N. R. Petersen, M. Ø. Jensen, and C. H. Nielsen. Interfacial Tryptophan Residues: A Role for the Cation-π Effect? *Biophys. J.*, 89:3985–3996, 2005. 59

[214] F. Petty, M. H. Trivedi, M. Fulton, and A. J. Rush. Benzodiazepines as Antidepressants: Does GABA Play a Role in Depression? *Bio. Psychiatry*, 38(9):578–591, 1995. 141

[215] M. Pigini, P. Bousquet, A. Carotti, M. Dontenwill, M. Giannella, R. Moriconi, A. Piergentili, W. Quaglia, S. K. Tayebati, and L. Brasili. Imidazoline Receptors: Qualitative Structure-activity Relationships and Discovery of Tracizoline and Benazoline. Two Ligands with High Affinity and Unprecedented Selectivity. *Bioorg. & Med. Chem.*, 5(5):833–841, 1997. 172, 206

[216] J. E. Piletz, G. A. Ordway, H. Zhu, B. J. Duncan, and A. Halaris. Autoradiographic Comparison of [3H]-Clonidine Binding to Non-adrenergic Sites and [alpha]2-Adrenergic Receptors in Human Brain. *Neuropsychopharm.*, 23(6):697–708, 2000. 172

[217] G. Placzek. The Scattering of Neutrons by Systems of Heavy Nuclei. *Phys. Rev.*, 86:377–388, 1952. 39

[218] P. Polc. Electrophysiology of Benzodiazepine Receptor Ligands: Multiple Mechanisms and Sites of Action. *Progress in Neurobio.*, 31(5):349–423, 1988. 141

[219] J. A. Pople. Molecular Association in Liquids. II. A Theory of the Structure of Water. *Proc. of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 205(1081):163–178, 1951. 4

[220] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, Peter M. P. M. Kasson, D. van der Spoel, B. Hess, and F. Lindahl. GROMACS 4.5: A High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinfo.*, 29. 67, 109

[221] S. Purser, P. R. Moore, S. Swallow, and V. Gouverneur. Fluorine in Medicinal Chemistry. *Chem. Soc. Rev.*, 37:320–330, 2008. 208

[222] L. Pusztai and E. Sváb. Structure Study of $Ni_{62}Nb_{38}$ Metallic Glass Using Reverse Monte Carlo Simulation. *J. Non-crystalline Solids*, 156:973–977, 1993. 6

[223] R. Pynn. *Neutron Scattering: A Primer*. Los Alamos Science, 1990. 8, 9, 27, 29, 30

[224] S. Rahman, M. E. Sowa, M. Ottinger, J. A. Smith, Y. Shi, J. W. Harper, and P. M. Howley. The Brd4 Extraterminal Domain Confers Transcription Activation Independent of pTEFb by Recruiting Multiple Proteins, Including NSD3. *Mol, and Cell. Bio.*, 31(13):2641–2652, 2011. 142

[225] H. Rattle. *An NMR Primer for Life Scientists*. Partnership Press, UK, 1995. 7

[226] M. Remko, M. Swart, and F. M. Bickelhaupt. Theoretical Study of Structure, pKa, Lipophilicity, Solubility, Absorption and Polar Surface Area of Some Centrally

Acting Antihypertensives. *Bioorg. & Med. Chem.*, 14(6):1715–1728, 2006. 200, 206

[227] M. Remko, O. A. Walsh, and W. G. Richards. Molecular Structure and Gas-phase Reactivity of Clonidine and Rilmenidine: Two-layered ONIOM Calculations. *Phys. Chem. Chem. Phys.*, 3:901–907, 2001. 172, 182, 200

[228] N. H. Rhys, A. K. Soper, and L. Dougan. The Hydrogen Bonding Ability of the Amino Acid Glutamine Revealed by Neutron Diffraction Experiments. *J. Phys. Chem. B*, 116(45):13308–13319, 2012. 1

[229] A. N. J. A. Ridder, S. Morein, J. G. Stam, A. Kuhn, B. de Kruijff, and J. A. Killian. Analysis of the Role of Interfacial Tryptophan Residues in Controlling the Topology of Membrane Proteins. *Biochem.*, 39:6521–6528, 2000. 59

[230] D. A. Rincón, M. Cordeiro, and R. A. Mosquera. On the Electronic Structure of Cocaine and its Metabolites. *J. Phys. Chem. A*, 113:13937–13942, 2009. 103

[231] J. Riss, J. Cloyd, J. Gates, and S. Collins. Benzodiazepines in Epilepsy: Pharmacology and Pharmacokinetics. *Acta Neurologica Scandinavica*, 118(2):69–86, 2008. 141

[232] F. Rived, I. Canals, E. Bosch, and M. Rosés. Acidity in Methanol-Water. *Analytica Chimica Acta*, 439(2):315–333, 2001. 93

[233] M. F. Roberts and M. Wink. *Alkaloids: Biochemistry, Ecology and Medicinal Applications*. Plenum Press, New York, USA, 1998. 103

[234] G. A. Ross, G. M. Morris, and P.C. Biggin. Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PLoS ONE*, 7(3):1–13, 2012. 170, 206

[235] M. Rovira-Esteva, N. A. Murugan, L. C. Pardo, S. Busch, J. L. Tamarit, G. J. Cuello, and F. J. Bermejo. Differences in First Neighbor Orientation Behind the Anomalies in the Low and High Density Trans-1,2-dichloroethene Liquid. *J. Chem. Phys.*, 136(12):124514, 2012. 47

[236] M. Rovira-Esteva, N. A. Murugan, L. C. Pardo, S. Busch, J. L. Tamarit, S. Pothoczki, G. J. Cuello, and F. J. Bermejo. Interplay Between Intramolecular and Intermolecular Structures of 1,1,2,2-tetrachloro-1,2-difluoroethane. *Phys. Rev. B*, 84(6):064202, 2011. 47

[237] J. S. Rowlinson and C. F. Curtiss. Lattice Theories of the Liquid State. *J. Chem. Phys.*, 19(12):1519–1529, 1951. 4

[238] R. R. Ruffolo, E. L. Yaden, and J. E. Waddell. Receptor Interactions of Imidazolines. V. Clonidine Differentiates Postsynaptic Alpha Adrenergic Receptor Subtypes in Tissues from the Rat. *J. Pharm. and Exp. Therapeutics*, 213(3):557–561, 1980. 171

[239] J. 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.*, 23(3):327–341, 1977. 147

[240] B. Saha, A. Mukherjee, C. R. Santra, A. Chattopadhyay, A. N. Ghosh, U. Choudhuri, and P. Karmakar. Alprazolam Intercalates into DNA. *J. Biomol. Structure and Dynamics*, 26(4):421–429, 2009. 142

[241] K. M. Sanchez, G. Kang, B. Wu, and J. E. Kim. Tryptophan-Lipid Interactions in Membrane Protein Folding Probed by Ultraviolet Resonance Raman and Fluorescence Spectroscopy. *Biophy. J.*, 100(9):2121–2130, 2011. 60

[242] R. Sanchez and M. M. Zhou. The Role of Human Bromodomains in Chromatin Biology and Gene Transcription. *Current Opinion in Drug Discovery and Devel.*, 12(5):659–665, 2009. 142, 206

[243] M. Sarkar, S. S. Paul, and K. K. Mukherjea. Interaction of Bovine Serum Albumin with a Psychotropic Drug Alprazolam: Physicochemical, Photophysical and Molecular Docking Studies. *J. Luminescence*, 142:220–230, 2013. 142

[244] K. T. Savjani, A. K. Gajjar, and J. K. Savjani. Drug Solubility: Importance and Enhancement Techniques. *ISRN Pharm.*, 2012, 2012. 2

[245] M. Schiffera, C. H. Chang, and F. J. Stevens. The Functions of Tryptophan Residues in Membrane Proteins. *Protein Eng.*, 5:213–214, 1992. 59, 60, 100

[246] V. D. Schmith, B. Piraino, R. B. Smith, and P. D. Kroboth. Alprazolam in End-stage Renal Disease: I. Pharmacokinetics. *J. Clin. Pharm.*, 31(6):571–579, 1991. 142

[247] E. Scoppola, A. Sodo, S. E. McLain, M. A. Ricci, and F. Bruni. Water-Peptide Site-Specific Interactions: A Structural Study on the Hydration of Glutathione. *Biophys. J.*, 106(8):1701–1709, 2014. 1

[248] V. F. Sears. Neutron Scattering Lengths and Cross Sections. *Neutron News*, 3:26–37, 1992. 19, 20, 33

[249] J. Shie, B. C. S. Chou, and Y. Chen. High performance Pirani Vacuum Gauge. *J. Vacuum Sci. & Tech. A*, 13(6):2972–2979, 1995. 62

[250] S. Shu, C. Y. Lin, H. H. He, R. M. Witwicki, D. P. Tabassum, J. M. Roberts, M. Janiszewska, S. Jin Huh, Y. Liang, J. Ryan, E. Doherty, H. Mohammed, H. Guo, D. G. Stover, M. B. Ekram, G. Peluffo, J. Brown, C. D'Santos, I. E. Krop, D. Dillon, M. McKeown, C. Ott, J. Qi, M. Ni, P. K. Rao, M. Duarte, S. Y. Wu, C. M. Chiang, L. Anders, R. A. Young, E. P. Winer, A. Letai, W. T. Barry, J. S. Carroll, H. W. Long, M. Brown, L. X. Shirley, S. A. Meyer, J. E. Bradner, and K. Polyak. Response and Resistance to BET Bromodomain Inhibitors in Triple-negative Breast Cancer. *Nature*, 529:413–7, 2016. 167

[251] W. Sieghart. Pharmacology of Benzodiazepine Receptors: An Update. *J. Psychiatry and Neurosci.*, 19(1):24–29, 1994. 141

[252] D. S. Sivia. *Elementary Scattering Theory.* Oxford University Press, Oxford, UK, 2011. 8, 9, 15, 30

[253] B. E. Smart. Fluorine Substituent Effects (on Bioactivity). *J. Fluorine Chem.*, 109(1):3–11, 2001. 208

[254] S. G. Smith, R. Sanchez, and M. M. Zhou. Privileged Diazepine Compounds and Their Emergence as Bromodomain Inhibitors. *Chem. & Bio.*, 21(5):573583, 2014. 167, 169

[255] A. K. Soper. Multiple Scattering from an Infinite Plane Slab. *Nuclear Instruments and Methods in Phys. Research*, 212(1):337–347, 1983. 37

[256] A. K. Soper. First Results from SANDALS - The Small Angle Neutron Diffractometer for Amorphous and Liquid Samples at ISIS. *ICANS-XI*, 1990. 31

[257] A. K. Soper. Empirical Potential Monte Carlo Simulation of Fluid Structure. *Chem. Phys.*, 202(2-3):295–306, 1996. 6, 40, 43

[258] A. K. Soper. The Radial Distribution Functions of Water and Ice from 220 to 673 K and at Pressures up to 400 MPa. *Chem. Phys.*, 258(2-3):121–137, 2000. 40, 43, 44

[259] A. K. Soper. Tests of the Empirical Potential Structure Refinement Method and a New Method of Application to Neutron Diffraction Data on Water. *Mol. Phys.*, 99(17):1503–1516, 2001. 6, 44

[260] A. K. Soper. Partial Structure Factors from Disordered Materials Diffraction Data: An Approach Using Empirical Potential Structure Refinement. *Phys. Rev. B*, 72:104204, 2005. 6, 44, 46

[261] A. K. Soper. Joint Structure Refinement of X-ray and Neutron Diffraction Data on Disordered Materials: Application to Liquid Water. *J. Phys.: Condensed Matter*, 19(33):335206, 2007. 112, 179

[262] A. K. Soper. *Empirical Potential Structure Refinement: A User's Guide.* 18 edition, 2009. 40, 42, 43, 44, 45, 46, 47, 204

[263] A. K. Soper. Inelasticity Corrections for Time-of-flight and Fixed Wavelength Neutron Diffraction Experiments. *Mol. Phys.*, 107(16):1667–1684, 2009. 18, 39

[264] A. K. Soper. GudrunN and GudrunX: Programs for Correcting Raw Neutron and X-ray Diffraction Data to Differential Scattering Cross Sections. RAL Technical Report (RAL-TR-2011-013). Rutherford Appleton Laboratory, STFC, UK, 2011. 32, 34, 37, 39, 40, 107

[265] A. K. Soper. Computer Simulation as a Tool for the Interpretation of Total Scattering Data from Glasses and Liquids. *Mol. Sim.*, 38(14-15):1171–1185, 2012. 40

[266] A. K. Soper. The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure? *ISRN Phys. Chem.*, 2013, 2013. 72, 73, 76, 89, 121, 122, 152, 153, 154, 182, 183, 184, 215

[267] A. K. Soper. Water and Ice Structure in the Range 220-365 K from Radiation Total Scattering Experiments. *ArXiv e-prints*, 2014. 40

[268] A. K. Soper and J. L. Finney. Hydration of Methanol in Aqueous Solution. *Phys. Rev. Lett.*, 71:4346–4349, 1993. 20, 73, 152, 182, 204

[269] A. K. Soper, W. S. Howells, and A. C. Hannon. ATLAS: Analysis of Time-of-flight Diffraction Data from Liquid and Amorphous Samples. 2012. 25, 35, 36, 37, 39

[270] A. K. Soper and A. Luzar. A Neutron Diffraction Study of Dimethyl Sulphoxide-water Mixtures. *J. Chem. Phys.*, 97(2):1320–1331, 1992. 74

[271] J. M. Sorenson, G. Hura, R. M. Glaeser, and T. Head-Gordon. What Can X-ray Scattering Tell us About the Radial Distribution Functions of Water? *J. Chem. Phys.*, 113(20):9149–9161, 2000. 27, 215

[272] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J. Comp. Chem.*, 26(16):1701–1718, 2005. 147, 177

[273] G. L. Squires. *Introduction to the Theory of Thermal Neutron Scattering.* Dover Publcations, New York, 1996. 15, 23

[274] T. Steiner. Hydrogen Bonds from Water Molecules to Aromatic Acceptors in Very High-resolution Protein Crystal Structures. *Biophys. Chem.*, 95:195–201, 2002. 76, 98, 100, 122, 123, 158, 187

[275] T. Steiner and G. Koellner. Hydrogen Bonds with π-acceptors in Proteins: Frequences and Role in Stabilizing Local 3D Structures. *J. Mol. Biol.*, 305:535–557, 2001. 98, 100

[276] R. Steingard, J. Biederman, T. Spencer, T. Wilens, and A. Gonzalez. Comparison of Clonidine Response in the Treatment of Attention-deficit Hyperactivity Disor-

der with and without Comorbid Tic Disorders. *J. Am. Acad. Child & Adolescent Psychiatry*, 32(2):350–353, 1993. 172, 206

[277] N. Steinke, R. Gillams, L. C. Pardo, C. D. Lorenz, and S. McLain. Atomic Scale Insights into Urea-Peptide Interactions in Solution. *Phys. Chem. Chem. Phys.*, 2016. 47, 50, 52

[278] S Suzuki, P. G. Green, R. E. Bumgarner, S. Dasgupta, W. A. Goddard, and G. A. Blake. Benzene Forms Hydrogen Bonds with Water. *Science*, 257(5072):942–945, 1992. 60, 125

[279] B. Szabo. Imidazoline Antihypertensive Drugs: A Critical Rreview on their Mechanism of Action. *Pharm. & Therapeutics*, 93(1):1–35, 2002. 201

[280] J. R. H. Tame, S. H. Sleigh, A . Wilkinson, and J. E. Ladbury. The Role of Water in Sequence-independent Ligand Binding by an Oligopeptide Transporter Protein. *Nat. Struct. Bio.*, 3:998–1001, 1998. 170, 206

[281] Y. Tanimura and S. Mukamel. Two-dimensional Femtosecond Vibrational Spectroscopy of Liquids. *J. Chem. Phys.*, 99(12):9496–9511, 1993. 7

[282] L. Tavagnacco, U. Schnupf, P. E. Mason, M. Saboungi, A. Cesàro, and J. W. Brady. Molecular Dynamics Simulation Studies of Caffeine Aggregation in Aqueous Solution. *J. Phys. Chem. B*, 115(37):10957–10966, 2011. 119

[283] J. L. Thomas, D. J. Tobias, and Jr. A. D. MacKerell. Direct Comparisons of Experimental and Calculated Neutron Structure Factors of Pure Solvents as a Method for Force Field Validation. *J. Phys. Chem. B*, 111(45):12941–12944, 2007. 27

[284] J. Tomkinson, J. Riesz, P. Meredith, and S. F. Parker. The Vibrational Spectrum of Indole: An Inelastic Neutron Scattering Study. *J. Chem. Phys*, 345:230–238, 2008. 59

[285] K. Tozawa, S. J. Ferguson, C. Redfield, and L. J. Smith. Comparison of the Backbone Dynamics of Wild-type Hydrogenobacter Thermophilus Cytochrome c_{552} and its b-type variant. *J. Biomol. NMR*, 62(2):221–231, 2015. 7

[286] J. Turner and A. K. Soper. The Effect of Apolar Solutes on Water Structure: Alcohols and Tetraalkylammonium Ions. *J. Chem. Phys.*, 101(7):6116–6125, 1994. 73, 89, 203

[287] J. P. Ulmschneider, M. B. Ulmschneider, and A. Di Nola. Monte Carlo vs Molecular Dynamics for All-Atom Polypeptide Folding Simulations. *J. Phys. Chem. B.*, 110(33):16733–16742, 2006. 6

[288] I. I. Vaisman and M. L. Berkowitz. Local Structural Order and Molecular Associations in Water-DMSO Mixtures. Molecular Dynamics Study. *J. Am. Chem. Soc.*, 114(20):7889–7896, 1992. 74

[289] C. M. Meerman van Benthem, K. van der Meer, and J. J. Mulder. Clonidine Base: Evidence for Conjugation of Both Ring Systems. *Mol. Pharm.*, 11:667–670, 1974. 200

[290] P. C. A. van der Wel, N. D. Reed, D. V. Greathouse, and R. E. Koeppe II. Orientation and Motion of Tryptophan Interfacial Anchors in Membrane-spanning Peptides. *Biochem.*, 46:7514–7524, 2007. 60

[291] S. Varughese, Y. Azim, and G. R. Desiraju. Molecular Complexes of Alprazolam with Carboxylic Acids, Boric Acid, Boronic Acids, and Phenols. Evaluation of Supramolecular Heterosynthons Mediated by a Triazole Ring. *J. Pharm. Sci.*, 99(9):3743–3753, 2010. 169

[292] D. S. Venables and C. A. Schmuttenmaer. Spectroscopy and Dynamics of Mixtures of Water with Acetone, Acetonitrile, and Methanol. *J. Chem. Phys.*, 113(24):11222–11236, 2000. 74

[293] L. Verlet and J. J. Weis. Perturbation Theory for the Thermodynamic Properties of Simple Liquids. *Mol. Phys.*, 24(5):1013–1024, 1972. 4

[294] S. Vilar, M. Chakrabarti, and S. Costanzi. Prediction of Passive Blood-brain Partitioning: Straight Forward and Effective Classification Models Based on *in silico* Derived Physicochemical Descriptors. *J. Mol. Graphics & Modelling*, 28(8):899–903, 2010. 11

[295] G. von Heijne. Membrane Proteins: From Sequence to Structure. *Annual Rev. of Biophys. and Biomol, Structure*, 23:167–192, 1990. 59

[296] H. P. Wacklin. Neutron Reflection from Supported Lipid Membranes. *Current Opinion in Colloid & Interface Sci.*, 15(6):445–454, 2010. 210

[297] L. J. Walport, R. Morra, E. J. Mancini, and C. Redfield. ^{1}H, ^{13}C, and ^{15}N Resonance Assignments for the Tandem PHD Finger Motifs of Human CHD4. *Biomol. NMR Assignments*, 9(2):239–242, 2014. 7

[298] H. Wang and A. Ben-Naim. A Possible Involvement of Solvent-induced Interactions in Drug Design. *J. Med. Chem.*, 39(7):1531–1539, 1996. 2

[299] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.*, 25:247–260, 2006. 66, 67, 107, 109, 145, 175

[300] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and Testing of a General Amber Force Field. *J. Comp. Chem.*, 25:1157–1174, 2004. 66, 67, 107, 109, 145, 175

[301] H. WaterBeemd, G. Camenisch, G. Folkers, J. R. Chretien, and O. A. Raevsky. Estimation of Blood-brain Barrier Crossing of Drugs Using Molecular Size and Shape, and H-bonding Descriptors. *J. Drug Targeting*, 6(2):151–165, 1998. 104

[302] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *J. Chem. Phys.*, 54(12):5237–5247, 1971. 4

[303] S. H. White and W. C. Wimley. Membrane Protein Folding and Stability: Physical Principles. *Annual Rev. of Biophys. and Biomol. Structure*, 28:319–365, 1999. 59

[304] P. J. Whittle and T. L. Blundell. Protein Structure-based Drug Design. *Annual Rev. Biophys. and Biomol. Structure*, 23(1):349–375, 1994. 1

[305] B. Wilson. Brain Targeting PBCA Nanoparticles and the Blood-brain Barrier. *Nanomed.*, 4:499–502, 2009. 3, 9

[306] T. B. Woolf, A. Grossfield, and J. G. Pearson. Indoles at Interfaces: Calculations of Electrostatic Effects with Density. *Internat. J. Quant. Chem.*, 75:197–206, 1998. 60

[307] J. D. Worley and I. M. Klotz. Near-Infrared Spectra of H_2O-D_2O Solutions. *J. Chem. Phys.*, 45(8):2868–2871, 1966. 7

[308] S. Y. Wu and C. M. Chiang. The Double Bromodomain-containing Chromatin Adaptor BRD4 and Transcriptional Regulation. *J. Bio. Chem.*, 282(18):13141–13145, 2007. 167

[309] www.atomsinmotioncom. June 2016. 42

[310] www.cambridgemedchemconsulting.com. June 2016. 10

[311] www.discoverymanagementsolutions.com. June 2016. 3

[312] Y. J. Yan and S. Mukamel. Pulse Shaping and Coherent Raman Spectroscopy in Condensed Phases. *J. Chem. Phys.*, 94(2):997–1005, 1991. 7

[313] W. M. Yau, W. C. Wimley, K. Gawrisch, and S. H. White. The Preference of Tryptophan for Membrane Interfaces. *Biochem.*, 37(42):14713–14718, 1998. 59, 100

[314] C. G. Zhan, S. X. Deng, J. G. Skiba, B. A. Hayes, S. M. Tschampel, G. C. Shields, and D. W. Landry. First-principle Studies of Intermolecular and Intramolecular Catalysis of Protonated Cocaine. *J. Comp. Chem.*, 26:980–986, 2005. 103, 124

[315] W. Zhang, B. N. Markiewicz, R. S. Doerksen, A. B. Smith III, and F. Gai. C≡N Stretching Vibration of 5-cyanotryptophan as an Infrared Probe of Protein Local Environment: What Determines its Frequency? *Phys. Chem. Chem. Phys.*, 18:7027–7034, 2015. 125

[316] F. Zheng and C. Zhan. Modeling of Pharmacokinetics of Cocaine in Human Reveals the Feasibility for Development of Enzyme Therapies for Drugs of Abuse. *PLOS Comp. Bio.*, 8:7, 2012. 104, 138, 205

[317] J. Zielkiewicz. Structural Properties of Water: Comparison of the SPC, SPCE, TIP4P, and TIP5P Models of Water. *J. Chem. Phys.*, 123(10), 2005. 64

[318] B. V. Zlokovic. The Blood-brain Barrier in Health and Chronic Neurodegenerative Disorders. *Neuron Rev.*, 178, 2008. 9