INVESTIGATIONS OF THE SOLUTION PROPERTIES OF BOTH UNMODIFIED AND MODIFIED CYCLODEXTRINS

A thesis submitted to the

UNIVERSITY OF CAPE TOWN

in fulfillment of the requirements for the degree of

MASTER of SCIENCE

By

Mogamat Riedaa Gamieldien

B.Sc. (Hons) (University of Cape Town)

July, 2008

Supervisor: Associate Professor Kevin J. Naidoo

"Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. It is our light, not our darkness that most frightens us. We ask ourselves, Who am I to be brilliant, gorgeous, talented, fabulous? Actually, who are you *not* to be? You are a child of God. Your playing small does not serve the world. There is nothing enlightened about shrinking so that other people won't feel insecure around you. We are all meant to shine, as children do. We were born to make manifest the glory of God that is within us. It's not just in some of us; it's in everyone. And as we let our own light shine, we unconsciously give other people permission to do the same. As we are liberated from our own fear, our presence automatically liberates others."

Our Deepest Fear

By Marianne Williamson

Investigations of the Solution Properties of both Unmodified and Modified Cyclodextrins

Mogamat Riedaa Gamieldien **July 2008**

ABSTRACT

The major part of this focuses on investigating the anomalous solubility of β-cyclodextrin using time correlation functions as the main tool for analysis. The primary focus is on the conformational metrics which govern the internal motions of the three main, unmodified cyclodextrins (namely α, β and γ). The conformational space and dynamics of the interglycosidic dihedral angles, individual glucose tilting and vertical librations of each glucose unit is investigated. Also the hydrogen bonding across the glycosidic linkage is examined.

We then investigate the solution properties of cyclodextrin solutions using both ultrasonic interferometry and molecular dynamics. The ultrasonic interferometry experiments have the unprecedented ability to accurately calculate the primary hydration number for each cyclodextrin solution. It is also gives an indication of the solute-solvent interactions by using the isentropic compressibility. Molecular dynamics simulations are used to investigate the solvent structuring, using both the pair and spatial distribution functions, residence time of water in the first solvation shell and "free energy" of interaction between water and the cyclodextrin.

Finally we switch mode and study, using molecular dynamics, a modified β-cyclodextrin (termed "Capped" β-cyclodextrin). Here we explore the associative and binding propensities of three dimer conformations; two dimer conformations have a head-to-head orientation and the third has a head-to-tail orientation. The angle and distance between these CDs and the interaction energy is used to determine how these dimers associate; and the MM-PBSA methodology is used to calculate the relative binding free energy.

ACKNOWLEDGMENTS

I would particularly like to thank my supervisor, Assoc. Prof. Kevin J. Naidoo, for his advice, encouragement and enthusiasm during the course of my MSc. studies. I am grateful for his understanding and concern during the period of my illness.

My family, especially my parents, provided invaluable support during the course of my studies. My parents, Achmat and Latiefa, were always there in support in every decision I made, and nursed me during my illness. My brother, Ashraf, for his encouragement and my sister, Shamiela, for her support and for proof-reading my thesis. Thank you all.

To my lab colleagues: Chris, Richard, Gerhard, Ranga, Bhajan and Melissa – thank you for being patient tutors and for interesting discussions on the fundamentals and philosophies of not only chemistry, but also about life as we know it. I really enjoy your company!

Finally, I extend my warmest and utmost gratitude and gratefulness to the National Research Foundation (NRF) for providing me with the funding I needed to perform my studies. I would also like to thank the University of Cape Town and Kevin J. Naidoo (Computational Chemistry bursary) for providing additional financial support.

PRESENTATIONS

Part of this work has been presented at the following conference:

- Carman National Physical Chemistry Symposium, Cape Town, South Africa, 23-27 September 2007.
    a. M. Riedaa Gamieldien, Kevin J. Naidoo. Glucose Orientation and Dynamics in α-, β- and γ-Cyclodextrins Affect their Relative Aqueous Solubility.
    b. Bhajan Lal, M. Riedaa Gamieldien, Kevin J. Naidoo. A Thermoacoustic Study of α-, β- and γ-Cyclodextrins.

List of Abbreviations

CD	Cyclodextrin
CGTase	Cyclodextrin Glycosyltransferase
CHARMM	Chemistry at Harvard Macromolecular Mechanics
MM	Molecular Mechanics
MD	Molecular Dynamics
kcal	Kilocalories
HGFB	Ha, Giammona, Field and Brady carbohydrate force field for CHARMM
PHLB	Palma, Himmel, Liang and Brady carbohydrate force field for CHARMM
Å	Angstrom (10^{-10} m)
nm	Nanometre (10^{-9} m)
fs	Femtosecond (10^{-15} s)
ps	Picosecond (10^{-12} s)
ns	Nanosecond (10^{-9} s)
TCF	Time Correlation Function
NMR	Nuclear Magnetic Resonance
PDF	Pair Distribution Function
SDF	Spatial (or Solvent) Distribution Function

*** Other abbreviations not shown here will be explained in the relevant chapters.**

TABLE OF CONTENTS

INDEX OF FIGURES AND TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

Chapter 1

Introduction

Traditionally chemists have been preoccupied with covalent bonds, in particular covalent bonds involving carbon. Covalent bonds are the bonds which hold the atoms of molecules and biopolymers together. Covalent bond energies are on the order of 100 kcal/mol. These bonds do not break when proteins fold (or unfold), when carbohydrates interact with proteins or when membranes assemble. These processes are controlled by non-covalent (or non-bonded) interactions, important for interactions between non bonded molecules. Each type of non-covalent interaction is relatively weak, often with a free energy of less than *RT*; but cumulatively the energies are huge. Take for instance the hydrogen bonding that takes place in water, not only does it result in the water having such an anomalously high boiling point but it also results in solid water (ice) having a lower density than the liquid, Figure 1.1.

Figure 1.1: Hydrogen bonding between water molecules which gives water its unique properties.

Other non-covalent interactions play a chief role in processes such as protein folding. The hydrophobic effect (which is entropy driven) plays a major role in the folding of globular proteins. The folded protein is also held together by other non-covalent interactions. Non-covalent interactions were discovered by van der Waals during his analysis of gases which showed deviations from the ideal gas law; he noticed that the molecules are "sticky". There are many different types of non-covalent interactions

such as: repulsive, charge-charge, dipole, fluctuating dipole and hydrogen bonding interactions and the hydrophobic effect. Without such non-covalent interactions cellular recognition and interaction would be impossible. Gaining insight into how molecules interact with each other will give us greater knowledge to study complex events such as protein folding.

Carbohydrates play a vast number of roles in nature, from energy storage to being host molecules for modifying the physical properties of guest molecules. Firstly, carbohydrates serve as energy stores, fuels, and metabolic intermediates for an organism's survival [1]. Secondly, the monosaccharides ribose and deoxyribose form the part of the structural framework of RNA (ribonucleic acid) and DNA (deoxyribonucleic acid), the genetic material responsible for the functionality of an organism [1]. Thirdly, carbohydrates are structural elements in the cell walls of plants and bacteria. In fact, cellulose, the main constituent of plant cell walls, is one of the most abundant organic compounds in the biosphere. Hydrogen bonding between cellulose strands results in a strong and sturdy cell wall [1]. Fourth, carbohydrates are linked to many proteins and lipids, where they play key roles in mediating interactions among cells (such as cell recognition, cell-cell interaction and adhesion) and interactions between cells and other elements in the cellular environment [1]. Finally, in the early to mid 20th century, a series of cyclic oligosaccharides with interesting and unprecedented properties and functionality were discovered. These molecules, named cyclodextrins (CDs), were found to have the amazing property in that they could form inclusion complexes with insoluble drugs making them more soluble in aqueous media. After the discovery of their incredible including power, CDs were labelled very seductive molecules, appealing to investigators in both pure research and applied technologies. They are most unusual in that they transcend traditional barriers separating many disciplines of chemical sciences [2].

1.1 Carbohydrates in Brief

Carbohydrates are aldehyde or ketone compounds with multiple hydroxyl groups. They make up most of the organic matter on earth because they play extensive roles in all forms of life [1]. Carbohydrates are generally composed of carbon, hydrogen and oxygen atoms and have the composition $C_xH_{2y}O_y$ (literally a "carbon hydrate"). The simplest carbohydrates are saccharides, the smallest of these being the

monosaccharides. Monosaccharides form the building blocks of more complex carbohydrates, much like amino acids form the building blocks of proteins. For monosaccharides with more than four carbons, there exists more than one chiral centre, diastereomers result in the various monosaccharides. Take as an example, glucose and galactose monosaccharides which are C4 epimers, i.e. their configuration at C4 is differ, Figure 1.2 [1]. The numbering of the monosaccharides begins at the carbonyl end, in the open chain form, Figure 1.2. Carbohydrates with ten or less linked monosaccharides are often called oligosaccharides, whereas polysaccharides is the name given when there are more than ten linked monosaccharides [1].

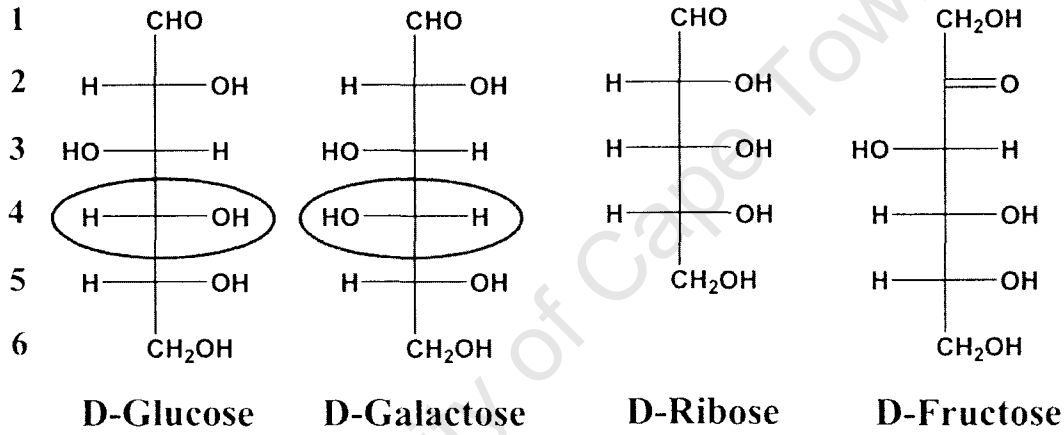

Figure 1.2. The numbering of monosaccharides, the C4 epimers (glucose and galactose), the five carbon monosaccharide, ribose, and the ketone monosaccharide fructose.

The open (or linear) chain forms, for the pentose and hexose monosaccharides, do not readily exist in solution, but rather they cyclise to form either five-membered (furanose) or six-membered (pyranose) rings [1], Figure 1.3.

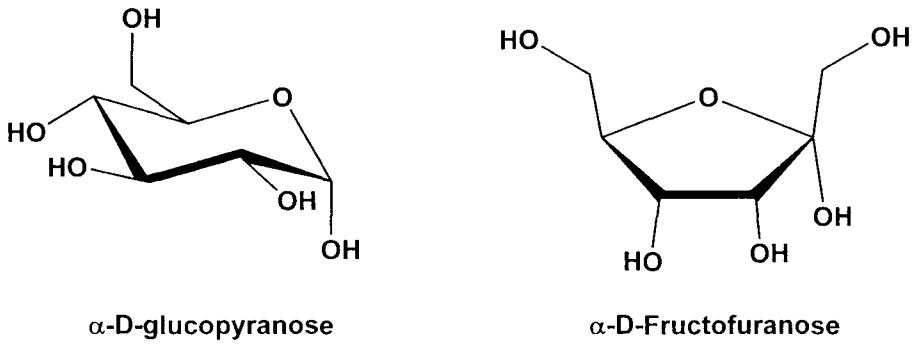

Figure 1.3: Six-membered (pyranose) and five-membered (furanose) monosaccharide rings.

In this thesis, the main focus is on the monosaccharide glucose, primarily in the pyranose ring form. The pyranose ring exists mainly in the chair conformation either with a $^4C_1$ or $^1C_4$ conformation, Figure 1.4.

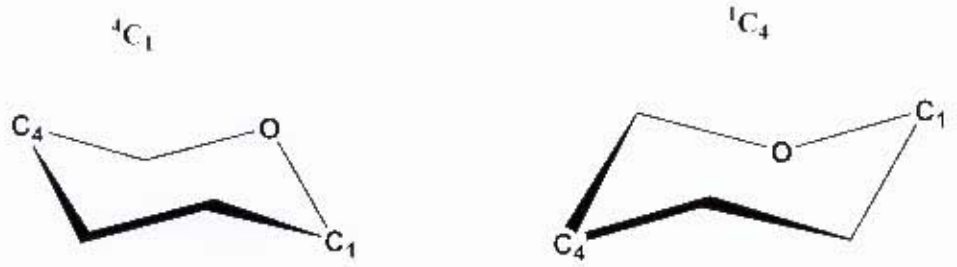

Figure 1.4: There are several types of conformations that a six-membered ring can form. The most stable and often observed conformation is the chair conformation. There are two types of chair conformations, namely the $^4C_1$ and $^1C_4$ conformations. Only the $^4C_1$ conformation is focused on in this thesis.

Due to the multiple hydroxyl groups in a monosaccharide, there are hence a multiple array of possible linkages between monosaccharides, i.e. between the C1-OH and C4'-OH (the prime denotes the adjacent monosaccharide), C1-OH and C6'-OH, etc. However, we mainly focus on the linkage between the C1-OH and C4-OH' (called the *1-4* linkages). The stereochemical configuration of the anomeric carbon (C1 for glucose) can take on two forms, namely the α and β, Figure 1.5, and is the major descriptor in distinguishing between the *1-4* linkages [1].

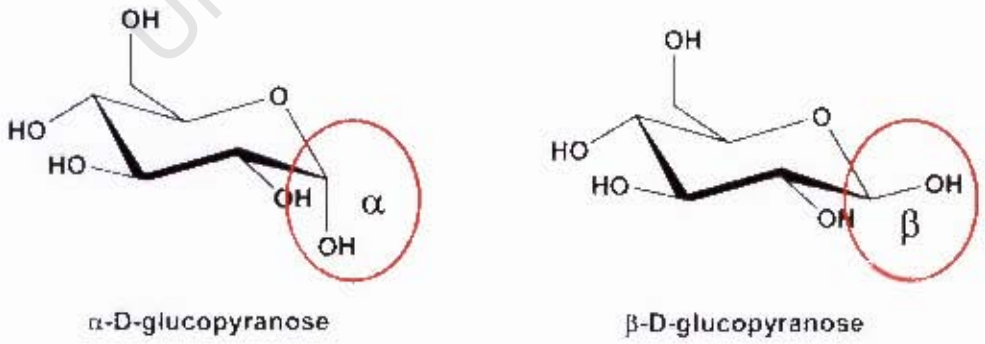

Figure 1.5: The two configurations (α and β) found at the anomeric (C1) carbon.

The *α(1-4)* linkage, the linkage in the molecules dealt with in this thesis, results in the carbohydrate forming a helical conformation, whereas the *β(1-4)* linkage results in a

linear structure and is the linkage found mainly in cellulose, the main structural component of cell walls, Figure 1.6. The free hydroxyl group, from the anomeric carbon, of carbohydrate is termed the reducing end of the carbohydrate, Figure 1.6 [1].

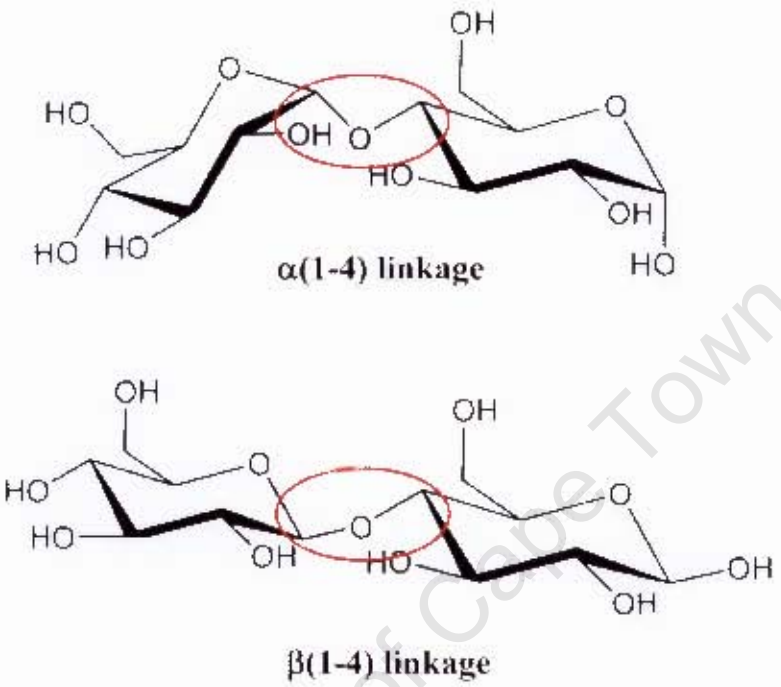

Figure 1.6: Two of the common types of linkages between monosaccharides. The α(1-4) linkage is linkage dealt with in this thesis; β(1-4) linkage is the linkage found in cellulose, it results in a cellulose having a linear conformation which is important from a structural stand point

1.2 Cyclodextrins

Cyclodextrins (CDs or cyclomaltooligosaccharides) are non-reducing cyclic oligosaccharides most commonly composed of 6, 7 or 8 glucose units bearing the names α-, β- and γ-CD, respectively; although larger CDs have been found. The glucose units are linked via α(1-4) glycosidic linkages, giving CDs a truncated cone shape with a central cavity. Other, usually smaller, molecules (called guests) can enter the CD cavity and form inclusion complexes with them; CDs have the ability to include a large variety of molecules in their cavity. CDs have found numerous and invaluable applications in many industrial sectors such as the pharmaceutical, food and cosmetics and agrochemical industries. Among all of the possible potential hosts, the CDs seem to be the most important ones, for the following reasons [3]:

1. They are semi-natural products produced from a renewable natural material (starch) by a relatively simple enzymatic conversion.
2. They are produced in the thousands of tons per year amounts by environmentally friendly technologies.
3. As a result from point 2, their initially high prices have dropped significantly where they now have become acceptable for most industrial applications.
4. Through their inclusion complex forming ability, important properties of the complexed host molecules can be modified significantly. This unprecedented "molecular encapsulation" is already widely utilised in many industrial products, technologies and analytical methods.
5. Any of their toxic effects is of secondary character and can be eliminated by selecting the appropriate CD type or modified CD or mode of application.
6. As a result of point 5, CDs can be consumed by humans as ingredients of drugs, foods or cosmetics.

According to Stoddart [4], "Cyclodextrins are all-purpose molecular containers for organic, inorganic, organometallic and metalloorganic compounds that may be neutral, cationic, anionic or even radical."

Discovery and History of Cyclodextrins

The first reference to a substance, which later proved to be a CD, was published by Villiers [5], in 1891. By digesting starch with *Bacillus amylobacter*, he isolated a crystalline compound and determined its composition to be $(C_6H_{10}O_5)_2 \bullet 3H_2O$, and named the product "cellulosine" because it resembled cellulose with regard to its resistance to acidic hydrolysis and because it did not show any reducing properties. Twelve years later, Schardinger [6], published a report in which he digested starch with a microorganism, which he named *Bacillus macerans*, resulting in the formation of small amounts of two different crystalline products. These substances seemed to be identical with the "cellulosine" of Villiers. Schardinger continued to study these two different crystallised dextrins, he observed that they formed characteristic iodine adducts upon the addition of iodine-iodide solution. The major crystalline product was the so-called α-dextrin. The simplest means to distinguish between the α- and β-dextrins was the iodine reaction. The crystalline α-dextrin/iodine complex in thin layers is blue when damp and grey-green when dry, while the crystalline β-dextrin/iodine complex is brownish (red-brown) when damp or dry [7]. It can be said

that the fundamentals of CD chemistry were laid down by Schardinger. Pringsheim and coworkers [8, 9] showed that the crystalline dextrins and their acetates have a high tendency to form complexes with various organic compounds.

In the 1930s it was found that the crystalline Schardinger dextrins are built from maltose units and contain only α(1-4) glycosidic linkages. Enzymatic hydrolysis [10], acetolysis with acetyl bromide and hydrolysis of the permethylated dextrins [11, 12] afforded these results; in 1936 the cyclic nature of these crystalline dextrins was postulated [13], and in 1948-50, the γ-CD had been discovered and its structure elucidated [14]. In the 1950s two groups, D. French et al. and F. Cramer et al., began to work intensively on the enzymatic production of CDs, on fractionating them to pure components and on characterising their true chemical and physical properties. French [15] discovered that there are even larger CDs, while Cramer's group mainly directed their attention toward the inclusion complex forming properties of the CDs [16]. By the end of the 1960s, the methods for the laboratory-scale preparation of CDs, their structure, chemical and physical properties had been discovered.

After it was discovered that CDs were not toxic to humans, the industrial production of CDs began to increase. In 1970 the price of 1 kg of β-CD was about $2000 US, and it was only available as a rare fine chemical. 25 years later, worldwide more than half a dozen companies were producing CDs. Their total output is in excess of 1000 tons/year, and the price of the key product, β-CD was (and still is) only several dollars per kilogram [3]. Also α- and γ-CD as well as other derivatives are all produced industrially. However, γ-CD is not readily produced by CGTase (cyclodextrin glycosyltranferase, a CD producing enzyme, Table 1.1) and thus γ-CD has to be synthetically made, hence the much higher price.

General Cyclodextrin Structure

The three main CDs are crystalline, homogeneous, nonhygroscopic substances, which are torus-like macrocyclic rings built up from only glucopyranose (glucose) units in the $^4C_1$ conformation [3]. The α-cyclodextrin (Schardinger's α-dextrin, cyclomaltohexaose, α-CD) comprises of six glucopyranose units, β-cyclodextrin (Schardinger's β-dextrin, cyclomaltoheptaose, β-CD) comprises of seven such units and γ-cyclodextrin (Schardinger's γ-dextrin, cyclomaltooctaose, γ-CD) comprises of eight such units, Figure 1.7a. The nomenclature for CDs is not exact, but for the three

main CDs their names remain (as it is unambiguous) as α-, β- and γ-CD. The important properties of the CDs are summarised in Table 1.1.

As a consequence of the $^4C_1$ conformation (Figure 1.4) of the glucopyranose units, all the secondary hydroxyl groups (i.e. the hydroxyl groups attached to C2 and C3) are situated on the one (upper and wider) edge of the two edges of the macrocyclic ring, whereas all the primary hydroxyl groups (i.e. the hydroxyl groups attached to C6) are placed on the other (lower and narrower) edge, Figure 1.7b. The C2 hydroxyl group of one glucopyranose unit can form a hydrogen bond with the C3 hydroxyl group on an adjacent glucopyranose unit.

Figure 1.7: (a) Shows the three common CDs, namely α-, β- and γ-CDs. (b) Shows the shape of the CDs, with the secondary hydroxyls lining the upper (wider) rim and the primary hydroxyls lining the lower (narrower) rim of the CD.

The macrocyclic ring, in reality, can be best described as having a truncated cone shape. The inside cavity is lined with the H3 and H5 protons and the lone pairs of the glycosidic linkage oxygen atoms. The lone pairs of the glycosidic linkage oxygens are directed toward the inside of the cavity producing a high electron density and gives it some Lewis base character [3]. Thus the cavity can be endowed with a hydrophobic

character and the exterior of CD (which has the secondary and primary hydroxyl groups) is hydrophilic in nature, Figure 1.8. A hydrophobic drug can thus be included in the cavity and its aqueous solubility increased when complexed with the CD (because of the hydrophilic exterior of the CD).

	α-CD	**β-CD**	**γ-CD**
No. of glucose units	6	7	8
Molecular Weight (g/mol)	972	1135	1297
Solubility in water at 25°C (g/100ml)	14.5	1.85	23.2
Cavity diametre (Å)	4.7-5.3	6.0-6.5	7.5-8.3
Height of torus (Å)	7.9 (0.1)	7.9 (0.1)	7.9 (0.1)
Diametre of outer periphery	14.6 (0.4)	15.4 (0.4)	17.5 (0.4)
Approx. volume of cavity ($Å^3$)	174	262	427
Crystal water, wt %	10.2	13.2-14.5	8.13-17.7
Diffusion constant at 40°C	3.443	3.224	3.000
Hydrolysis by *A. oryzae* α-amylase	Negligible	Slow	Rapid
V_{max} value (min^{-1})	5.8	166	2300
Size of guest molecules which can be included	Small Alkyls	Alkyls and Phenyls	Large Alkyls and Phenyls
Cost (R/g)[a]	155.02	22.65	1678.29

Table 1.1: Important properties and characteristics of α-, β- and γ-CDs. Taken from reference 3.

[a] Prices obtained from Sigma-Aldrich on 6 July 2008.

The size of the cavity for each CD is of significance importance, because this determines the size of the guest molecule that can be included, Table 1.1. In the case of α-CD, it has a small cavity and can only accommodate small alkyls; β-CD, with a larger cavity, can include larger alkyls and some phenyl molecules; and γ-CD, which has the largest cavity, can include even larger alkyls and phenyl molecules. With this in mind, one can easily see the limitations of the applications for α-CD, while β- and γ-CD would have greater applicability in industry because many molecules, of interest, contain either long alkyl chains or phenyl groups. However, there exists a cost limitation when it comes to γ-CD because it has to produced synthetically (and hence the use of it in industry is economically unfeasible), thus β-CD has become the most

commonly used CD for industrial use in inclusion complexes. But, there exists a problem with the aqueous solubility of β-CD, Table 1.1.

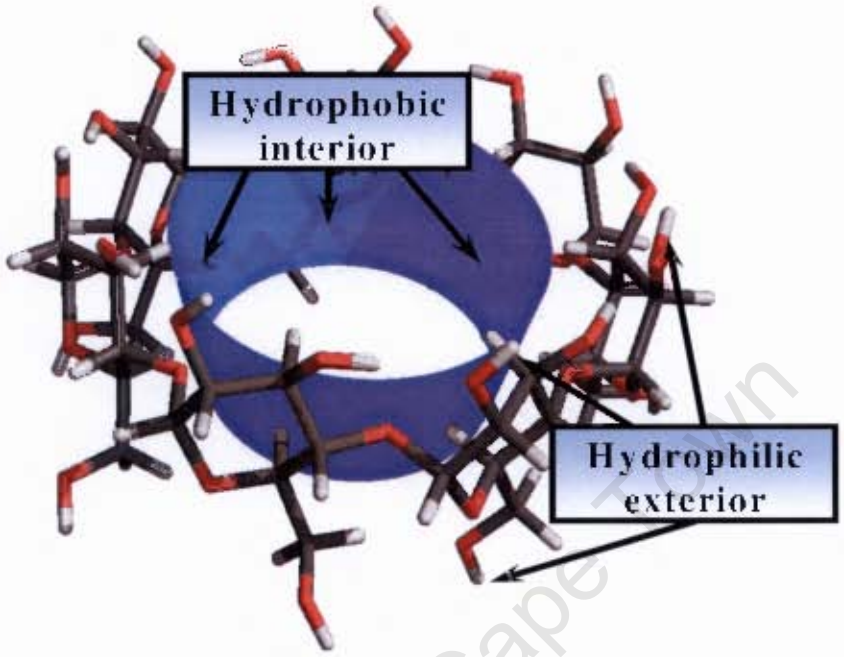

Figure 1.8: Diagram depicting the hydrophobic interior and hydrophilic exterior of the CD.

The anomalous solubility trend for the three main CDs is that one would expect the solubility to increase with increasing size (and flexibility of the macrocyclic ring), but it was found that β-CD (the intermediate in size) is at least 9 times less soluble than α-CD (the smallest in size) and 11 times less soluble than γ-CD (the largest in size), Table 1.1. Many speculations and postulations have been put forward concerning this anomalous behaviour of β-CD in solution, from the hydrogen bonding between the secondary hydroxyls groups [3] to the asymmetric arrangement of seven glucose residues for β-CD. The hydrogen bonding postulate comes from observations in solid state with the use of single X-ray crystallographic studies; the postulate states that hydrogen bonding between the secondary hydroxyl groups are responsible for the anomalous solubility behaviour of β-CD. In the case of α-CD, one glucose unit is always inclined in toward the cavity of the CD and hence only four of the six hydrogen bonds between the secondary hydroxyl groups are possible, which is postulated to be the reason for its relatively high solubility. For β-CD, it is argued that all the glucose units are in the correct orientation for a complete hydrogen bond belt, between the secondary hydroxyls, around the entire CD, resulting in a more rigid

structure and hence a lower solubility. Due to the size and relative (for a lack of a better word) "floppiness" of the macrocyclic ring for γ-CD, the hydrogen bonds between the secondary hydroxyls cannot be completed around the entire CD ring, resulting in its relatively high solubility [3]. This anomalous solubility has been overcome by making derivatives of β-CD, which increases its solubility and applicability in industry.

Inclusion Properties

The great significance of CDs both in research and industry lies in their ability to selectively form inclusion complexes with other molecules, ions or even radicals [3]. This phenomenon bears the name molecular recognition while the selectivity in the formation of complexes with enantiomeric species is called chiral recognition. In aqueous solution, the slightly hydrophobic CD cavity is occupied by water molecules which are energetically unfavoured (polar-nonpolar interactions), and therefore can be readily displaced or substituted by appropriate "guest molecules" which are less polar than water, Figure 1.9.

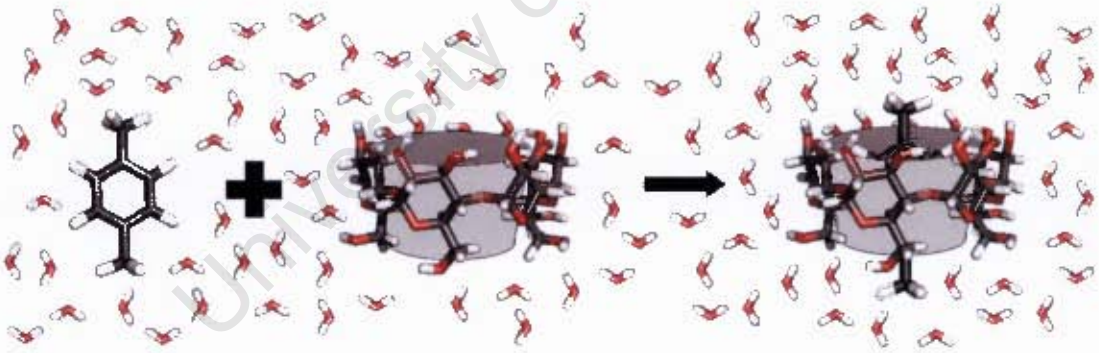

Figure 1.9: Schematic representation of CD inclusion complex formation. *P*-Xylene is the guest molecule, and the waters are shown. Adapted from reference 3.

The dissolved CD is the "host" molecule, and the "driving force" of the complex formation is the substitution of the high-enthalpy water molecules by an appropriate "guest" molecule. Most frequently the host:guest ratio is 1:1, but there also exists other ratios such as a host:guest ratio of 2:1, 1:2, etc.

The ratio of 1:1 is the essence of "molecular encapsulation". It is the simplest and most frequent case, however, 2:1, 1:2, 2:2 or even more complicated associations, and

higher order equilibria exist, almost always simultaneously [3]. The formed inclusion complex can be isolated as stable crystalline substances. Upon dissolving these complexes, an equilibrium is established between the dissociated and associated species, and this is expressed by the complex stability constant K_a. The association of the CD and guest (G) molecules, and the dissociation of the formed CD-guest complex are governed by a thermodynamic equilibrium [3].

$$CD + G \leftrightarrow CD \cdot G$$

$$K_{1:1} = \frac{[CD \cdot G]}{[CD][G]} \tag{1.1}$$

The most important primary consequences of the interaction between the poorly soluble guest and the CD in aqueous solution are as follows [3]:

- The concentration of the guest in the dissolved phase increases significantly, while the concentration of the dissolved CD decreases. This latter point is not always true, however, because ionised guests or hydrogen bonding compounds may enhance the solubility of the CD.
- The spectral properties of the guest are modified. For example, the chemical shifts of the anisotropically shielded atoms are modified in the NMR spectra. Also when achiral guests are inserted into the chiral CD cavity, they become optically active and show strong induced Cotton effects on the circular dichromism. Sometimes the maxima of the UV spectra are shifted by a few nm and the fluorescence is strongly improved, because the fluorescing molecule is transferred from the aqueous environment into a hydrophobic surrounding.
- The reactivity of the included molecule is modified. In most cases the reactivity decreases i.e. the guest is stabilised, but in many cases the CD behaves as an artificial enzyme, accelerating various reactions and modifying the reaction pathway.
- The diffusion and volatility of the included guest is decreased strongly.
- The formerly hydrophobic guest, upon complexation, becomes hydrophilic; therefore its chromatographic mobility is also modified.

- The complexed substance is molecularly dispersed in a carbohydrate matrix, forming a microcrystalline or amorphous powder. Even with gaseous guest molecules.
- The complexed substance is effectively protected against any type of reaction, except that with the hydroxyls or reactions catalysed by the CD.
- The complex is hydrophilic, easily wettable and rapidly soluble.

The two main components of the driving force of the inclusion process are the repulsive forces between the included water molecules and the hydrophobic CD cavity, and between the bulk water and the nonpolar guest. When, in an aqueous system, the formation of the complex can be detected e.g. by NMR or circular dichromism, and in solid state it can be detected by powder X-ray diffraction or single crystal diffraction.

1.3 Enzymatic Production of Cyclodextrins

There are numerous synthetic routes which one can follow when trying to synthesize CDs, but they are not nearly as efficient as an enzyme catalysed synthesis of CDs. Enzyme catalysed reactions have exact stereochemical control and can yield the desired product(s) using cheap and renewable bio-materials such as starch. Cyclodextrin glycosyltransferase (CGTase, EC 2.4.1.19) is a unique member of the α-amylase family of glycosylases with a low hydrolytic activity. It is usually considered to be an exo-acting enzyme and unable to bypass branching points. Its main products, when acting on starch, are cyclic oligosaccharides consisting of 6, 7 or 8 glucose (named α-, β- and γ-CD, respectively) and highly branched high molecular weight dextrins (CGTase limit dextrins). The CDs are produced via an intramolecular transglycosylation (cyclisation) reaction in which the CGTase cleaves an α(1-4) glycosidic linkage in the starch molecule, concomitantly linking the reducing and non-reducing ends. The enzyme also catalyses two intermolecular transglycosylation reactions: coupling, in which the ring of a CD (donor) is opened and transferred to linear oligosaccharide (acceptor), and disproportionation, in which a part of a linear oligosaccharide chain (donor) is transferred to another linear oligosaccharide (acceptor) [17, 18]. A possible explanation for the existence of CD enzymes, such as CGTases, is that by producing CDs an organism builds up an external storage form of glucose, not accessible for most other organisms because they unable to metabolize

CDs. Alternatively, CDs may protect bacterial cells against toxic compounds in the environment by forming inclusion complexes as in biological waste water treatment, where the addition of small amounts of β-CD in activated sludge's increases the tolerance level to toxic chemicals [19]. Also the availability of compounds needed for growth may improve when present in an inclusion complex with CDs [20]. These alternatives, however, do not explain the presence of the specific uptake and degradation routes for CDs.

The industrial importance of CGTases are not only for the production of CDs, but recent developments have also concentrated on the use of the CGTase catalysed coupling and disproportionation reactions for the synthesis of modified oligosaccharides by using alternative acceptor substrates. Furthermore, the applications of CGTase limit dextrins are being explored. Increasingly, the use of alternative acceptors is reported, resulting in novel glycosylated compounds [21, 22, 23]. A commercial application of this method is found in glycosylation of the intense sweetener, stevioside. This bitter compound is isolated from the leaves of the plant *Stervia rebandiana* and has a low solubility. Glycosylation decreases bitterness and increases solubility [24]. Other applications are found for the CGTase limit dextrins. Due to the inability of CGTase to bypass α(1-6) glycosidic linkages in gelatinised starches, degradation of these substrates leads to a reduction in viscosity without a corresponding decrease of the high-molecular character of starch. This CGTase limit dextrin is applied in the processes of surface sizing or coating of paper, to improve the writing quality of the paper and to obtain a glossy and well printable surface [25].

The enzyme catalysed synthesis of CDs by CGTase proceeds with the retention of the substrate's anomeric (α-) configuration. Since substitution at a chiral centre results in an inversion of configuration, catalysis must proceed through a double displacement reaction, Figure 1.10 [26]. The first step involves a protonation of the glycosidic oxygen by a general acid catalyst, creating an oxo-carbonium transition state which collapses into a covalent β(1-4) glycosyl-enzyme intermediate [27, 28]. In the second step, the covalent bond of the intermediate is cleaved and an α(1-4) is reformed with an acceptor, such as water or the C4-OH group of another sugar residue, Figure 1.10 and 1.11. CGTase has the unique ability to use the free hydroxyl group at the non-reducing end of the sugar chain as an acceptor, thus forming a cyclic oligosaccharide (cyclodextrin). The roles of the three carboxylic amino acids in this mechanism have been clarified by X-ray crystallography studies. Glu257 is the

general acid catalyst, acting as a proton donor; Asp229 serves as the nucleophile, stabilising the intermediate; and Asp328 has an important role in substrate binding. The intermediate could either be an oxo-carbonium ion which is electrostatically stabilised by a carboxylate, or involves formation of a covalent bond, in which one of the catalytic aspartates is presumed to act as a nucleophile, Figure 1.10.

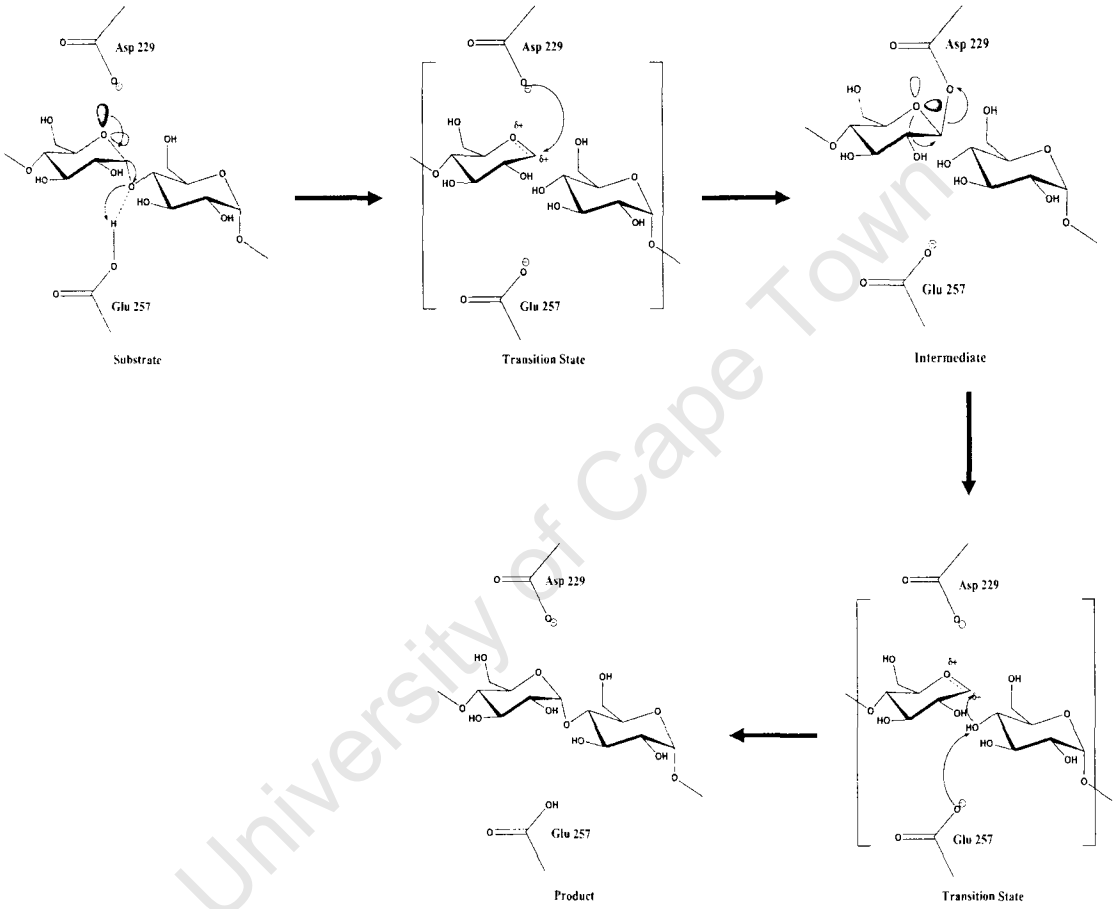

Figure 1.10: The reaction mechanism of Cyclodextrin glycosyltransferase (CGTase). Taken from reference 30.

Although initially the nature of the intermediate was disputed, it is now generally accepted that the reaction proceeds via a covalent intermediate. Clear evidence for a covalent glycosyl-enzyme intermediate has been obtained from rapid trapping studies with natural substrates. Low-temperature ^{13}C NMR experiments have provided evidence for the formation of a β-carboxyacetal ester covalent adduct between maltotetraose and porcine pancreas α-amylase [29].

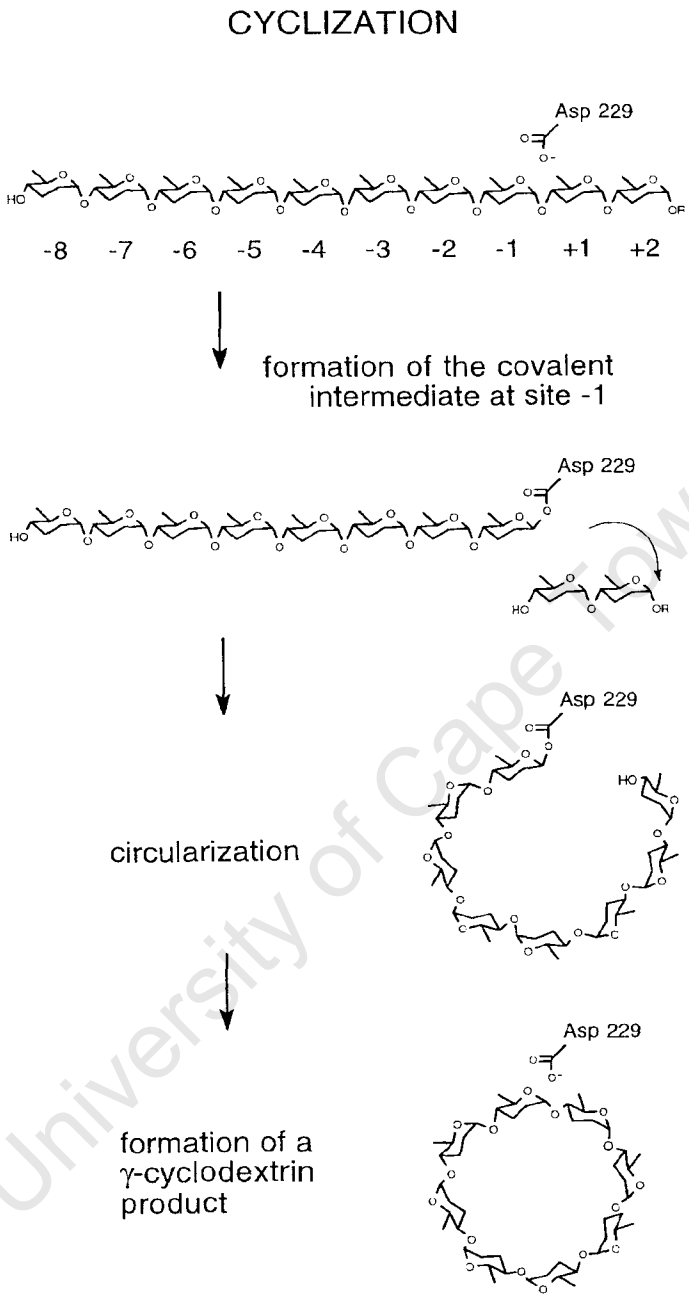

Figure 1.11: A schematic diagram for the cyclisation of a linear oligosaccharide into γ-CD. The same cyclisation procedure occurs for the other two CDs (α- and β-CD) as well. Taken from reference 31.

1.4 Modifications of Cyclodextrins

The exterior of the CD (or more specifically the secondary and primary hydroxyl groups) interacts with the solvent and thus is the determining factor in the overall solubility of the CD-guest complex has in an aqueous medium. In many cases the contribution of the outer surface to the solubility of the complex is not sufficient to

obtain the desired solubility properties. Modification of the hydroxyl groups of the outer surface of the CD can alter the solubility properties markedly. Substitution with neutral groups, such as hydroxylpropyl, carboxymethyl or tertiary amine groups, or an ionic group, such as a quaternary amine group, increases the solubility of the CD to 60% or greater in water, which is particularly true for hydroxylpropyl β-CD. This has been a highly successful modification to β-CD, and it has been widely exploited in industry. The methylated form of β-CD has also proved to be a successful modification for increasing the solubility of β-CD as well; methylated modifications to the hydroxyl groups can be random (RAMEB), two hydroxyl groups can be modified (DIMEB) and all three hydroxyl groups can be methylated (TRIMEB). All increase the solubility of β-CD dramatically and are extensively used in supramolecular chemistry and industry to increase the aqueous solubility of β-CD. Modification with aliphatic groups, such as hexyl groups, or nearly complete substitution with smaller groups, such as acetyl groups, result in decreased, nondetectable solubility in water and increased solubility in organic solvents [32]. Modification with a naphthalene ring to one of the C3 hydroxyl groups is also possible and these CDs have been called "Capped" CDs because the naphthalene ring (or some other bulky aromatic functional group) behaves much like a cap on the upper (wider) rim of the CD. It is also possible for the pendant arm (or naphthalene or bulky aromatic ring) to self include inside the CD cavity, this type of behaviour is called self inclusion of the pendant arm (or naphthalene ring) [38].

The modifications made to CDs offer both enormous opportunities and challenges for chemists and thus specialised synthetic techniques are required. The challenges that are provided are the hydrophobic cavity and a large number hydroxyl groups. Hydroxyl groups present at the 2-, 3- and 6-positions compete for the reagent and make selective modification extremely difficult [33]. The hydrophobic cavity often has a tendency to interfere with the well laid out plan of a chemist by complexing with the reagent to direct its activity to an unexpected place [34].

CDs are modified for a variety of reasons ranging from achieving solubility in a desired solvent to investigating the mechanism of enzyme-catalysed reactions. The strategy for modification depends on the purpose for the final product. For example, if a highly water soluble CD is desired for application in a drug formulation, then a random conversion of hydroxyl groups to sulfate groups can be easily achieved and the product will have the desired solubility in water [35, 36]. Similarly, if a CD with a

high solubility in organic solvents is desired, one can easily convert the hydroxyl groups to silyl ethers in a random fashion [37]. There is a general scheme for selectively modifying the hydroxyl groups of CDs, Figure 1.12.

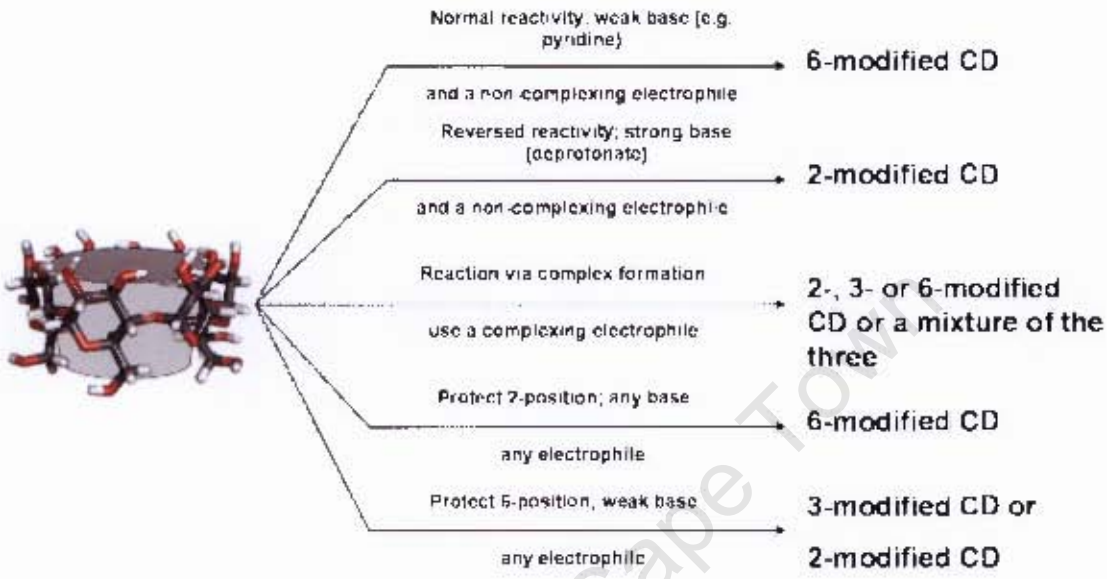

Figure 1.12: Brief scheme and overview for selectively modifying CDs. Taken from reference 38.

These modification methods usually fall into three classes: (1) the "clever" method, where the chemistry of the CD is exploited to get the desired product by the shortest route; (2) the "long" method, where a series of protection and deprotection steps have taken place in order to selectively reach positions which would otherwise not be selectively accessible; (3) the "sledgehammer" method, where the CD is indiscriminately reacted to give a mixture of products and then the desired product is separated out from the other isomers by chromatographic methods [38].

For further reading on selectively modifying CDs, there is an outstanding review article [38] which has further references.

1.5 Techniques for Studying Cyclodextrins

In 1891 when Villiers [3] discovered CDs, he used elemental analysis to determine the composition of the crystalline compound he obtained from bacterial degradation of starch. He also noted that it did not posses a reducing end, but he never knew that the saccharide he had isolated was a cyclic oligosaccharide. Freudenberg

[11] and Cramer [12] later postulated that these oligosaccharides were cyclic in nature and with use of X-ray crystallography and NMR this was proved.

CD as a host molecule has become one the most popular molecules to use in industry and thus, as a means of quality control, structural determining techniques have become prominent for studying CDs and their complexes. The most obvious and widely used are: X-ray diffraction, Nuclear Magnetic Resonance (NMR) spectroscopy and computational methods. Structural characterisation is of particular importance for suparmolecular host – guest complexes, which are the basis for most CD applications in medicine, catalysis, or in food chemistry, separation and sensor technology. Pharmaceutical uses of CDs for drug protection or targeting now legally require structural characterisation of the administered compounds [39].

Single crystal X-ray diffraction is a commonly used supramolecular technique for elucidating the structure of the CD host – guest complex in solid state. This method is an important means of quality control for solid state compounds in the pharmaceutical, catalysis and food chemistry industries [39]. However, it is not possible, using this technique, to study the solution properties of the CD supramolecular host – guest complexes, and hence NMR spectroscopy and computational methods are used to study CDs and their complexes in solution state.

NMR spectroscopy has become the most important tool for structural elucidation of organic compounds, particularly in solution state [39]. NMR is also becoming an important method for in vitro, in future perhaps even for in vivo, studies of CD interactions with biological macromolecules such as nucleic acids, proteins or cell membranes [39]. The most obvious incentive, however, to use NMR techniques for the investigation of CD complexes is the interest to understand the driving forces and binding modes in these non-covalent associations, and then to make optimal use of these factors for new applications [39]. It should be remembered that the association between the CD and a guest molecule is of a hydrophobic nature and thus requires a liquid matrix, which again emphasizes the significant role that NMR spectroscopy plays for studying CD host – guest complexes. The advent of high-field instruments, and in particular of 2-D methods, which were applied to CDs from 1987 on [40-42], has completely changed the situation for studying the 3-D structure of CDs and their complexes. The possibilities of the now available NMR techniques are far from being exploited. The spectacular advances of NMR techniques during the last years have led to a much more detailed structural elucidation and solution properties of CDs and

their complexes. Nuclear Overhauser effects (NOEs) have become a major tool in structural studies of CDs [39]. Another factor greatly facilitating the elucidation of complex structures is the development of computer-aided molecular modelling methods, based on reliable force fields. NMR is a much too indirect method to derive complex three-dimensional structures without the use of realistic models, which then must be tested against experimentally observed NMR data. This fitting step also is now significantly improved by the availability of modern computer techniques. Only for selected simpler questions can one safely proceed without using more rigorous computational methods [39].

There have been a number of reviews and reports dealing with computational methods for studying CDs [43-48], but an overview of the application of computational chemistry for studying CDs and their complexes is provided here. Computational methods for studying CDs and their complexes have become increasingly popular. Both quantum mechanical based techniques and application of empirical force field methods have been used, with the most applied method being molecular mechanics and molecular dynamics [43]. A significant number of these applications have been directed toward understanding the shapes of CDs, especially for comparison to those structures derived form X-ray crystallography studies. Assessing CD structural features such as geometry, charge distributions, lipophilicity etc. represents the largest percentage of published papers where molecular modelling tools have been applied to issues in CD chemistry [43]. The energetics of host – guest binding has been investigated using both quantum- and force field-based methods. Another widely used application involves assessing the structure and dynamics of host – guest complexes. Some applications have been used to explain observed chemical reactions taking place inside the cavity as well as to rationalise the catalytic abilities of the CD microenvironment [43]. Most studies, though, have been directed toward understanding the regioselectivity of host – guest binding using relatively simple nonpolar guests as probes [43]. Related to this are the many applications of molecular modelling tools, which have been used to help scientists understand their spectroscopic data. Most applications have been used to interpret NMR data, especially Overhauser enhancements from ROESY spectra, by providing working models related to averaged locations and orientations of included guests and of the CD itself. Many studies have also addressed the molecular recognition properties of CDs. In these studies, researchers are interested in the ability of the CDs to selectively

bind stereoisomers including diastereomers but more interestingly enantiomers. The enantiodiscriminating abilities of CDs have been studied especially by separation scientists interested in the chromatographic properties of CDs, and most of those studies used molecular mechanics and dynamics to compute differential energies of analyte binding and to describe the forces responsible for chiral recognition [43]. Recently Naidoo et al. have showed similar trends between their computational and NMR results for the three main unmodified (α-, β- and γ-) CDs. They studied the diffusion of the CDs, both the computational and NMR experimental results showed an anomalous behaviour for the diffusion coefficient of β-CD. A further look into this anomalous behaviour resulted in them showing that in fact β-CD has a higher water probability density around it, thus it diffuses slower than theoretically predicted by the Einstein-Stokes theorem [48]. Two more recent papers, also investigating the anomalous solubility of β-CD, have focused their research on the results obtained by Naidoo et al [48]. Molecular dynamics was also employed in these studies using periodic boundary conditions. To date, it is computationally unfeasible to study CDs, including the waters and periodic boundary conditions, quantum mechanical simulations [43]. Cai et al. [57] showed that the probability of waters around the CDs is greatest for β-CD, which can rationalized by β-CD having greater macrocyclic rigidity conferred by the participating intramolecular hydrogen bonds and the higher density of water molecules of lesser mobility. The hydration free energy, for α-, β- and γ-CD, was also computed using the free energy perturbation method. The hydration free energy was shown to increase with the number of glucose units, thereby suggesting that the anomalous solubility of β-CD cannot be explained by its hydration free energy alone [57]. In the paper by Perez-Miron et al. [58], all possible methylated β-CDs with C7 symmetry, in solution and vacuum, were studied using molecular dynamics. CD flexibility increased with the degree of methylation, very likely due to the concomitant reduction in the number of possible intramolecular hydrogen bonds. Solvation energy was computed using the MM-GBSA method and was found to increase as more methyl groups were added to the CD. An analysis of the radial distribution functions was used to determine the distribution of the solvent around the O2, O3 and O6 atoms. The number of solvent molecules around these oxygen atoms decreased with an increase in the degree of methylation. The entropic, ΔS, contribution from the solvent thus becomes more positive with an increased

number of methyl groups on the CD and, consequently, the overall free energy, ΔG, in water diminishes [58].

The following excellent review articles [39, 43] are suggested for the reader who is more interested in the above mentioned techniques and their uses in the study of CDs.

1.6 Applications of Cyclodextrins

CD inclusion complexes are used to obtain certain benefits from the included molecule, which result from complexation with the CDs. These include alteration of the solubility of the guest molecule, stabilisation against the effects of light, heat and oxidation, masking of unwanted physiological effects, reduction of volatility, catalysing various organic reactions and others. In some applications, more than one benefit is obtained by complexation with CDs [49]. Obviously all the applications of CDs and there derivatives cannot be discussed and suggested further reading and references I refer the reader to [49, 50].

Control of Solubility

Hespiridin is found in the juice of some oranges and imparts an undesirable cloudiness to the juice of canned orange slices. By adding β-CD to the juice resulted in the complexation and thus solubilisation of the hespiridin and a clear syrup was formed [51]. In addition to solubilising the hespiridin, the bitter taste of the hespiridin was masked.

Process Aids

CDs can also be used as process aids, where the benefits of the control of solubility of the guest are used, but the CD is removed and not present in the final product [49]. The most widely used example of the use of CDs is that of removal cholesterol from animal products such as eggs, dairy products and animal fats such as lard and tallow [52].

Stabilisation

CDs can be used to stabilise compounds. The cavity of CD is a finite space. Once this space has been occupied by a molecule of the guest, other molecules are excluded

from occupying the space at the same time. This prevents interaction and reaction with other molecules. Although the ends of the cavity are open, some steric hindrance is provided to prevent the approach of other molecules at the exposed portion of the molecule. Association with the cavity or hydroxyl groups surrounding the cavity can also stabilise the guest in its less reactive forms [49]. Penicillin G is stabilised in aqueous solution in chloroacetate buffer with hydroxypropyl β-CD [53]. The rate of degradation is about 9 times slower for the complex as compared to the uncomplexed drug.

Masking the Effects of the Guest

When a guest is included in a molecule of CD, it is isolated and prevented from coming into contact with surfaces of the body where it could cause unwanted side effects such as irritation or an off flavour. Release of the guest from the complex is slow. As a result, the amount of free guest is lower than if the free guest alone is used so that the amount of guest to elicit a physiological response is reduced, resulting in masking or decrease in the intensity of the unwanted side effects of the guest [49]. The irritating or toxic effects of insecticides can be reduced or eliminated by complexing the insecticide with CDs. Azinphos-methyl forms an odourless complex with β-CD [54]. No systemic toxic effects were observed when the insecticide was administered dermally at a dose 4000 mg/kg in the complex, while the free insecticide shows toxicity when administered dermally at 17.84 mg/kg. Complexation resulted also in a 3.8 fold increase in solubility, and the insecticidal activity was comparable to that of the commercial formulation.

Reduction of Volatility

Compounds can be complexed with CDs to reduce their volatility. Interaction of the guest with CD produces a higher energy barrier to volatilise. A dry CD powder can be used in products such as menstrual products, diapers, tissues, paper towels, etc [55]. A small particle size of less than 12 μm works best. As the particle size becomes smaller, the surface area increases with respect to the volume, resulting in more rapid dissolution rates. Because of the rapid dissolution rates, only small amounts of water are needed for dissolution and dispersion of the molecules of CD, making them effective scavengers of unwanted and unpleasant odours. Perfumes can also be

complexed with the CD, which is released upon dissolution of the complex and displacement by odiferous compounds [49].

Organic Reactions Mediated by Cyclodextrins

The effects of CDs in the organic reactions are divided into two types. The first is the effect on covalent bonds where the reaction proceeds according to the Michaelis-Menten type reaction. CD and the reactant initially form a CD-reactant reaction intermediate involving a covalent bond which then leads to the product. These catalytic effects have been studied and reported as the enzyme model. Various advanced "enzyme models" and "artificial enzymes" have been envisioned through the chemical modification of CDs by using detailed molecular design. The development of techniques for the precise chemical modification of CDs has undoubtedly made an essential contribution to progress in this field. The second effect does not involve a covalent bond. The hydrophobic cavity of the CD gives the reactant access to a new reaction environment, an "extra reaction field", in which the reactivity, such as the rate or selectivity, changes. In these cases, the role of the CD is not always defined as a catalyst, but more correctly mediates the reactions [50].

Control of Fluorescence and Light Absorption

Some compounds when included in the cavity of a CD exhibit an increase or decrease in the intensity of fluorescence or light absorption. The obvious benefit, in the areas of analytical chemistry, is the increase in sensitivity of assays or to block the effects of interfering compounds [49]. An optical brightener was complexed with a methylated CD for use in a photographic application [56]. The increased fluorescent intensity made the white areas of the photograph whiter. Additional benefits of decreased migration and heat stability were also obtained.

All of the amazing applications of CDs could not be given here, but it is most certainly clear that CDs have a vast range of applications in industry. The remarkable including power of CDs has created bridges between seemingly unrelated areas of industry such as the pharmaceutical, agrochemicals, food and cosmetics industries.

1.7 Objectives

The objectives of this thesis are to investigate the anomalous solubility of β-CD and also the solution conformation and energetics of the naphthalene modified, so called "Capped" β-CDs. The former is examined using molecular dynamics simulations, NMR spectroscopy and ultrasonic interferometry, while the latter exclusively uses molecular dynamics simulations to examine the dimer conformations and binding energetics.

In Chapter 2 we discuss in general the molecular dynamics simulation methods we used. Chapter 3 gives an overview of the analytical techniques employed to investigate the data provide by these simulations and a brief discussion of the experimental techniques is provided.

Molecular dynamics simulations and NMR spectroscopy were employed to investigate the conformations of the CDs in aqueous solution, the results and analyses are discussed in Chapter 4. Molecular dynamics simulations and ultrasonic interferometry were used to investigate the solution properties, such as hydration number, "free energy" of interaction between the CDs and water, pair and spatial distribution functions of the CDs and the results are discussed in Chapter 5. Chapter 6 deals with the naphthalene modified or "Capped" *β*-CD, we examine and discuss the three dimer conformations, solvent structuring around the monomers and the energetics and binding propensities in solution.

Finally, concluding remarks and comments on the future work appear in Chapter 7.

References

1. (a) J. M. Berg, J. L. Tymoczko, L. Stryer. *Biochemistry*, 5th Ed., W. H. Freeman and Company, New York, **2001**, Chapter 11. (b) N. Sharon. *Sci. Am.* **1980**, *245(5)*, 90. (c) A. Varki, R. Cummings, J. Esko, H. Freeze, G. Hart, J. Marth. *Essentials of Glycobiology.* Cold Spring Harbor Laboratory Press, **1999**. (d) H. S. El Khadem. *Carbohydrate Chemistry.* Academic Press, **1988**.
2. V. T. D'Souza, K. B. Lipkowitz. *Chem. Rev.* **1998**, *98*, Editorial.
3. J. Szejtli. *Chem. Rev.* **1998**, *98(5)*, 1743.
4. J. F. Stoddart. *Carbohydr. Res.* **1989**, *192*, xii.
5. A. Villiers. *Compt. Rend.* **1891**, *112*, 536.
6. F. Schardinger. *Z. Unters. Nahr. U. Genussum.* **1903**, *6*, 865.
7. F. Schardinger. *Zentralbl. Bakteriol. Parasitenk. Abt. 2* **1905**, *14*, 772.
8. H. Pringsheim. *Chemistry of the Saccharides*, McGraw-Hill, New York, **1932**.
9. H. Pringsheim. *A Comprehensive Survey of Starch Chemistry*, Walton, R. P., Ed., Chemical Catalogue Co., Inc., New York, NY, **1928**.
10. K. Freudenberg, W. Rapp. *Ber. Dtsch. Chem. Ges.* **1932**, *65*, 69.
11. K. Freudenberg, H. Boppel, M. Meyer-Delius. *Naturwissenschften* **1938**, *26*, 123.
12. K. Freudenberg, M. Meyer-Delius. *Ber. Dtsch. Chem. Ges.* **1938**, *71*, 1596.
13. K. Freudenberg, G. Blomquist, L. Ewald, K. Soff. *Ber. Dtsch. Chem. Ges.* **1936**, *69*, 1258.
14. K. Freudenberg, F. Cramer. *Z. Naturforsch.* **1948**, *3b*, 464.
15. D. French. *Adv. Carbohydr. Chem.* **1957**, *12*, 189.
16. F. Cramer. *Einschulssverbindungen (Inclusion Compounds)*, Springer-Verlag, Berlin, **1954**.
17. A. Nakamura, K. Haga, K. Yamane. *Biochemistry* **1993**, *32*, 6624.
18. B. A. van der Veen, G. J. W. M. van Alebeek, J.C. M. Uitdehaag, B. W. Dijkstra, L. Dijkhuizen. *Eur. J. Biochem.* **2000c**, *267*, 658.
19. M. Allegre, A. Deratani. *Agro Fd. Ind.* **1994**, Jan/Feb, 9
20. F. Aeckersberg, F. Bak, F. Widdel. *Arch. Micribiol.* **1991**, *156*, 5.
21. T. Kometani, Y. Terada, T. Nishimura, S. Okada. *Biosci. Biotechnol. Biochem.* **1994**, *58*, 1990.

22. T. Kometani, T. Nishimura, T. Nakae, H. Takii, S. Okada. *Biosci. Biotechnol. Biochem.* **1996a**, *60*, 645

23. T. Kometani, Y. Terada, T. Nishimura, T. Nakae, H. Takii, S. Okada. *Biosci. Biotechnol. Biochem.* **1996b**, *60*, 1176.

24. S. Pedersen, L. Dijkhuizen, B. W. Dijkstra, B. F. Jensen, S. T. Jorgensen. *Chemtech* **1995**, *25*, 19.

25. P. M. Bruinenberg, A. C. Hulst, A. Faber, R. H. Voogd. Inventors (**1996**). *A process for surface sizing or coating paper*. The Netherlands. EP690170A1.

26. D. E. Koshland. *Biol. Rev.* **1953**, *28*, 416.

27. J. D. McCarter, S. G. Withers. *Curr. Opin. Struct. Biol.* **1994**, *4*, 885.

28. J. D. McCarter, S. G. Withers. *J. Am. Chem. Soc.* **1996**, *118*, 241.

29. B. Y. Tao, P. J. Reilly, J. F. Robyt. *Biochim. Biophys. Acta* **1989**, *995*, 15729.

30. J. C. M. Uitdehaag, R. Mosi, K. H. Kalk, B. A. van der Veen, L. Dijkhuizen, S. G. Withers, B. W. Dijkstra. *Nature Structural Biology* **1999**, 6(5), 432.

31. J. C. M. Uitdehaag, K. H. Kalk, B. A. van der Veen, L. Dijkhuizen, B. W. Dijkstra. *The Journal of Biological Chemistry.* **1999**, *274(49)*, 34868.

32. A. R. Hedges. *Chem. Rev.* **1998**, *98*, 2035.

33. J. Szejtli. *Cyclodextrin Technology*, Kluwer Academic Publisher, Dordrecht, The Netherlands, **1988**.

34. A. Ueno, R. Breslow. *Tetrahedron Lett.* **1982**, *23*, 3451.

35. P. B. Weisz, M. M. Joullie, C. M. Hunter, K. M. Kumor, Z. Zhang, E. Levine, E. Macarak, D. Weiner, E. S. Barnathan. *Biochem. Pharm.* **1997**, *54*, 149.

36. J. Folk, P. B. Weisz, M. M. Joullie, W. W. Li, W. R. Ewing. *Science* **1989**, *243*, 1490.

37. D. W. Armstrong. U. S. Patent 4,539,399, **1985**.

38. A. R. Khan, P. Forgo, K. J. Stine, V. T. D'Souza. *Chem. Rev.* **1998**, *98*, 1977.

39. H. –J. Schneider, F. Hacket, V. Rüdiger, H. Ikeda. *Chem. Rev.* **1998**, *98*, 1755.

40. Y. Yamamoto, M. Onda, Y. Takahashi, Y. Inoue, R Chûjô. *Carbohydr. Res.* **1987**, *170*, 229.

41. C. M. Spencer, J. F. Stoddart, R. Zarzycki. *J. Chem. Soc. Perkin Trans. 2* **1987**, 1323.

42. I. Tabushi, T. Nabeshima, K. Yamamura, H. Fujita. *Bull. Chem. Soc. Jpn.* **1987** *60*, 3705.

43. K. B. Lipkowitz. *Chem. Rev.* **1998**, *98*, 1829.

44. L. Catoire, V. Michon, L. Monville, A. Hocquet, L. Jullien, J. Canceill, J.-M. Lehn, M. Piotti, C. Hervé du Penhoat. *Carbohydr. Res.* **1997**, *303*, 379.
45. A. Deratani, E. Renard, F. Djedaïni-Pilard, B. Perly. *J. Chem. Soc. Perkin Trans. 2*, **1997**, *419*, 133.
46. K. B. Lipkowitz, R. Coner, M. A. Peterson, A. Morreale, J. Shackleford. *J. Org. Chem.* **1998**, *63*, 732.
47. J. Qian, R. Hentschke, W. Knoll. *Langmuir* **1997**, *13*, 7092.
48. K. J. Naidoo, Y. –J. Chen, J. L. M. Jansson, G Wildmalm, A. Maliniak. *J Phys. Chem. B.* **2004**, *108*, 4236.
49. A. R. Hedges. *Chem. Rev.* **1998**, *98*, 2035.
50. K. Takahashi. *Chem. Rev.* **1998**, *98*, 2013.
51. A. Konno, M. Misaki, J. Toda, T. Wada, K. Yasumatsu. *Agric. Biol. Chem.* **1986**, *46*, 2203.
52. W. Shieh, A. Hedges. U.S. Patent 5,371,209, **1994**.
53. J. K. Ong, V. B. Sunderland, C. McDonald. *J. Pharm. Pharmacol.* **1993**, *45*, 666.
54. M. Marzona, R. Carpignano, S. Girelli, M. Dolci. *In Proceedings of the 8th International Symposium on Cyclodextrins*; J. Szejtli, L. Szente, Eds.; Kluwer Academic Publishers, Dordrecht, The Netherlands, **1996**, p. 619.
55. T. Trinh, D. Phan. WO 94/22501, **1994**.
56. P. A. Martic, T. R. Skochdopole. U.S. Patent 5,340,854, **1994**.
57. W. Cai, T. Sun, X. Shao, C. Chipot. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3236-3243.
58. J. Perez-Miron, C. Jaime, P. M. Ivanov. *Chirality.* **2008**, 20, 1127.

Chapter 2

Molecular Dynamics Simulation Methods

A molecular dynamics simulation gives a detailed insight into atomic positions, dynamics and molecular interactions – and these details can often not be determined using classical experimental techniques.

There are two main ways of studying molecules using computational methods: (i) electronic structure approach methods and (ii) force field methods. Although the conformations of molecules have been studied extensively using *ab initio* methods, with the increase in size of the molecule and the addition of explicit water molecules increases the size of the calculations to such an extent that it becomes practically implausible to use. Force field methods, which include Molecular Mechanics (MM) and Molecular Dynamics (MD) as well as Monte Carlo simulations, with accurately derived force fields, can dramatically simplify the calculation, by ignoring electronic effects.

In MM and MD (the most widely used modelling method), the energy of the system is calculated as a function of the nuclear coordinates only, using a classical "ball-and-spring" model to represent the bonds and interactions between atoms. Regardless of the simplicity of the MD calculation, appropriately designed force fields can often be more accurate for large macromolecules than unreliable low levels of quantum theory; and thus MD is an efficient and suitable choice for studying large, solvated complex systems. An advantage of the method is that a detailed time evolution of the molecular interactions and conformational changes is obtained. Thus it can be used as a structural tool for studying ambiguous experimental data and can be used in elucidating the complex roles which biomolecules play in biology.

With the increase in popularity of MD simulations, there has been a concomitant increase in the number of software programs and packages that have become available for biomolecular modelling. Commonly used molecular dynamics simulation packages are: CHARMM [1] (the program used in this thesis), AMBER [2], GROMOS [3] and DL_POLY [4]; however many institutions have developed their own in-house MD programs.

This chapter provides a brief overview of the MM and MD simulation methods and their application to the study of CDs. The focus is on the techniques used in this thesis. There are many introductory texts available on the theory of MD and its application to inorganic and biological systems [5, 6, 7, 8].

2.1 Background

The complete description of systems on the atomic and molecular scale is provided by quantum mechanical theory. However, species consisting mainly of heavy atoms, a great many properties may be successfully calculated without the use of the computationally intensive task of solving the Schrödinger equation for a realistically sized problem.

This convenient simplification is achieved by the application of the Born-Oppenheimer approximation [9], which treats the nuclei and electrons separately. The molecular Hamiltonian may be written as equation 2.1:

$$H = T_{elec}(\{r_i\}) + T_{nucl}(\{R_i\}) + V_{elec-elec}(\{r_i\}) + V_{nucl-nucl}(\{R_i\}) + V_{elec-nucl}(\{r_i, R_i\}) \quad (2.1)$$

where T_{elec} and T_{nucl} are the electronic and nuclear kinetic energies; $V_{elec\text{-}elec}$, $V_{nucl\text{-}nucl}$ and $V_{elec\text{-}nucl}$ are the electron-electron, nucleus-nucleus and electron-nucleus electrostatic potentials, and $\{\boldsymbol{R_i}\}$ and $\{\boldsymbol{r_i}\}$ being the nuclear and electron positions, respectively.

For the purposes of the electronic problem, the nuclei move much more slowly than electrons (the mass of the lightest nucleus is about 1840 times heavier than an electron), the nucleus may be considered fixed and thus we can neglect T_{nucl}, while $V_{nucl\text{-}nucl}$ is simply a constant. Since the addition of a constant to the Hamiltonian will only result in adding a constant to the eigenvalues, with no effect on the eigenfunctions, the electronic wavefunctions can then be obtained by solving the stationary state Schrödinger equation with the electronic Hamiltonian, given in equation 2.2:

$$H_{elec} = T_{elec}(\{r_i\}) + V_{elec-elec}(\{r_i\}) + V_{elec-nucl}(\{r_i, R_i\}) \quad (2.2)$$

This gives a set of "electronic" wavefunctions, which depend explicitly on the electron positions only (since they are functions of r_i only), while implicitly depending on the set of fixed nuclear coordinates chosen (since these appear as parameters in the Hamiltonian). Thus for each set of nuclear coordinates, a different set of electronic wavefunctions would need to be generated. This is the domain of electronic structure theory or *ab initio* calculation methods. The approximation is usually very good, leading to errors in the energy of the order of 10^{-4} for the hydrogen molecule [9], and better for molecules with heavier nuclei. It is certainly negligible compared to the other approximations involved in actually solving the Schrödinger equation for the electrons.

Having solved the electronic problem by considering the nuclei fixed, we now consider the rapidly moving electrons to induce a time-averaged potential, which is a function of nuclear coordinates (the implicit dependence mentioned above) demonstrated in equation 2.3:

$$\langle E_{elec}(\{R_i\})\rangle = \langle T_{elec} + V_{elec-elec} + V_{elec-nucl}(\{R_i\})\rangle \tag{2.3}$$

Since the nucleus-nucleus repulsion also depends on the $\boldsymbol{R}_i$, the nuclei are effectively moving in a potential entirely described by their coordinates, given in equation 2.4.

$$V_{tot}(\{R_i\}) = \langle E_{elec}(\{R_i\})\rangle + V_{nucl-nucl}(\{R_i\}) \tag{2.4}$$

Whereas, the nuclear Hamiltonian is given by equation 2.5, which can be solved in terms of the nuclear coordinates only. The resultant nuclear wavefunctions $\Phi_{nucl}(\{\boldsymbol{R}_i\})$ describe all the vibrational, translational and rotational properties of the molecule, and therefore form a basis for the calculation of macroscopic properties from statistical mechanics.

$$H_{nucl}(\{R_i\}) = T_{nucl}(\{R_i\}) + V_{tot}(\{R_i\}) \tag{2.5}$$

The actual form of the potential energy described by the terms on the right in equation 2.4 has a similar form to the classic Lennard-Jones potential in the special case of diatomic systems. This consists of long-range attractive

interactions due to the electronic term and short-range repulsive interactions due to the nucleus-nucleus repulsion, as well as contributions from the electronic potential related to the Pauli's Exclusion Principle [10].

2.2 Force Fields

Force field methods require that an appropriately parameterised force field be available, which in turn requires specification of both the functional form of the potential energy function and the numerical values of the functional parameters based upon experimental results or *ab initio* calculations.

2.2.1 Molecular Mechanics Potential Energy Functions

For molecular mechanics, the significance of the Born-Oppenheimer approximation is that the potential for the nuclear Hamiltonian (equation 2.5) is a function of the nuclear coordinates only. However, the system is not treated quantum mechanically using equation 2.5, but by a set of empirical potential functions known as a force field. These potentials are derived from experimental results and *ab initio* calculations. Since the nuclei are sufficiently heavy, they can be treated classically: a common criterion which is applied is that the de Broglie wavelength (h/p) is much less than the distance between the nuclei; alternatively, one can observe that at room temperature the available nuclear energy states form a continuum ($\Delta E << k_b T$), which is not the case for electrons [10]. These potential functions determine the relative stabilities of the various possible stable and meta-stable structures of a system. Most potential functions for CDs and other biological macromolecules treat the molecular potential as a sum of the valence and non-bonded interactions. A typical energy function for a molecular system, like the one used in CHARMM [1], has the form:

$$E = E_{bonds} + E_{angles} + E_{Urey-Bradley} + E_{improper-dihedrals} + E_{dihedrals} + E_{VdW} + E_{elec} \qquad (2.6)$$

Conceptually this force field can be divided up into three parts: vibrational modes, torsional flexibility and intermolecular interactions.

The vibrational part of the force field is represented by the harmonic potentials for the internal coordinates: bond lengths, bond angles, Urey-Bradley and out of plane

distortions (improper dihedral angles); the bond-stretching and angle-bending terms are often referred to as a hard degrees of freedom, in that quite substantial energies are required to cause significant deformations from their equilibrium values. At normal temperatures and pressures the deformations are sufficiently small for the harmonic approximations to apply (Hooke's Law) [12]. The functional forms of these terms are:

- Bond potential:

$$E_{bonds} = \sum_{bonds} K_b (b - b_0)^2 \tag{2.7}$$

The equilibrium or ideal bond length is represented by b_0. The force constant K_b can be evaluated from infrared stretching frequencies or quantum mechanical calculations, while b_0 can be obtained from crystal structures or microwave data.

- Bond angle potential:

$$E_{angles} = \sum_{angles} K_\theta (\theta - \theta_0)^2 \tag{2.8}$$

Here θ is the angle between atoms separated by two covalent bonds, the K_θ is the force constant for angle-bending and θ_0 is the equilibrium bond angle.

- Urey-Bradley potential:

It is also a harmonic potential to model the Urey-Bradley deformations of the distance between atoms that are separated by two covalent bonds. Here S is the distance between atoms separated by two covalent bonds (1, 3 distance), k_{UB} the corresponding force constant and S_0 the equilibrium distance.

$$E_{Urey-Bradley} = \sum_{UB} K_{UB} (S - S_0)^2 \tag{2.9}$$

- Improper dihedral angle potential:

$$E_{improper-dihedrals} = \sum_{improper-dihedrals} K_{\omega}(\omega - \omega_0)^2 \tag{2.11}$$

The improper dihedral angles are intended to model linearity about a tetrahedrally extended heavy atom and to maintain planarity about certain planar atoms. Again the K_{ω} is the improper dihedral angle force constant and ω_0 is the equilibrium angle.

These terms are essentially representing the vibrational frequencies that can be experimentally observed.

The torsional flexibility contribution to the potential energy function is the dihedral angles, which represents most of the molecular flexibility. They are described by the Fourier term:

- Dihedral angle potential:

$$E_{dihedrals} = \sum_{dihedrals} K_{\phi}(1 - \cos(n\phi)) \tag{2.12}$$

where $n \in N$. This is a four atom term based on the dihedral angle about an axis defined by the middle pair of atoms. K_{ϕ} is the force constant and ϕ is the dihedral angle achieved during the simulation.

The third and last contributions to the potential energy are the intermolecular interactions represented by the Lennard-Jones (12, 6) and Coulombic pairwise interactions. These interactions are essential for modelling long-range and solute-solvent interactions.

- Van der Waals interactions: this term uses the Lennard-Jones (12, 6) potential.

$$E_{VdW} = \sum_{i>j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \tag{2.13}$$

This has a minimum at an interatomic distance of the van der Waals radii of the two interacting atoms. Parameters A and B depend on the atoms involved and have been determined by a variety of measurements, including non-bonded distances in crystals and gas-phase scattering measurements.

- Electrostatic interaction potential:

$$E_{elec} = \sum_{i>j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.13}$$

The electrostatic interactions are represented by a Coulombic potential where q_i and q_j is the (partial) charge on atom i and j, ε_0 is the relative permittivity of a vacuum and r_{ij} is the distance between the two charges.

The potential energy function described here is the form of the CHARMM [1, 8] potential energy function used in our simulations. However, there are many variants of the expression for the different parts of the potential energy function in use. For instance, although hydrogen bonding can be adequately modelled with the van der Waals interactions and electrostatic terms potential above [7], some force fields include special hydrogen bonding terms to ensure that this interaction is correctly modelled [12].

2.2.2 Carbohydrate Force Fields

There are numerous force fields and parameters sets available for modelling carbohydrates. Carbohydrate-specific force fields were developed as structural models to assist in interpreting ambiguous NMR data. Most of the more sophisticated biomolecular force fields in use currently were originally designed for the study of

proteins, but have also been extended to include other classes of molecules. Carbohydrate-specific parameter sets have been developed for each of the three principal biomolecular force fields: CHARMM [13, 14, 15], AMBER [16, 17-21] and GROMOS [22, 23, 24]. Only the force fields available for CHARMM will be discussed, as the CHARMM simulation program was used exclusively in this thesis.

Carbohydrate parameter set for CHARMM

The first carbohydrate parameter set for use with the CHARMM program, was developed by Brady and co-workers [13], is referred to as the HGFB force field, after the authors' initials. This force field was parameterised using the CHARMM parameter optimisation feature to fit the experimental vibrational and crystallographic information for α-D-glucose and the force field was optimised for use with the TIP3P water model. Partial charges were based on standard CHARMM values. The glycosidic angle was treated the same way as the ring C-O-C angle and there was no specific term to treat the *exo*-anomeric effect. The different α and β linkages are known to prefer different minima about the linkage dihedral angles, this phenomenon is known as the *exo*-anomeric effect. Naturally, the HGFB force field gives good agreement with vibrational data, and the energy minimum in vacuum agrees with the crystal structure.

Palma et al. [14] then developed a force field (the PHLB force field) that replaces the general dihedral angle term in the HGFB force field with a specific dihedral angle term, using up to three terms for each dihedral (1-, 2- and 3-fold cosine terms). The dihedral angles were parameterised to reproduce the experimental vibrational frequency data and to fit *ab initio* energy surfaces for small molecules such as ethylene glycol [25].

Kuttel et al. recently adjusted the PHLB force field to correct the hydroxymethyl behaviour and the new CSFF (Carbohydrate Solution Force Field) was validated by calculating the potential of mean force (PMF) profiles both in solution and vacuum for β-D-glucose and β-D-galactose [26]. The PMF profiles confirm that the MD simulations using this force field will produce equilibrium population distributions and rotational frequencies in solution that are consistent with experimental evidence. The secondary hydroxyl groups rotational frequencies had been increased, which should aid in the equilibration of these groups in solution. In addition, normal modes calculated for α-D-glucose in vacuum compare favourably with experimental data.

The hydroxymethyl group rotational frequency had been increased, resulting in the hydroxymethyl groups equilibrating faster and the equilibrium populations are now achievable with nanosecond simulations. This has far reaching implications for simulations of molecules where the rotation of the hydroxymethyl group is important in molecular conformations and dynamics. In addition, the CSFF should display improved equilibration and dynamics for α(1-6) linked saccharides such as isomaltose and dextran [26].

2.2.3 Water Models

Computational solution studies on the atomic level can take two forms: one where the solvent is included explicitly as separate molecules that interact with the solute and another where the solvent is implicitly included in the parameterisation of the terms of the force fields. Simulations where the solvent is implicitly included are attractive because they are computationally far cheaper to perform. However, because of the major structural roles that specific hydrogen bonds play in water-polysaccharide solutions, implicit (or continuum) models may be inadequate for modelling aqueous states of carbohydrates [12]. The use of explicit water in carbohydrate solution simulations have been shown to be important for reproducing NMR results and to adequately model the available conformational space in simulations of oligosaccharides.

Various explicit water models have been developed for use in MD simulations: from the simple rigid site-interaction models, through flexible models that allow for conformational changes in the water molecule, to complex models that include polarisation and many-body effects. The simplest water models treat the water as a rigid molecule and depend on non-bonded interactions. The electrostatic interaction is modelled using Coulomb's law and the attraction and repulsion forces using the Lennard-Jones potential [27, 29]. The potential function for simple models such as TIP3P and TIP4P is given by:

$$E_{ij} = \sum_{i} \sum_{j \neq i} \frac{k_C q_i q_j}{r_{ij}} + \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^{6}} \qquad (2.14)$$

where k_C is the electrostatic constant, which typically has a value of 332.1 Å·kcal/mol; q_i and q_j are the partial charges on atoms i and j; r_{ij} is the distance between the two atoms; and A and B are the Lennard-Jones parameters. The charge sites may reside on either the atoms or on dummy sites (such as lone pairs). Figure 2.1 shows the general shape for the 3- to 5-site water models; where the exact geometric parameters (OH distance and HOH angle) of the water depend on the model. The parameters for these water models are shown in Table 2.1.

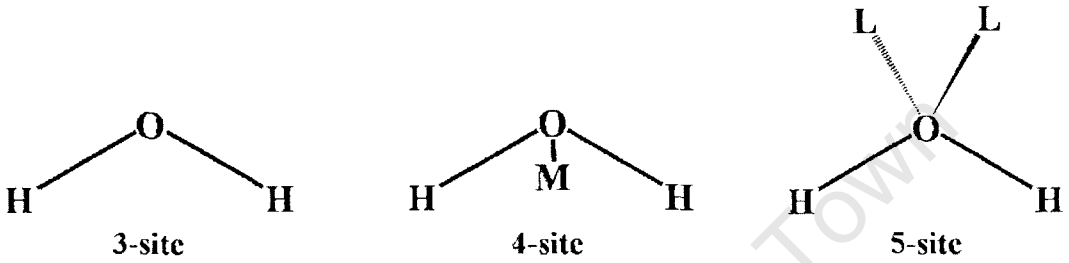

Figure 2.1: Illustration of the 3- to 5-site water models and the location of their sites.

The three-site water models are the simplest of the water models, in that each site corresponds to the three atoms in the water molecule. The most commonly used three site water models are: the transferable intermolecular potential (TIPS, TIP3P) [27, 28] and the simple point charge (SPC and SPC/E) [29, 30] force fields. Each atom is assigned its point charge and only the oxygen atom is assigned Lennard-Jones parameters. These models use a rigid geometry that matches the known geometry of the water molecule, except the SPC and SPC/E models which assume an ideal tetrahedral shape (HOH angle of 109.47°) instead of the observed angle of 104.5°. The SPC/E model adds and average polarization correction potential energy function

$$E_{pol} = \frac{1}{2}\sum_i \frac{(\mu - \mu^o)^2}{\alpha_i} \tag{2.15}$$

where μ is the dipole of the effectively polarised water molecule, μ^o is the dipole moment of an isolated water molecule, α_i is the isotropic polarisability constant, with a value of 1.608 x 10^{-40} F m and the charges are constant; the correction adds 1.25 kcal/mol to the total energy. The SPC/E model results in a better density and diffusion constant than the SPC model [30]. The TIP3P model implemented in the CHARMM force field is slightly modified in that it also gives a small van der Waals radius and

places Lennard-Jones parameters on the hydrogen atoms; the charges are not modified [31].

The four-site model that is most commonly used is the TIP4P force field. Instead of placing the negative charge on the oxygen atom, it is placed on the dummy atom (labelled M, Figure 2.1) placed near the oxygen atom at the bisector of the HOH angle. This improves the electrostatic distribution around the water molecule [28].

The five-site water model, which includes TIP5P force field, is not commonly used due to its high computational costs. This model places the negative charge on the dummy atoms (labelled L, Figure 2.1) representing the lone pairs of the oxygen atom, with tetrahedral-like geometry, with an angle of θ_{LOL}. When compared to the three- and four-site models, the TIP5P model results in improvements in the geometry for the water dimer. The consequence of this is that a more "tetrahedral" water structure is achieved, which reproduces the experimental radial distribution function from neutron diffraction and the temperature of the maximum density of water [32].

	TIP3P	**SPC**	**SPC/E**	**TIP4P**	**TIP5P**
***r(OH)* (Å)**	0.9572	1.0	1.0	0.9572	0.9572
∠HOH	104.52	109.47	109.47	104.52	104.52
***A* x 10^{-3} (kcal·Å^{12}/mol)**	582.0	629.4	629.4	600.0	544.5
***B* (kcal·Å^{6}/mol)**	595.0	625.5	625.5	610.0	590.3
q(O)	-0.834	-0.82	-0.8476	0.0	0.0
q(H)	0.417	0.41	0.4238	0.52	0.241
q(M)	-	-	-	-1.04	0.0
q(L)	-	-	-	-	-0.241
***r(OM)* (Å)**	-	-	-	0.15	0.70
θ_{LOL}	-	-	-	-	109.47

Table 2.1: Parameters for water models: TIP3P [28], SPC [29], SPC/E [30], TIP4P [28], TIP5P [32].

Of the popular models, TIP4P, TIP5P and SPC/E give the best reproduction of the solvent structuring, with the closest correspondence of their pair distribution functions Of the popular models, TIP4P, TIP5P and SPC/E give the best reproduction of the solvent structuring, with the closest correspondence of their pair distribution functions to experiment. This is thought to be because they favour a more tetrahedral local

structuring of water [33]. The recent calculation of a three-dimensional distribution function for SPC/E has shown that the average numbers of nearest neighbours occupying tetrahedral sites are in fact exactly 4.0 [34]. Of the two families, it is generally found that the SPC-type models reproduce structural and diffusional characteristics best, while the lack of structure in TIP-type models seem to give it elevated self-diffusion rates. On the other hand, the TIP-type models reproduce best the experimental results over a wide range of temperature and pressure [33].

2.3 Molecular Dynamics

Once a satisfactory potential energy function and force fields have been developed, it is then possible to generate configurations of a system in a suitable ensemble. The two main simulation techniques for dynamics are MD, where Newton's equations of motion are integrated over time, and Stochastic Dynamics, in which the Langevin equation of motion for Brownian motion is integrated over time.

Molecular Dynamics Simulations

In the MD approach, Newton's equations of motion, equations 2.16 and 2.17, are integrated simultaneously, over time, for all atoms in the system.

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i}{m_i} \tag{2.16}$$

$$F_i = -\frac{\partial V(r_i,...r_N)}{\partial r_i} \tag{2.17}$$

The force on atom i is denoted by F_i and t denotes the time. The gradient of the potential energy, $V(r)$, are the forces, and therefore $V(r)$ must be a differentiable function of the atomic coordinates r_i. The integration of equation 2.17 is performed using small time steps, typically 1 to 10 fs.

Stochastic Dynamics

Stochastic dynamics is an extension of MD [35]. It involves integrating the stochastic Langevin equation of motion, equation 2.18:

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i}{m_i} + \frac{R_i}{M_i} - \gamma_i \frac{dr_i(t)}{dt} \tag{2.18}$$

This equation has two added terms in it: a stochastic force, R_i, that models random collisions with solvent molecules and a frictional drag force (proportional to γ_i, the frictional coefficient) that dampens the solute's motion through the solvent. The stochastic term introduces energy and the frictional term removes energy from the system.

2.3.1 Integration of the Equations of Motion

Given that the generation of MD data involves integrating a large number of differential equations, it is of the utmost importance that the algorithm be efficient. The most computationally intensive part of a MD simulation is the calculation of the forces at each timestep, which far exceed the computational requirements of any available integration algorithms. Consequently, the most important feature of any integration algorithm is the ability to support a long integration timestep with a small numerical error, rather than being fast itself. One of the most established and most commonly used method is the Verlet algorithm [36, 6] given in equation 2.19:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t) \tag{2.19}$$

The deficiency, however, of the Verlet method is that a very small term ($\delta t^2 a(t)$) is being added to a very large term ($2r(t) - r(t - \delta t)$), which can lead to a loss of accuracy.

An improvement upon this is the leap-frog algorithm (equations 2.20 and 2.21) [6].

$$r(t+\delta t) = r(t) + \delta t v\left(t + \frac{1}{2}\delta t\right) \tag{2.20}$$

$$v\left(t + \frac{1}{2}\delta t\right) = v\left(t - \frac{1}{2}\delta t\right) + \delta t a(t) \tag{2.21}$$

The stored variables are the current coordinates $\boldsymbol{r}(t)$ and accelerations $\boldsymbol{a}(t)$ together with the mid-step velocities $\boldsymbol{v}(t - 1/2\delta t)$. The velocity equation 2.21 is implemented first, and then the velocities leap over the coordinates to give the next mid-step values $\boldsymbol{v}(1 + 1/2\delta t)$. During this step, the current velocities may be calculated using equation 2.22

$$v(t) = \frac{1}{2}\left(v\left(t + \frac{1}{2}\delta t \right) + v\left(t - \frac{1}{2}\delta t \right) \right) \tag{2.22}$$

This is required because the energy ($\boldsymbol{H} = \boldsymbol{K} + \boldsymbol{V}$) at time t must be calculated, as well as any other variables that require both the coordinates and velocities at the same time. Following this, equation 2.20 is then used to leap the coordinates over the velocities once again. After this, the new accelerations may be calculated ready for the next step [6]. This method is equivalent to the Verlet algorithm, but it avoids the need to include the δt^2 terms and concomitant inaccuracies.

2.3.2 Potential Truncation

The time taken for a MD simulation significantly increases as the size of the system increases and this is mainly due to the calculation of the non-bonded interactions. For a MD package that uses a pairwise interaction model, such as the CHARMM empirical potential energy function, generally the number of non-bonded interaction terms increases with the square of the number of atom and hence the computation time is proportional to N^2 (due to the double loop used for calculating the forces), rather than N. In addition to this practical consideration, it is also undesirable to have long-range forces, especially for a system which uses periodic boundary conditions (section 2.3.3) since it would lead to long range order (effectively modelling a crystal), which is not wanted in a simulation of bulk solvent.

A common practice for treating non-bonded interactions is to use a non-bonded cutoff (or potential truncation of the non-bonded van der Waals and Electrostatic interactions) and apply the minimum image convention (section 2.3.3). The minimum image convention is a natural consequence of periodic conditions. It requires that interactions for molecule i are only calculated between that molecule and its closest image of every other molecule in the system, Figure 2.3 (section 2.3.3) [6]. The present cutoff radii are generally between 9 and 12 Å, since the pair potential is

negligible at this distance. However, this introduces some severe errors, since the potential is no longer continuous, which in turn means that the total energy is not conserved. The cutoff introduces a discontinuity in the potential function at the cutoff distance, which artificially increases the kinetic energy and temperature of the system. As a result of this it is necessary to either scale the velocities or couple the system strongly to a heat bath to keep the temperature constant. Both methods are unphysical and potentially detrimental, since they may obscure underlying defects in the simulation setup [10]. To avoid this, a smoothing function is applied that "smoothes" the potential to zero. There are two commonly used methods namely the "shifting" (equation 2.23) and "switching" (equation 2.24) functions.

$$S_f(r_{ij}) = \begin{cases} \left(1 - \frac{2r_{ij}^{\,2}}{r_{cut}^{\,2}} + \frac{r_{ij}^{\,4}}{r_{cut}^{\,4}}\right) & r_{ij} < r_{cut} \\ 0 & r_{ij} > r_{cut} \end{cases} \tag{2.23}$$

$$S_w(r_{ij}) = \begin{cases} 1 & r \le r_{on} \\ \dfrac{(r_{off} - r_{ij})^2 (r_{off} + 2r_{ij} - 3r_{on})}{(r_{off} - r_{on})^3} & r_{on} < r_{ij} \le r_{off} \\ 0 & r > r_{off} \end{cases} \tag{2.24}$$

The switching function are polynomials in the separation distance that smoothly "turns off" the interaction potential over the given range between the "cut-on" (distance where the potential begins to be modified) and "cutoff" (distance where the potential is set to zero) distances. The interaction potential is unaffected for distances less than the "cut-off" distance and zero for distances greater than the "cutoff" distance. In contrast, the shifting functions modifies the interaction potential over the entire range to the "cutoff" distance, and goes to zero beyond the "cutoff" distance. Smoothing functions can be applied on an atom-by-atom basis or can be applied to a predefined neutral group of atoms within the molecule. Switching functions applied on a group-by-group basis have been shown to be the superior method, since it gives excellent energy conservation [37]. However, there are many other methods of producing good cutoff behaviour, for example, applying the switching function

directly to the forces, so that the potential is now given by the integral of the switched forces ("force-based" methods [37]).

2.3.3 Periodic Boundary Conditions and Minimum Image Convention

Periodic boundary conditions enable macroscopic properties to be calculated from simulations using a relatively small number of particles, and hence reduced simulation time. The major problem with MD simulations is that there are a large number of molecules that lie on the surface, which result in unwanted edge or surface effects: these molecules experience forces very different to the forces in the bulk solvent. This problem can be overcome by implementing periodic boundary conditions; hence the simulation can be performed using a small number of particles in such a way that the particles experience forces as if they were in the bulk solvent. The atoms of the system are put into a "box" or "cell", of a shape that will be space filling, and repeatedly tiled throughout space. The cube and the truncated octahedron are the most commonly used shapes, but certainly the cubic simulation box is much more versatile, allowing easy implementation of constant pressure dynamics and the application of sheer stresses to the system. If a cube is used, the box is treated as if it was surrounded by 26 ($3^3 - 1^3$) identical images of itself. When a particle leaves the box, it re-enters it with the identical velocity at the opposite side, Figure 2.2 [6, 12].

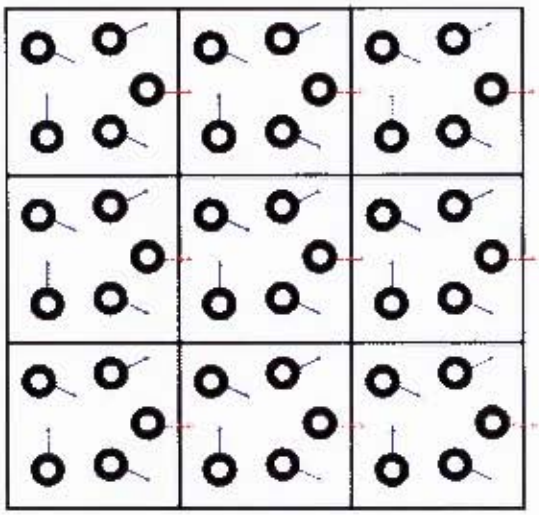

Figure 2.2: Two dimensional illustration of periodic boundary conditions. The central box is the system that is repeated throughout space.

The formulation of the potential is exactly the same, except that the atoms are now able to interact with image atoms. Thus in principle, each atom now interacts with all of its infinite set of images as well as those of all other atoms. This is most undesirable for both practical and theoretical reasons: firstly the calculation of such forces in a periodic system is possible, but more computationally intensive, and secondly long-range interactions in the potential will give rise to long range order, which is undesirable in a simulation of bulk liquid. For this reason, it is common to restrict the interaction of atom *i* to its closest image of every other atom in the system, a condition known as the minimum image convention, Figure 2.3 [6].

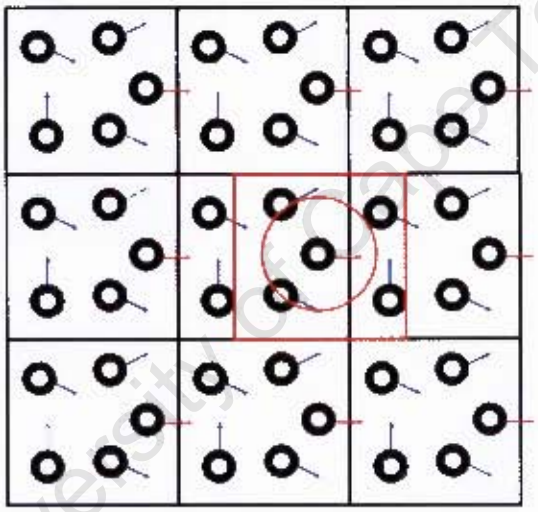

Figure 2.3: The red box is a two-dimensional illustration of the minimum image convention and spherical potential cutoff. The central atom, in the red box (with the red arrow), is shown with its closest image of every other atom in the system.

There is an inherent risk in setting up a simulation of a solute in a periodic box such that the system does not behave too much like a crystal. Thus the cutoff radius, in a cubic box of length L, should not exceed ½L. A satisfactory guess to the minimum acceptable box size may be estimated by using a sphere, large enough to always contain the solute (or any atom with a spherical cutoff), placed inside the simulation box, Figure 2.4. To ensure that the molecule does not interact with itself, the minimum box length should be at least the diametre of the sphere plus two cutoff radii for the non-bonded interactions. Two cutoff radii are used to ensure that the solvent molecules do not interact twice with the solute.

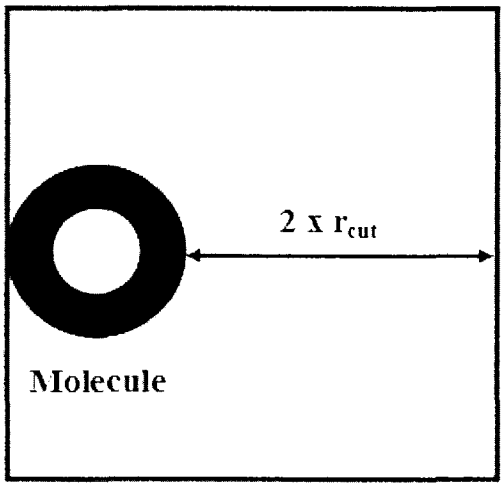

Figure 2.4: Minimum cell size for a cubic simulation box.

Although the use of periodic boundary conditions is best for dealing with homogeneous solutions under equilibrium conditions, it is not suitable for modelling solid/liquid interfaces or a restricted region of macromolecules such as an enzyme active site, where a water sphere, or droplet, (using spherical boundary conditions, section 2.3.4) is used. However, for water under biological conditions, periodic boundary conditions are adequate [6, 10].

2.3.4 Spherical Boundary Conditions

When studying an enzyme reaction, in the active site, or a very large macromolecule, using a cubic simulation box with periodic boundary conditions is computationally unfeasible due to the size of not only the box but also the size of the system (with its large number of explicit solvent molecules). Thus a new approach to studying localised chemical events and very large macromolecules in condensed phase was developed and is known as deformable stochastic boundaries.

Deformable or spherical stochastic boundary conditions provide a simple and convenient method for reducing the total number of solvent particles explicitly included in the simulations of localised processes while decreasing edge effects. Both energy flow across the boundary and density fluctuations in the simulation region are included; this makes it possible to treat nonequilibrium processes, such as thermal gradients and exothermic and endothermic chemical reactions [39]. The system that is

being studied is divided up into three regions; an inner (reaction) region in which the particles are treated explicitly by MD, a surrounding buffer region in which the particles are treated explicitly using stochastic dynamics [38], and a deformable boundary, whose forces on the particles in the reaction and buffer regions arise from average structure of the system beyond the boundary. The addition of a "deformable boundary force" arises from mean field interactions with particles beyond the boundary, where only the knowledge of the radial distribution function is required. The system is illustrated in Figure 2.5.

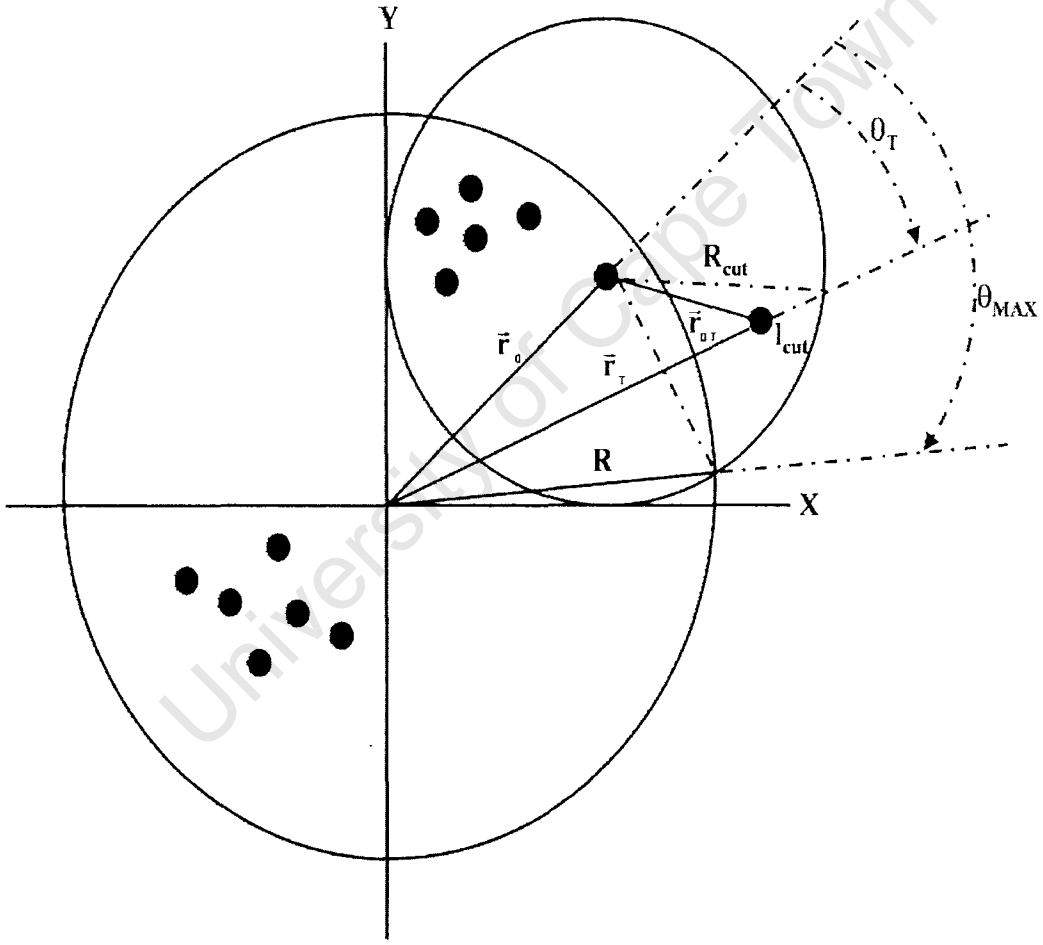

Figure 2.5: Two-dimensional conceptual illustration for the geometric details for spherical boundary force calculation. The particle of interest (particle zero) is located r_0 within the simulation sphere of radius, R. The "test particle" is located at r_T outside the simulation sphere, but inside the cutoff radius, R_{cut}. The effective force is calculated by the Lennard-Jones force arising from the interaction at separation r_{0T} multiplied by the probability of the test particle being there, given by the radial distribution function, $g(r_{0T})$. Taken from reference 39.

The three-dimensional radial boundary force, or spherical boundary force, is calculated as follows:

$$F_W(r_0) = 4\pi \int_{x_{\max}}^{1} dx_T \int_{R}^{l_{cut}(x_T)} dr_T\, r_T F(\vec{r}_{0T})\rho g(\vec{r}_{0T}) \tag{2.25}$$

with

$$x_T = \cos\theta_T \tag{2.26}$$

$$F(\vec{r}_{0T}) = \left(\frac{r_T \cos\theta_T - r_0}{r_{0T}}\right)\frac{\partial u(\vec{r}_{0T})}{\partial \vec{r}_{0T}} \tag{2.27}$$

and

$$\vec{F}_B(\vec{r}_0) = \hat{r}_0 F_W(r_0) \tag{2.28}$$

Spherical boundary conditions allow a small volume system to be studied without the need for periodic boundary conditions, which is useful for studying in particular reactions occurring in the enzyme active site. It allows the reduction in the size of the system, thus reducing the computational requirements. Also, systems with thermal gradients, such as are generated by exothermic processes, can be treated by using spherical boundary conditions that avoids unphysical heating effects that arise in conventional periodic boundary methods [39].

2.3.5 Constraint Dynamics

The timestep to be used in MD simulations is dictated by the highest frequency motion present in the system. Bond vibrations involving light atoms such as hydrogens occur so quickly that an extremely short timestep would be required to solve the equations of motion. These very high frequency motions are usually of little interest than the lower frequency modes, which often correspond to major conformational changes. Constraint MD enables the selected coordinates to be constrained during a simulation without affecting the other degrees of freedom, thus allowing the timestep to be increased.

The most commonly used constraint method in MD simulations is Ryckaert, Ciccotti and Berendsen's SHAKE algorithm [40]. SHAKE is a special technique developed to treat the dynamics of molecular system where certain arbitrarily selected

degrees of freedom (such as bond lengths) are constrained, while others are free to evolve under the influence of intermolecular and intramolecular forces.

SHAKE is most useful for large molecules. The most common use of SHAKE is for constraining bonds involving hydrogen atoms, which vibrate at very high frequencies. In addition, SHAKE algorithm is often used to maintain both fixed bond lengths and angles for water molecules in solution simulations [31].

2.3.6 Simulation Ensembles

There are four common types of statistical ensembles used in MD simulations: the microcanonical, or constant-NVE; the canonical, or constant-NVT; the isothermal-isobaric, or constant-NPT; and the grand-canonical, or constant-μVT, ensembles. For each ensemble, the aforementioned thermodynamic variables are specified, i.e. fixed. Other thermodynamic quantities must be determined by ensemble averaging and, for any particular state point, the instantaneous values of the appropriate phase function will deviate from this average value, i.e. fluctuations occur, which are crucial for calculating certain thermodynamic parameters such as the heat capacities etc. [6].

Computer simulations are most credible when they can be compared with the results from a physical experiment. It would be ideal to run the dynamics simulation in an ensemble that corresponds closely to the experimental conditions. Hence, the most obvious choice would be to run either the canonical, constant-NVT, or the isothermal-isobaric, constant-NPT, ensemble. There are various methods that have been proposed to implement these ensembles [7, 8, 41]. The canonical, constant-NVT, ensemble is the easiest of the two to implement in a program, because the periodic box has a fixed volume. In the case of running the isothermal-isobaric, constant-NPT, ensemble, the size of the box size is constantly scaled to ensure that a constant pressure and temperature is maintained. The NVT ensemble corresponds to the Helmholtz free energy and the NPT ensemble corresponds to the Gibbs free energy, and thus NPT is therefore the ideal ensemble for comparison with experiment. In the simulations performed in this work, both NPT and NVT ensembles were used; not only because of the ease of implementation, but also most of the experimental data available, or which will become available in the future, are performed under either constant pressure and temperature or constant volume and temperature conditions.

References

1. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus. *J Comput. Chem.* **1983**, *4(2)*, 187.
2. P. K. Weiner, P. A. Kollman. *J. Comput. Chem.* **1981**, *2*, 398.
3. W. F. van Gunsteren, H. J. C. Berendsen. *GROMOS: GROningen MOlecular Simulation Software.* Technical Report, Laboratory of Physical Chemistry, University of Groningen, **1988**.
4. W. Smith, T. R. Forester. *J. Molecular Graphics.* **1996**, *14*, 136.
5. A. R. Leach. *Molecular Modelling. Principles and Applications.* Addison Wesley Longman Limited, **1996**.
6. M. P. Allen, D. J. Tildesley. *Computer Simulations of Liquids.* Oxford University Press, **1987**.
7. W. F. van Gunsteren and H. J. C. Berendsen. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 992.
8. M. Karplus, G. A. Petsko. *Nature*. **1990**, *347*, 631.
9. A. Szabo, N. S. Ostlund. *Modern Quantum Chemistry*. Dover Publications, London, 1st Ed., **1996**.
10. R. Best, "Combined NMR and Simulation Study of Carbohydrate Linkage Dynamics", University of Cape Town, **2000**.
11. P. C. Reidi, Thermal Physics: *An Introduction to Thermodynamics, Statistical Mechanics and Kinetic Theory*. Oxford University Press, Oxford, 2nd Ed., **1988**.
12. R. U. Lemieux, S. Koto, D. Voisin. The Exo-Anomeric Effect. *The Origin and Consequence of the Anomeric Effect.* **1979**, ACS Symposium, *17*.
13. M. M. Kuttel, "Developing Analytical Tools for Saccharides in Condensed Phase", MSc. Thesis, University of Cape Town, **1999**.
14. S. N. Ha, A. Giammona, M. Field, J. W. Brady. *Carbohydr. Res.* **1988**, *180*, 207.
15. R. Palma, M. E. Himmel, G. Liang, J. W. Brady. *Molecular Mechanics Studies of Cellulases.* In M. E. Himmel, editor, *Glycosyl Hydrolases in Biomass Conversion: ACS Symposium Series.* ACS, Washington, **2001**.
16. S. Reiling, M. Schlenrick, J. Brickamn. *J. Comput. Chem.* **1996**, *17(4)*, 450.

17. T. M. Glennon, Y. Zheng, S. M. le Grand, B. A Shutzberg, K. M. Jnr. Merz. *J. Comput. Chem.* **1994**, *15(9)*, 1019.
18. R. J. Woods, R. A. Dwek, C. J. Edge. *J. Phys. Chem.* **1995**, *99*, 3832.
19. H. Senderowitz, C. Parish, W. C. Still. *J. Am. Chem. Soc.* **1996**, *118*, 2078.
20. H. Senderowitz, W. C. Still. *J. Org. Chem.* **1997**, *62*, 1427.
21. S. K. Gregurick, J. H.-Y. Liu, D. A Brant, R. B. Gerber. *J. Phys. Chem. B.* **1999**, *103*, 3476.
22. F. A. Momany, J. L. Willet. *Carbohydr. Res.* **2000**, *326*, 194.
23. K. Ott, B. Meyer. *J. Comput. Chem.* **1996**, *17(8)*, 1068.
24. W. Damm, A. Frontera, J. Tirado-Rives, W. L. Jorgensen. *J. Comput. Chem.* **1997**, *18(16)*, 1955.
25. D. Kony, W. Damm, S. Stoll, W. F. van Gunsteren. *J. Comput. Chem.* **2002**, *23(15)*, 1416.
26. M. A. Murcko, R. A. DiPaola. *J. Am. Chem. Soc.* **1992**, *114*, 10010.
27. M. Kuttel, J. W. Brady, K. J. Naidoo. *J. Comput. Chem.* **2002**, *23(13)*, 1236.
28. W. L. Jorgensen. *J. Am. Chem. Soc.* **1981**, *103*, 335-340.
29. W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. W. Impey, M. L. Klein. *J. Chem. Phys* **1983**, *79*, 926-935.
30. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, In *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, **1981**), p. 331.
31. H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. *J. Phys. Chem.* **1987**, *91*, 6269.
32. M. M. Kuttel. "Simulations of Carbohydrate Conformational Dynamics and Thermodynamics", PhD. Thesis, University of Cape Town, **2003**.
33. M. W. Mahoney, W. L. Jorgensen. *J. Chem. Phys.* **2000**, *112*, 8910.
34. D. N. Bernardo, Y. Ding, K. Krogh-Jespersen, R. M. Levy. *J. Phys. Chem.* **1994**, *98*, 4180.
35. K. A. Soper, M. G. Phillips. *Chem. Phys.* **1986**, *107*, 47.
36. H. C. Andersen. *J. Chem. Phys.* **1980**, *72(4)*, 2384.
37. C. Molteni, M. J. Parrinello. *J. Am. Chem. Soc.* **1998**, *120*, 2168.
38. M. Berkowitz, J. A. McCammon. *Chem. Phys. Lett.* **1982**, *90*, 215.
39. C. L. Brooks III, M. Karplus. *J. Chem. Phys.* **1983**, *79(12)*, 6312.
40. J. P. Ryckaert, G. Ciccotti. H. J. C. Berendsen. *J. Comp. Phys.* **1977**, *23*, 327.
41. H. Bekker. *J. Comput. Chem.* **1997**, *18*, 1930.

Chapter 3

Computational and Experimental Analytical Techniques

3.1 Computational Analysis

3.1.1 Pair Distribution Function

In analysing the local structuring of liquids, the first and most commonly used approach is the calculation of the pair distribution function (PDF), *g(r)*. This function gives the probability of finding pairs of atoms a distance *r* apart, relative to the probability expected for a completely random distribution at the same density [1]. The pair distribution function can easily be extended to molecular liquids, such as water, by calculating the site-site distributions, for example O-O and O-H *g(r)*. It can be further extended to the solvent distribution around solute sites using the site-site *g(r)*, with a solute atom or residue as one of the sites.

The pair distribution function, *g(r)*, for solvent molecules around a sugar (or any solute) molecule is defined [2] in equation 3.1:

$$g(r) = \frac{1}{4\pi\rho r^2}\frac{dN(r)}{dr} \tag{3.1}$$

In this equation, r is the interatomic distance from the selected atoms in the solute molecule and the solvent molecule, ρ is the bulk solvent number density and $N(r)$ is the number of water molecules within a sphere of radius r around the solute. The factor of $(4\pi\rho r^2)^{-1}$ is used to normalise the $g(r)$ to unity at positions in the bulk solvent that are far away from the selected atom in the solute [3].

The pair distribution function can be determined experimentally using X-ray, neutron or electron diffraction structure factors [4], or computed from the trajectories of a molecular dynamics simulation. This function is most significant in the case of solute-solvent interactions, which cannot be obtained directly from the experiment

due to the radial averaging of the experimental observables. However, computer simulations are able to provide the information in such cases [4].

Gases, liquids and solids all have very different pair distribution functions. In the gas phase pair distribution function, there is only one definite peak and then a rapid decay to unity indicating no long-range structure (or interactions). In a crystal, the pair distribution function has an infinite number of sharp peaks with well defined separations and peak heights, which indicates long-range structure (or interactions) typical of a crystal structure. The liquid pair distribution function is the intermediate case, in that it has small number of peaks, less than the solid but more than the gas, at short distances and then rapidly decays to unity at positions further from the solute [3, 4], Figure 3.1.

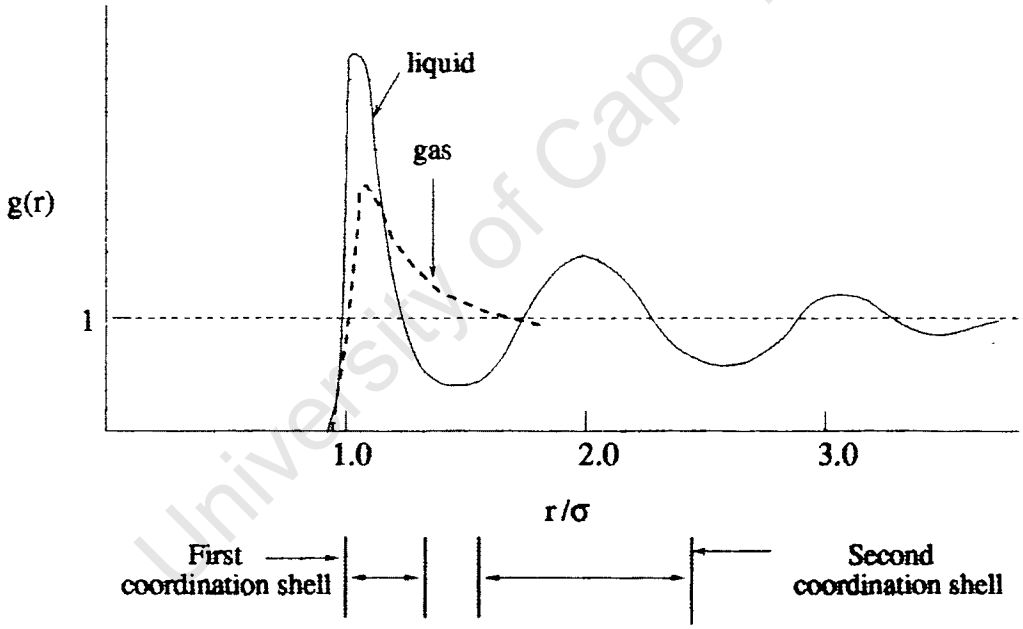

Figure 3.1: Schematic pair distribution functions for a hypothetical liquid and gas of the same molecule.

From the pair distribution function, it is possible to obtain the distance of the nearest neighbour to an atom of interest, as well as the total number of atoms that make up the solvation shell by using a running integral of the density.

Pair distribution functions are used to give invaluable insight into the structuring of liquids. However, they have the disadvantage of being radially averaged, and thus are unsuitable for investigating anisotropic liquid structuring. Other techniques have

thus been employed in order to investigate non-spherically symmetrical systems (Section 3.1.2).

3.1.2 Spatial Distribution Function

The spatial distribution function or water probability density calculation is a non-radially averaged method used to elucidate the water structuring about solute molecules, as radially averaged methods (pair distribution function, Section 3.1.1) obscure the anisotropic detail. Information available from the one-dimensional pair distribution functions are the number of neighbours for each site and the distance from each site. However, we would like to know, more precisely, the location of the water molecules around the solute and the role that they play in solvation, and this requires a more detailed three-dimensional picture. Analytical methods to map the average three-dimensional, anisotropic water structuring around carbohydrates have been developed in some extensive computational investigations [5, 6, 7, 8].

The method we employ is similar to that of Lui et al. for trehalose [7], with some adaptations to make the method more applicable for studying more flexible carbohydrates. The method produces a three-dimensional density matrix – a picture of the non-uniform distribution of water molecules – that may be contoured using standard three-dimensional graphing packages.

The Method

To generate the map of the anisotropic structure of water about a carbohydrate (or solute molecule), we first have to remove all of the rotational and translational diffusion throughout the simulation. To achieve this the instantaneous positions and orientations of the solute in each coordinate set, which make up the trajectory, are translated and rotated to obtain a best least-squares overlap with a reference frame, as shown in Figure 3.2. The coordinate transformations performed during this reorientation procedure are applied to all the atoms in the system. If two or more trajectories are to be compared, then the original reference frame is used in order that the resulting density maps will have the same relative orientation. Care must be taken when choosing the atoms for the least-squares fit, opting to use atoms that move around to much will result in unsatisfactory overlay of solute molecules and

concomitant blurring of the water densities [3].

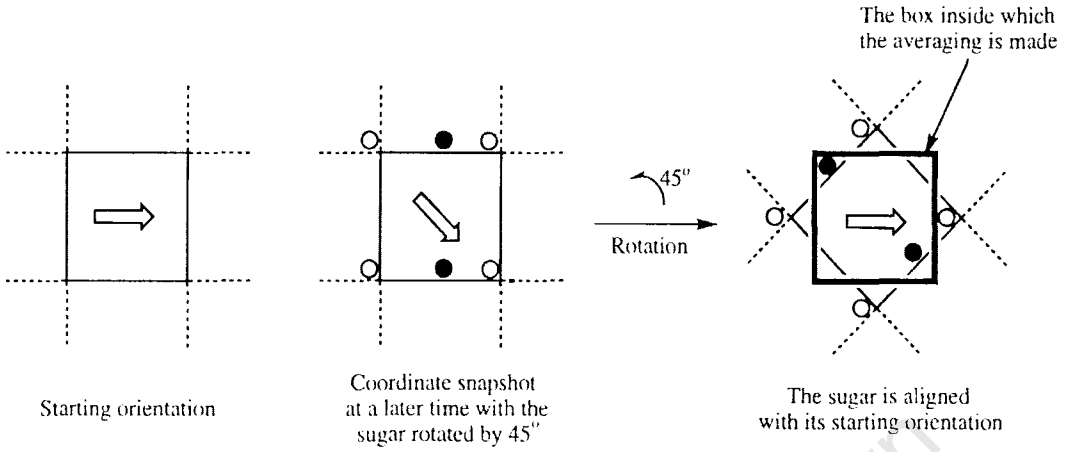

Figure 3.2: Schematic of the reorientation procedure used for the spatial distribution function. Taken from reference 3.

Previous studies using the method [7] were done on relatively inflexible carbohydrate molecules, such as monosaccharides. As there was little internal solute motion in these cases, all of the frames from the dynamics simulation could be used to generate the map of the water density. However, the carbohydrates we investigate are polysaccharides linked via a flexible $\alpha(1\text{-}4)$ glycosidic linkage. Using all the frames in the dynamics simulation (regardless of the ϕ, ψ conformations) results in unacceptable blurring of the very anisotropic water structuring one wishes to investigate. Therefore, it was necessary to have selection criteria for choosing frames with solute conformations that are similar enough to provide a clear, distinct picture of the anisotropic water structure. The obvious choice was to use the atoms that make up the ϕ, ψ dihedral angles that define the glycosidic linkage orientations, as these are the source of most of the internal molecular motions and thus have the greatest effect on the overall conformations. Then, when generating the map of the average water densities around the polysaccharide molecules, ranges about these are decided upon for all the ϕ, ψ dihedrals and only frames falling within these criteria are used to calculate the water density map. In choosing the range for ϕ, ψ angles, care must be taken. If too much molecular motion is allowed, the resulting water density maps will be blurred. This approach has the advantage that the water-structuring ability of different conformations of a polysaccharide molecule can be compared. This can help in explaining conformational preferences of polysaccharides in water: conformations

that impose greater structure on the surrounding solvent may be relatively entropically disfavoured [3].

When calculating the distribution of water around the solute, only the oxygen atoms were used to calculate the electron densities. The electron densities were used in order to obtain smooth distributions for the water probabilities around the solute, whereas an atom approach would give a scattered and obscured picture of the water probability density. A Gaussian distribution function centered on each water oxygen atom was used to approximate the distribution of electrons for each oxygen atom. The Gaussian distribution function used is of the form:

$$G(r) = elec \times \left(\frac{a}{\pi}\right)^{\frac{3}{2}} \times e^{-a \times r^2} \tag{3.2}$$

where *elec* is the number of electrons for the atom and a is calculated so that the function drops to 10% of its maximum value at the atom's van der Waals radius (1.4 Å for oxygen).

The simulation boxes are divided into bins approximately 0.5 Å wide, Figure 3.3, and the densities in each box is summed for all the selected frames from the dynamics simulation. The final water density matrix is normalised using equation 3.3 so that the density of the bulk water corresponds to a value of 1, and 50% above bulk density corresponds to a value of 1.5 etc.

$$dens_{norm}(i,j,k) = dens(i,j,k) \times \frac{n_{xbin} \times n_{ybin} \times n_{zbin}}{n_{electroins} \times n_{atom} \times n_{frames}} \tag{3.3}$$

In this equation, n_{xbin}, n_{ybin} and n_{zbin} are the number of divisions in the x, y and z directions in the simulation box, respectively. These are usually adjusted so that the box size is 0.5 Å. The number of atoms whose density is being calculated is represented by n_{atom} (in our case the water oxygens) and $n_{electrons}$ is the number of electrons for that atom (i.e. 8 for oxygen, etc). This method of calculating the water densities is essentially the same as a three-dimensional pair distribution function but without the radial averaging [3]. The resultant electron densities are contoured in

three-dimensions to produce diagrams of the water structuring around the solute molecule.

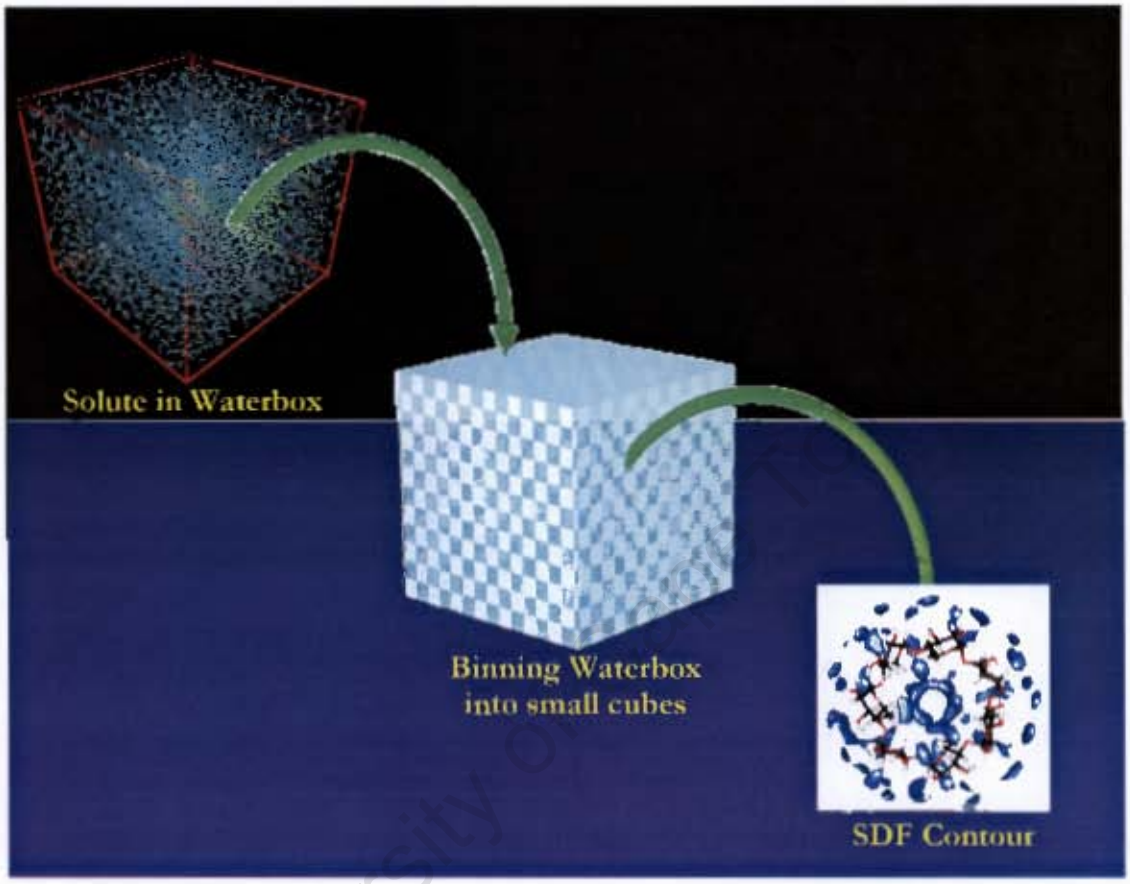

Figure 3.3: Schematic depiction f the binning procedure to produce the spatial distribution function.

3.13 Binding Free Energy

The relative binding free energies, of an enzyme with its ligand or between two monomers in solution, can be determined by using the Molecular Mechanics-Poisson-Boltzmann Solvent-Accessible surface area (MM-PBSA) methodology, which is an implicit solvent method. It is rapidly becoming one of the most popular methods for calculating relative binding free energies because of its statistical mechanics foundations and overall good accuracy [9].

The MM-PBSA calculation of the relative binding free energy takes into account the gas phase free energy, solvation free energies of the complex and free monomers and the entropy of all conformations in the simulation. The overall equation is as follows:

$$\Delta G_{bind} = \Delta G_{gas} + \Delta G_{solv}{}^{AB} - \left(\Delta G_{solv}{}^{A} + \Delta G_{solv}{}^{B}\right) \tag{3.4}$$

The gas phase free energy can be decomposed into internal energy and entropy components using the following approximation and neglecting volume changes upon binding:

$$\Delta G_{gas} = \Delta H_{gas} - T\Delta S = \left(\Delta U_{gas} + p\Delta V\right) - T\Delta S \approx \Delta U_{gas} - T\Delta S \tag{3.5}$$

and hence the MM/PBSA equation becomes:

$$\Delta G_{bind} = \Delta U_{gas} + \Delta G_{solv}{}^{AB} - \left(\Delta G_{solv}{}^{A} + \Delta G_{solv}{}^{B}\right) - T\Delta S \tag{3.6}$$

The internal energy, ΔU_{gas}, is equal to the potential energy, of the solute, which can be obtained from the molecular mechanics potential energy function (section 2.2.1). The complex's solvation free energy is subtracted by the solvation free energies of the monomers, which make up the complex. The entropy is calculated, at a specified temperature, for all the conformations in simulation.

The solvation free energy, ΔG_{solv}, can be decomposed into two parts: the polar and non-polar contributions.

$$\Delta G_{solv} = \Delta G_{solv}{}^{polar} + \Delta G_{solv}{}^{non-polar} \tag{3.7}$$

The polar or electrostatic contribution is calculated using the Poisson-Boltzmann (PB) equation and the non-polar contribution is calculated using the solvent-accessible surface area (SASA) of the solute molecule.

The PB equation (equation 3.8) is a differential equation that describes the electrostatic interaction between molecules and ionic solutions. It treats the solvent using a continuum model which lowers the computational cost by implicitly accounting for solvent effects using a simple dielectric model [10].

$$\Delta G_{solv}{}^{polar} \propto \vec{\nabla} \cdot \left[\varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})\right] = -4\pi\rho_{sol}(\vec{r}) - 4\pi\sum_i c_i{}^{\infty} z_i \lambda(\vec{r}) q e^{-\frac{z_i q \phi(\vec{r})}{k_B T}} \quad (3.8)$$

In this equation $\varepsilon(\vec{r})$ is the position dependent dielectric constant and $\phi(\vec{r})$ the electrostatic potential. $\rho_{sol}(\vec{r})$ is the charge density of the solute, c_i^{∞} is the concentration of ion, *i* (a distance infinity from the solute), z_i is the charge on the ion and $\lambda(\vec{r})$ is a factor that is position dependent to the accessibility of position *r* to ions in solution. At low electrostatic potentials, the PB equation can be linearised, which is mostly the case for neutral biomolecules such as carbohydrates and CDs. The non-linear and linear PB equations are solved numerically using the finite difference method that splits the system into uniform grids; a dielectric is defined for solute and solvent with a definition of a boundary between the two. There are three ways of calculating the boundary conditions: focussing, Coulombic and dipolar. The focussing boundary condition is a series of calculations for the system that occupies a greater fraction of the grid and uses the potential map from the previous calculation, and thus it focuses to the solution. The Coulombic boundary condition is calculated using the sum of the Debye-Hückel potentials generated by all charges and the dipolar boundary condition calculations approximate the Debye-Hückel dipole resultant from the molecular charge distances [10, 11]. The Poisson-Boltzmann equation can be used to generate an electrostatic potential map for the solute.

The solvent-accessible surface area (SASA) is the surface area of a biomolecule that is accessible to the solvent, and has the units of Å^2. SASA is typically calculated using a "rolling ball" algorithm developed by Shrake and Rupley [12]. The algorithm uses a sphere (of solvent) of a particular radius to "probe" the surface of the solute molecule. The choice of the probe radius does have an affect on the observed surface area, as using a smaller probe radius detects more surface details and therefore reports a larger surface area. A typical value of 1.4 Å is used, which approximates the radius of a water molecule. Another factor that affects the SASA is the definition of the van der Waals radii of the atoms of the solute molecule. If the hydrogens atoms are given a van der Waals radius then they are used in the calculation of the SASA, whereas the hydrogen atoms may be implicitly included in the atomic radius of the heavy atoms, with a measure called the "group radii" [12]. The non-polar contribution to the free

energy of solvation is then proportional to the SASA weighted by a heuristic surface tension term.

$$\Delta G_{solv}^{\ non-polar} = 0.00542 \times SASA + 0.92 kcal/mol \tag{3.9}$$

When calculating the entropy of the system, we approximate it using the quasiharmonic which assumes the solute to be a system of coupled harmonic oscillators, i.e. "ball and spring" model, atoms are represented by balls or spheres and bonds are represented as springs. The entropy is then calculated using

$$\Delta S \approx \frac{k_B}{T} \ln \det \left[I + \frac{k_B T e^2}{\hbar^2} M\sigma \right] \tag{3.10}$$

where I and M are the identity and mass matrices respectively and σ the covariance. This method has the advantages, over other methods, in that one obtains in an exact manner the entropy; it can give the quasiharmonic contribution to the enthalpy and also a decrease in computation time is achieved [13].

3.1.4 Time Correlation Functions

The 1950s saw the beginning of the development of a new approach to transport properties that has now become the most interesting and productive area of non-equilibrium statistical mechanics. This area showed that the phenomenological coefficients describing many transport processes and time-dependent phenomena in general could be written as integrals over a type of function called a time correlation function. Time correlation functions play a similar role in non-equilibrium statistical mechanics as the partition function plays in equilibrium statistical mechanics [14]. The advantage of time correlation functions is that the resulting formulas for the transport coefficients do not depend on the details of any particular model and are not limited to any particular density region. For example, the self-diffusion coefficient can be expressed in terms of a velocity time correlation function by:

$$D_s = \frac{1}{3} \int_0^\infty \langle \mathbf{v}(t) \cdot \mathbf{v}(0) \rangle dt \tag{3.11}$$

This expression is valid for any density, for angle-dependent intermolecular forces, for polyatomic molecules and generally any classical system in which the diffusion (or some transport process) is governed by the diffusion equation (or by some transport process equation) [14]. Since the state of equilibrium is unique, a single partition function gives all the thermodynamic properties, but since there are many different kinds of non-equilibrium states, we shall need a different time correlation function for each type of transport process. Hence the various time correlation functions play a similar role as the single partition function.

Time correlation functions are a time-dependent measure of the correlation of spontaneous fluctuations in a system. Let $\delta A(t)$ represent the instantaneous deviation or fluctuation in $A(t)$ from its time-independent equilibrium average, $\langle A \rangle$

$$\delta A(t) = A(t) - \langle A \rangle \tag{3.12}$$

Its time evolution is governed by microscopic laws. For classical systems, A depends upon time via the time dependence of the coordinates and the momenta in the system:

$$\delta A(t) = \delta\left(t; r^N, p^N\right) = \delta A\left[r^N(t), p^N(t)\right] \tag{3.13}$$

Unless A is constant of the motion (e.g. the energy), $A(t)$ will look chaotic even in an equilibrium system. While the equilibrium average of $\delta A(t)$ may seem "uninteresting" (i.e. $\langle \delta A \rangle = 0$), the generalised order parameter, S^2, is an important average of the Legendre polynomial. Also, one can extract nonchaotic information by considering the equilibrium correlations between the fluctuations at different times. The correlation between $\delta A(t)$ and an instantaneous or spontaneous fluctuation at time zero is:

$$C(t) = \langle \delta A(t) \delta A(0) \rangle = \langle A(0) A(t) \rangle - \langle A \rangle^2 \tag{3.14}$$

and the averaging is performed over the initial conditions. Thus for a classical system

$$C(t) = \int dr^N dp^N f\left(r^N, p^N\right) \delta A\left(0; r^N, p^N\right) \delta A\left(t; r^N, p^N\right) \tag{3.15}$$

where $f(r^N, p^N)$ is the equilibrium phase space distribution function.

In an equilibrium system, the correlations between the dynamical variable, *A*, at different times should depend upon the separation between these times only and not the absolute value of time [15]. Thus the non-normalised time correlation function is:

$$C(t) = \langle \delta A(t') \delta A(t'') \rangle, \text{ for } t = t''-t' \tag{3.16}$$

and the normalized time correlation function

$$c(t) = \frac{\langle \delta A(t') \delta A(t'') \rangle}{\langle (\delta A(0))^2 \rangle} \tag{3.17}$$

At short times,

$$c(0) = \frac{\langle \delta A(0) \delta A(0) \rangle}{\langle (\delta A(0))^2 \rangle} = \frac{\langle (\delta A(0))^2 \rangle}{\langle (\delta A(0))^2 \rangle} = 1 \tag{3.18}$$

At large times $\delta A(t)$ will become uncorrelated to $\delta A(0)$. Thus

$$c(t) \to \frac{\langle \delta A(t) \rangle \langle \delta A(0) \rangle}{\langle (\delta A(0))^2 \rangle}, \text{ as } t \to \infty \tag{3.19}$$

and since $\langle \delta A \rangle = 0$,

$$c(t) \to 0, \text{ as } t \to \infty \tag{3.20}$$

This is confirmed by the Schwartz inequality, which guarantee that the absolute value of *c(t)* lies between 0 and 1, with values close to 1 indicating a high degree of correlation, between the dynamical variables. Values tending toward zero indicating

lesser and lesser correlation and a value of zero indicating no correlation, Figure 3.4 [15].

This decay of correlations with increasing time is the "regression of spontaneous fluctuations" referred to as Onsager's hypothesis. The hypothesis says that the relaxation of macroscopic non-equilibrium disturbances is governed by the same laws as the regression of spontaneous microscopic fluctuations in an equilibrium system [15]. In other words a system, which has been disturbed, will relax or tend toward the equilibrium state much in the same way as spontaneous fluctuations occur in an equilibrium system.

For a correlation with the same dynamical variable, A, with a time integral of $t = 0$ to $t = \infty$, one obtains an *autocorrelation function*, Figure 3.4. These functions are of particular interest because:

- they give a clear picture of the dynamics in the solution,
- their time integrals τ may often be directly related to macroscopic transport properties
- and their Fourier transform $I_{AA}(\omega)$ may often be related to the experimental spectra [3].

Correlation between two different dynamical variables, A and B, results in a *cross-correlation function*:

$$C(t) = \langle \delta A(t) \delta B(0) \rangle \tag{3.21}$$

and the normalised cross-correlation function

$$c(t) = \frac{\langle \delta A(t) \delta B(0) \rangle}{\langle \delta A(0) \delta B(0) \rangle} \tag{3.22}$$

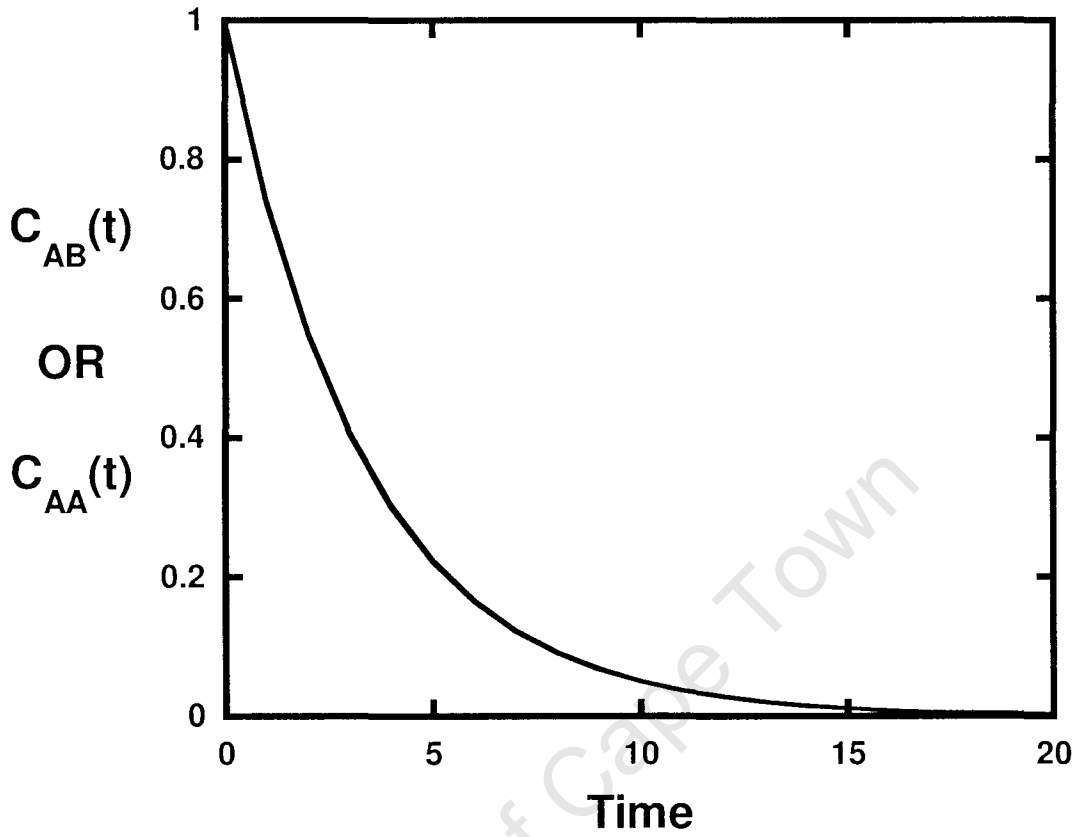

Figure 3.4: Normal exponentially decaying nature of time correlation function (AB) and autocorrelation function (AA).

If the decay, of the time correlation function, is assumed to be exponential, with the general form $Cexp(-t/\tau)$; with the time, τ, is known as the correlation time. The time correlation function gives an indication of how quickly the variables become uncorrelated (the correlation time) as well as any regular motion (manifested as later peaks in the function). Further insight into the frequency of such regular motion may be obtained by calculating the power spectrum of the correlation function by Fourier transformation [4]:

$$I(\omega) = \int_0^\infty \cos \omega t \langle \delta A(t) \delta B(0) \rangle dt \tag{3.23}$$

This spectrum contains peaks corresponding to the regular motions in the time correlation function, giving an indication of what motions are responsible for the decay and their frequency.

RMSD Time correlation Functions

The root mean square deviation (RMSD) time correlation function is calculated in a similar manner to the time correlation functions. The function is calculated as the correlation that is the best RMSD fit between structures at time *t'* and *t"*, with a time integral of $t = 0$ to $t = \infty$, and the correlation time, τ_{rmsd}, which relates how quickly a solutes equilibrium structure is obtained, Figure 3.5 [4]. The resulting c_{rmsd} takes on values between 0 and $RMSD_{max}$ (which is the maximum RMSD between structures of the solute). A value of 0 indicates perfect correlation or no RMS deviation; values tending toward $RMSD_{max}$ indicate an increase in RMS deviation and a value of $RMSD_{max}$ indicate that a maximum RMS deviation and equilibrium structure has been reached. The rise of the function indicates how quickly the internal degrees of freedom or variables become uncorrelated with time. The normalised RMSD time correlation function is:

$$c_{rmsd} = \left(\frac{\langle \delta_{rmsd}(0) - \delta_{rmsd}(t) \rangle}{\langle \delta_{rmsd}(0)^2 \rangle} \right) \qquad (3.24)$$

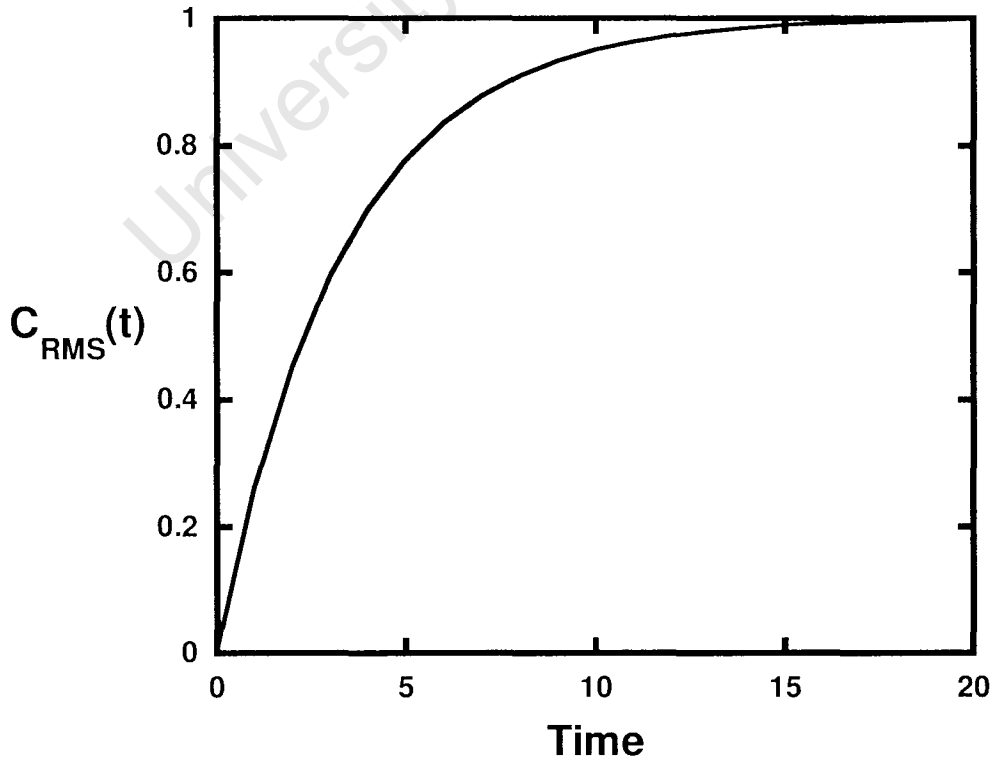

Figure 3.5: Exponential growth typical of normalised RMSD time correlation function (RMSD TCF).

Residence Time

The time that a water molecule spends in the first solvation shell, of the solute, before diffusing back into the bulk solution is known as the residence time. The residence time is calculated from a number correlation function, similar to a time correlation function, except that the dynamical variable has a value of either 0 or 1. The correlation time, τ, calculated from the time correlation function is the average residence time that a water molecule will spend in the first solvation shell. A spherical cutoff is used in the calculation; the cutoff radius extends to slightly more than the distance of the first solvation shell from the centre of the solute molecule.

$$C_{res}(t) = \frac{1}{N_t}\sum_{n=1}^{N_t}\sum_{j} C_j(t_n, t; t^*) \tag{3.25}$$

To obtain a precise definition for the residence time, we introduce a number correlation function $C_j(t_n, t; t^*)$. This is a property of the water molecule j and is equal to either 1 or 0. Water molecule j takes a value of 1 if the water molecule lies within the first solvation shell of the solute at both time steps t_n and $t + t_n$, and in the interim does not leave the solvation shell for any continuous period longer than t^*. A value of 0 is assigned to water molecule j if it is outside the first solvation shell or leaves the solvation shell for a period longer than t^*. A time period, called t^*, is used to take account of water molecules that leave the first solvation shell only temporarily and return to it without ever having properly entered the bulk solution. A typical value for t^* is 2 ps, since this is the residence time for pure water [16].

Transport Properties

Time correlation functions are not only useful for analysing conformational events; they can also be used to calculate transport properties such as diffusion and viscosity and physical properties such as the shear strain. Transport coefficients are defined in terms of the response of a system to a perturbation. For example, the diffusion coefficient relates the particle flux to a concentration gradient, while the shear viscosity is a measure of a shear stress induced by an applied velocity gradient [1].
One useful relation is the self-diffusion coefficient, D_s, defined by:

$$D_s = \frac{1}{3}\int_0^{\infty} \langle \mathbf{v}(t) \cdot \mathbf{v}(0) \rangle dt \tag{3.27}$$

where $\mathbf{v}(t) = \frac{d\mathbf{r}(t)}{dt}$ is the velocity of the particle or molecule in question. At long times, this may be approximated by the Einstein relation:

$$2tD_s = \frac{1}{3}\left\langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \right\rangle \tag{3.28}$$

where $\mathbf{r}(t)$ is the displacement and $|\mathbf{r}(t) - \mathbf{r}(0)|^2$ is the mean square displacement of the particle.

3.2 Experimental Analysis

3.2.1 NMR Spectroscopy

Nuclear magnetic resonance spectroscopy, commonly known as NMR spectroscopy, is the name given to the technique which utilizes the magnetic properties of certain nuclei. All nuclei containing odd numbers of protons or neutrons have an intrinsic magnetic moment and angular momentum, most often termed the spin of the nucleus. In principle, NMR is applicable to any nucleus possessing spin. The most commonly measured nuclei are hydrogen-1 (^{1}H, the most abundant isotope at natural abundance) and carbon-13 (^{13}C), although nuclei from isotopes of other elements (e.g. ^{15}N, ^{19}F, ^{23}Na, ^{35}Cl, ^{195}Pt) can also be observed.

In an NMR experiment a changing magnetic field is applied to the sample, which induces transitions between different energy levels. The energy absorbed from the field B_o is then observed as a signal, where the intensity is proportional to the population difference N_β-N_α and thus to the sample concentration.

A spinning charge generates a magnetic field, Figure 3.6a, and the resulting spin-magnet has a magnetic moment (μ) proportional to the spin.

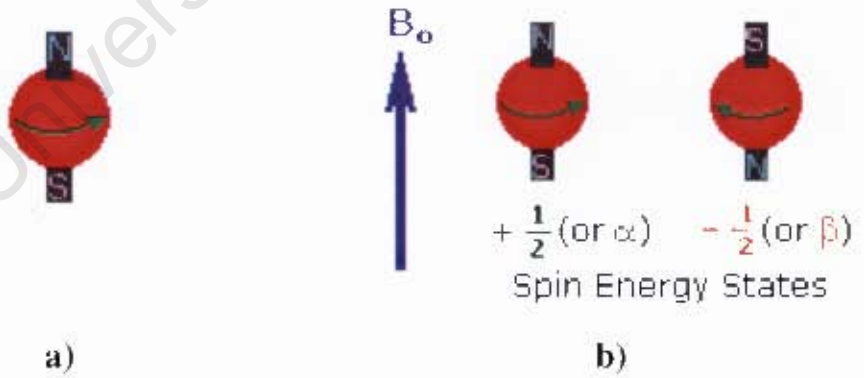

Figure 3.6: (a) Any nucleus which possesses spin generates a magnetic field, as shown. (b) When an external magnetic field is applied, the spins can be aligned (+½, lower energy state) or opposed to the external field (–½, higher energy).

In the presence of an external magnetic field (B_0), two spin states exist, +½ and –½. The magnetic moment of the lower energy +½ state is aligned with the external field, but that of the higher energy –½ spin state is opposed to the external field, Figure 3.6b.

The difference in energy between the two spin states is dependent on the external magnetic field strength, and is always very small. The diagram in Figure 3.7a illustrates that the two spin states have the same energy when the external field is zero, but diverge as the field increases. At a field equal to B_x a formula for the energy difference is given (remember I = ½ and μ is the magnetic moment of the nucleus in the field). Strong magnetic fields are necessary for NMR spectroscopy. The international unit for magnetic flux is the Tesla (T). Modern NMR spectrometers use powerful magnets having fields of 1 to 20 T. Even with these high fields, the energy difference between the two spin states is less than 0.1 cal/mol. To put this in perspective, recall that infrared transitions involve 1 to 10 kcal/mol and electronic transitions are nearly 100 times greater.

For NMR purposes, this small energy difference (ΔE) is usually given as a frequency in units of MHz, ranging from 20 to 900 Mz, depending on the magnetic field strength and the specific nucleus being studied, Figure 3.7a. Irradiation of a sample with radio frequency (rf) energy corresponding exactly to the spin state separation of a specific set of nuclei will cause excitation of those nuclei in the +½ state to the higher –½ spin state. Note that this electromagnetic radiation falls in the radio and television broadcast spectrum. NMR spectroscopy is therefore the energetically mildest probe used to examine the structure of molecules.

Figure 3.7b displays the energy differences for the proton spin states, the approximate frequency that correspond to the spin state energy separations for a proton in an external magnetic field of 2.35 T [17].

NMR is extensively used for spatial conformational analysis of molecules ranging from disaccharides to macromolecular protein structures. In the liquid phase the most useful NMR spectral parameters related to geometry are the nuclear Overhauser effects (NOEs), rotating nuclear Overhauser effects (ROEs) and scalar and dipolar spin-spin coupling constants reflecting, respectively, internuclear distances and torsion angles. Other properties such as the diffusion coefficient, rotational diffusion relaxation and hydrogen bonding (using variable temperature NMR) can also be calculated [20, 21].

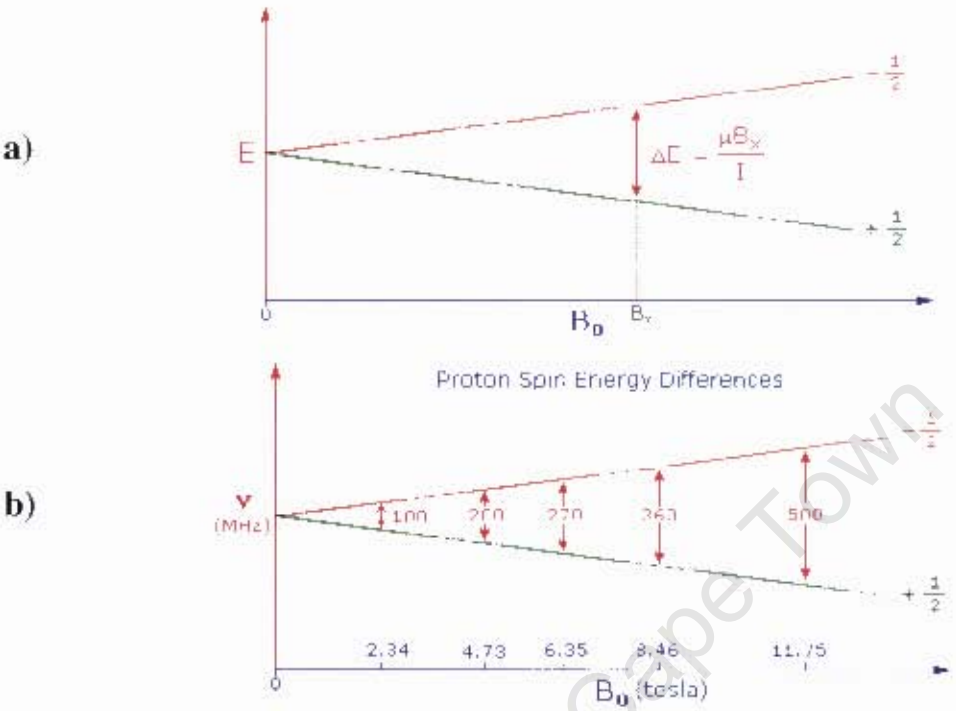

Figure 3.7: (a) General energy difference as a function of the strength of the external magnetic field. (b) Proton spine energy difference as a function of the external magnetic field. The energy difference between the aligned and opposed states are highly dependent on the strength of the magnetic field applied [17].

Spectral Parameters

The spectral parameters consist of numerous observables obtained from an NMR experiment [20]. Spin-spin (or J-) and dipolar couplings, as well as Nuclear Overhauser Enhancements (NOEs), are discussed below.

Spin-Spin Coupling

The nomenclature for the common coupling constants i.e. proton-proton and proton-carbon, are J_{HH} and J_{CH} and the number of bonds through which the nuclei are coupled is given by n preceding the J, such that we have ${}^{n}J$. For the purpose of this study, only the ${}^{3}J$-vicinal coupling is considered for determining the torsional or dihedral angles.

The importance of spin-spin coupling constants comes into context when one considers the factors that influence the coupling constants. They are [20, 21]:

1. The hybridisation of the atoms involved

2. Bond lengths
3. Bond angles and dihedral angles
4. The effects of neighbouring bonds and lone-pairs
5. Substituent effects

More specifically, if we obtain the $^3J_{HH}$ and $^3J_{CH}$ vicinal couplings, we can determine the bond angles, as well as dihedral angles and use this information to help in structure determination. To better understand vicinal couplings as a function of the dihedral angles in saturated systems, it is best to apply a Karplus relationship curve, which relates the value of the ^3J-coupling to the dihedral angle ϕ, Figure 3.8.

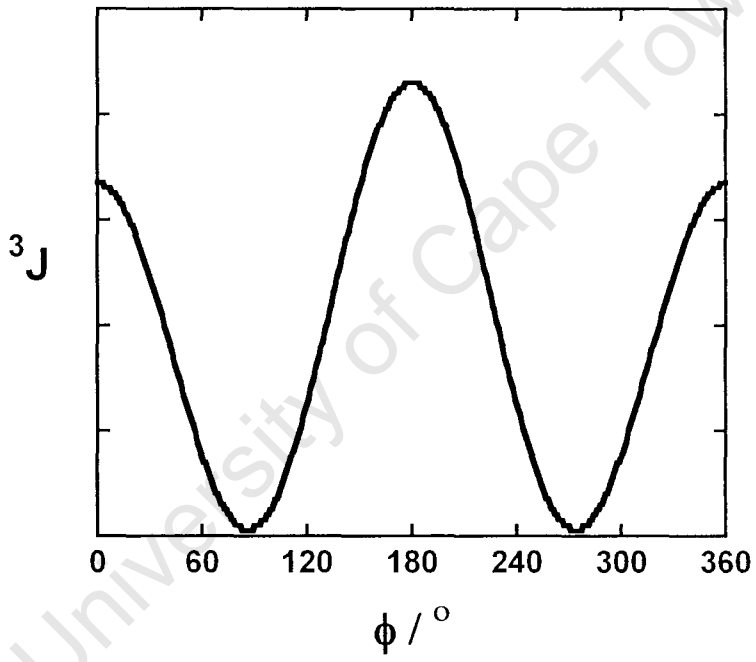

Figure 3.8: Graphical representation of the Karplus relationship between the ^{3}J coupling constant and the dihedral angle.

The Karplus equation, which was introduced by Martin Karplus in 1959 [22], describes the correlation between ^3J-coupling constants and dihedral angles in NMR spectroscopy. It can be used as a tool in determining the conformations and configurations of saturated systems and has been extensively used in determining dihedral angles across numerous linkages, which include carbon, oxygen and nitrogen atoms [23-27], such as for determining backbone dihedral angles in proteins. The Karplus equation's general form is:

$$J(\phi) = A\cos^2\phi - B\cos\phi + C \tag{3.29}$$

where J is the ^3J-coupling constant, ϕ is the dihedral angle, and A, B and C are empirically-derived parameters whose values depend on the atoms and substituents involved [22, 23]. The relationship may be expressed in a variety of ways e.g. involving $cos2\phi$ rather than $cos^2\phi$ - this leads to different numerical values of A, B and C but does not change the nature of the relationship. The magnitude of these couplings is generally smallest when the dihedral angle is close to 90° and largest at angles of 0 and 180°. The ^3J-couplings in this thesis were calculated using the Karplus equation with Serianni et al. parameterisation [27]:

$$J(\phi) = 7.49\cos^2\phi - 0.96\cos\phi + 0.15 \tag{3.30}$$

Magnetic Dipole-Dipole Interaction

The magnetic dipole-dipole interaction, also called dipolar coupling, refers to the direct interaction between two magnetic dipoles. The energy of the interaction is as follows:

$$H = -\frac{\mu_0}{4\pi r_{ij}^3}\left(3(m_i \cdot e_{ij})(m_j \cdot e_{ij}) - m_i \cdot m_j\right) \tag{3.31}$$

where e_{ij} is a unit vector parallel to the line joining the centres of the two dipoles. r_{ij} is the distance between two dipoles, m_i and m_j.

For two interacting nuclear spins:

$$H = -\frac{\mu_0 \gamma_i \gamma_j}{4\pi r_{ij}^3}\left(3(I_i \cdot e_{ij})(I_j \cdot e_{ij}) - I_i \cdot I_j\right) \tag{3.32}$$

γ_i, γ_j and r_{ij} are gyromagnetic ratios of two spins and spin-spin distance respectively.

The direct dipole-dipole coupling is very useful for molecular structural studies, since it depends only on known physical constants and the inverse cube of internuclear distance. Estimation of this coupling provides a direct spectroscopic route to the distance between nuclei and hence the geometrical form of the molecule. Although internuclear magnetic dipole couplings contain a great deal of structural information, in isotropic solution, they average to zero as a result of rotational diffusion. However, their effect on nuclear spin relaxation results in measurable nuclear Overhauser effects (NOEs).

The residual dipolar coupling (RDC) occurs if the molecules in solution exhibit a partial alignment leading to an incomplete averaging of spatially anisotropic magnetic interactions i.e. dipolar couplings. RDC measurement provides information on the global folding of the protein-long distance structural information. It also provides information about "slow" dynamics in molecules [17].

Nuclear Overhauser Effect (NOE)

NOE's arise between two spins that possess dipolar coupling (i.e. through-space coupling) and not from scalar coupling (spin-spin coupling which produces the multiplets in proton spectra). Since NOE depends on the inverse of the distance to the sixth power, only short-range dipolar couplings are heavily weighted. This dependence on distance allows for an estimation of the internuclear distance. Geometric and structural information can also be extracted from the NOESY spectra. Care must be taken as many processes may lead to reduced NOEs, including spin-lattice relaxation, temperature, increased solvent viscosity, increased molecular weight, and dissolved paramagnetic impurities including oxygen [4].

The homonuclear ^{1}H 2D NOESY (Nuclear Overhauser Enhancement SpectroscopY) experiment measures the cross-relaxation of hydrogen nuclei in a magnetic field. "Cross relaxation" simply refers to the relaxation of one nucleus due to a chemically different one, while "direct relaxation" would result from other relaxation mechanisms in the sample or from relaxation due to chemically equivalent nuclei. This can be illustrated by an uncoupled AX system, shown as a state diagram in Figure 3.9 (spin states α and β for each nucleus i and j). Relaxation can occur via either the normal single quantum transitions (w_1^i and w_1^j) or by the double or zero quantum transitions, w_2^{ij} and w_0^{ij}, respectively, where the w_i refers to the transition rate [4].

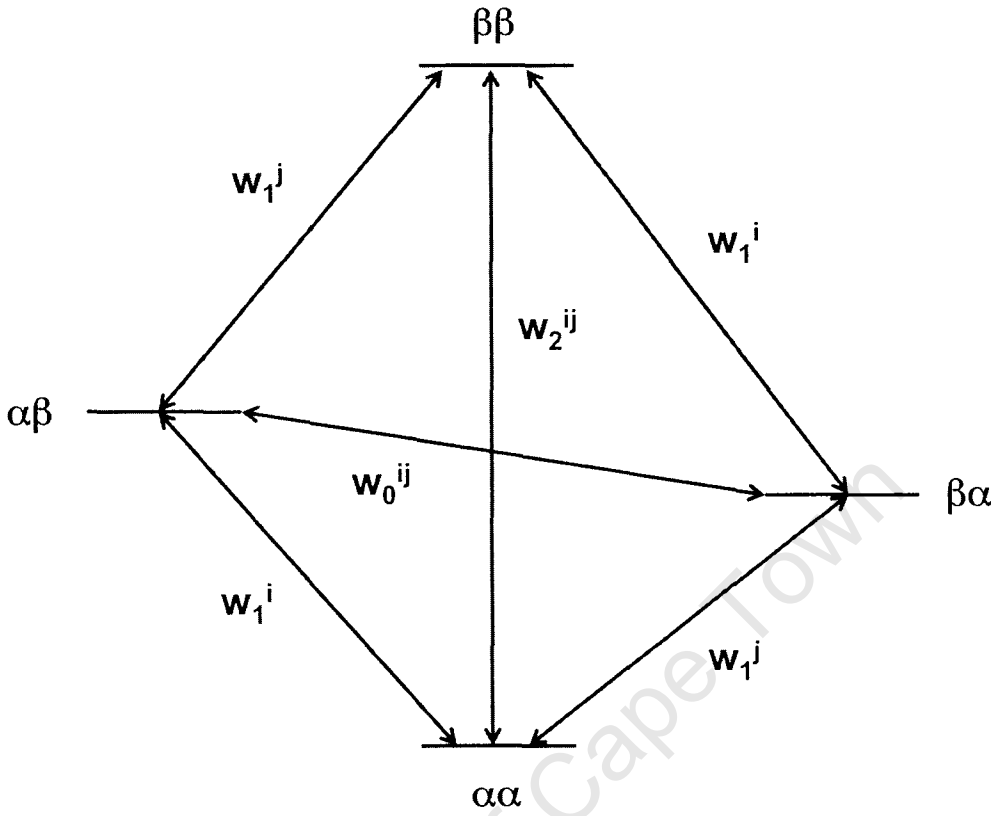

Figure 3.9: State diagram for an uncoupled AX system, shown as a state diagram. Both direct and cross relaxation processes are shown [4].

These latter two processes are responsible for cross relaxation [18]. The net cross relaxation and direct relaxation are given by:

Cross relaxation:

$$\sigma_{ij} = n_i\left(w_2{}^{ij} - w_0{}^{ij}\right) \tag{3.33}$$

Direct relaxation:

$$\rho_{ii} = 2(n_i - 1)\left(w_1{}^{ii} - w_2{}^{ii}\right) + \sum_{i \neq j}\left(w_0{}^{ij} + 2w_1{}^{i} + w_2{}^{ij}\right) \tag{3.34}$$

Where n_i is the number of equivalent protons in a group and w_i and w_j are the transitions rates [19]. A transition corresponding to a given frequency is promoted by molecular motion at the same frequency. Small molecules in non-viscous solvents tumble at rates around 1011 Hz, while larger molecules such as proteins tumble at rates around 107 Hz. For small molecules, w_2 will be greater than w_0 and NOE enhancements will be positive. For larger molecules w_0 will become greater than w_2 and NOE enhancements will be negative.

3.2.2 Ultrasonic Interferometry

The propagation of sound in a medium is a wave phenomenon and thus exhibits interference; in fact, interference is observed when two coherent waves are allowed to be superimposed in space. When two waves interact with a particle of the propagation medium, the resultant displacement, velocity and pressure are the vector sum of the effects due to the separate waves. If the final distribution of energy for this particle is different from what it would have been for each separate wave, interference has occurred. This phenomenon may be used to determine the relative physical properties of the media through which these waves have travelled prior to their superimposition. Thus a properly designed acoustic interferometer (in this case an ultrasonic interferometer) behaves as a differential device, since it transforms the phase difference of the initial coherent wave into intensity modulation [28].

Ultrasonic interferometry, based on the aforementioned phenomenon of interference, presents several advantages as far as accuracy and reproducibility of the measurements are concerned. In fact, they are not based on the measurement of absolute values, like other techniques, but rather on the determination of differences between the reference signals and the perturbed ones (e.g. phase shifts, frequency shifts) [28, 29]. Such differences are responsible for destructive or constructive interference between signals, which can be measured with high precision by using simple experimental setups. Other advantages of ultrasonic interferometry are:

1. Its applicability to very small samples
2. Its ability to simultaneously measure both velocity and attenuation
3. The possibility of using noncontact detectors
4. The higher precision that may be obtained from direct measurement of the difference between unperturbed and perturbed quantities
5. High accuracy at high pressures and temperatures.

Ultrasonic Interferometry is a simple and direct device used to determine the ultrasonic velocity in liquids with a high degree of accuracy. A number of thermodynamic properties of a liquid system can be estimated directly from the measurement of ultrasonic velocity and density data. Ultrasonic propagation assumes

(i) that there is no loss in the fluid and (ii) an adiabatic nature [30]. Ultrasonic interferometry (ultrasonic speed and attenuation) is beginning to establish itself as an important tool for studying molecular interactions in solutions, as well as for investigating structural and physio-chemical behaviour of solutions.

Ultrasonic Speed Measurement

The ultrasonic interferometer is studied in detail in a comprehensive paper by W. G. Cady [31] and further extensive analysis was made by D. W. Dye [32]. G. W. Pierce [33] for the first time developed methods of high frequency interferometry. This is a continuous wave method of ultrasonic measurements in solution in the MHz region, which has been used for determination of wavelength, speed [34-37] and absorption coefficient [37, 38] in solution. Hubbard and co-workers [39] showed that the output current of an ultrasonic interferometry is a complicated function of the absorption coefficient, the reflection coefficient of the reflector and the distance moved by the reflector.

Description of the Apparatus

The speed of the ultrasonic waves, which travels through the solution, is measured using a multi frequency ultrasonic interferometer of single crystal, variable path type, Figure 3.10.

The apparatus consists of two parts:

1. the high frequency generator, and
2. the measuring cell

The High Frequency Generator

This is a high frequency crystal controlled oscillator in the form of a modified circuit, operating in the megacycle range. It is designed to excite the transducer, which is a quartz crystal fixed at the bottom of the measuring cell, to produce ultrasonic waves at its resonant frequency in the solution. The generator is provided with a DC micrometer to detect the current. There are two controls, one for the initial adjustment of the micrometer and the other for the purpose of controlling the sensitivity. There are provisions for two condensers marked as A and B, respectively, on the backside of the generator. These are required for the purpose of adjustment or

turning the instrument where the deflection in the micrometer is zero or becomes insufficient for any solution [30].

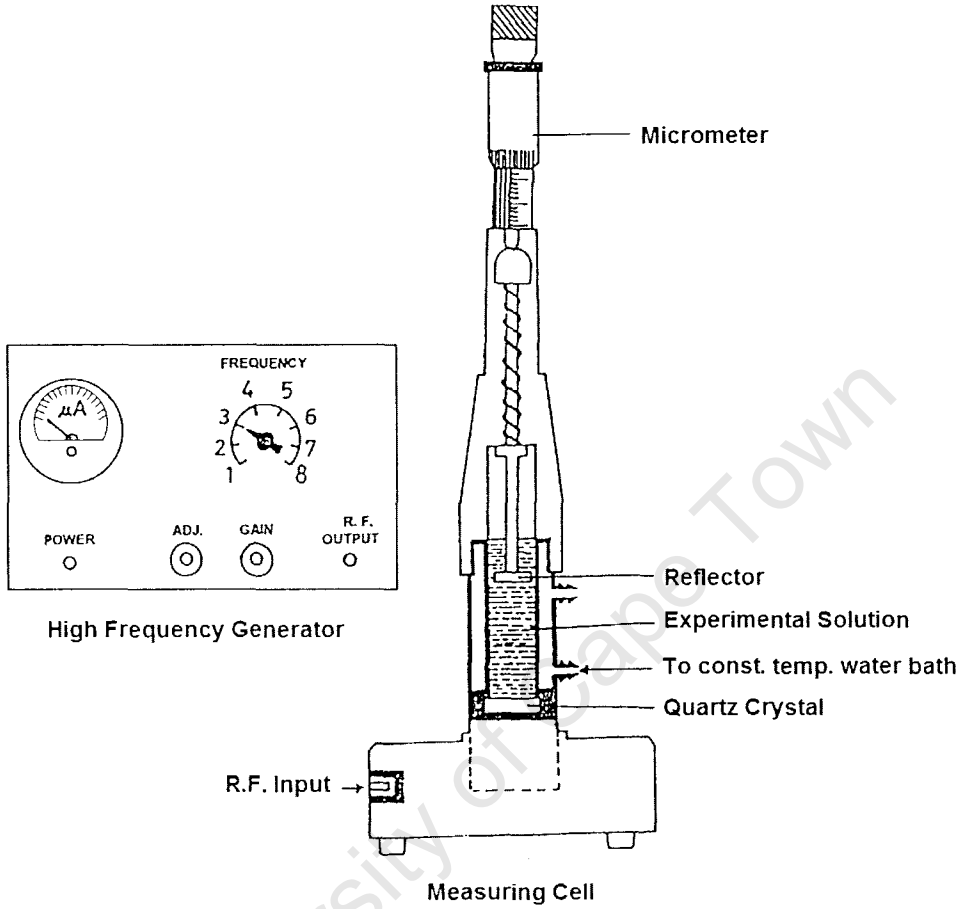

Figure 3.10: The ultrasonic interferometer instrument used in the experiments in this thesis.

The Measuring Cell

This is a cylindrical metal container placed vertically on a heavy metal base, which serves as the coupler between the quartz crystal and the high frequency generator. The measuring cell is a double-walled cell containing the experimental solution and around it there is also provision for circulation of water from the thermostatically regulated water bath, to maintain a constant temperature. The temperature can be controlled and kept constant with an accuracy of ± 0.02 ^{o}C. A quartz crystal acting as a transducer is fixed at the bottom of the cell and a movable metallic reflecting plate, controlled from the outside by micrometer screw, is kept parallel to the quartz crystal. The measuring cell can be easily dismantled into three parts: (i) metal base, (ii) container and (iii) metallic plate reflector – such that the

solution can be changed. The measuring cell is connected to the high frequency generator through a short co-axial cable [30].

Working of the Ultrasonic Interferometer

An ultrasonic wave is produced when a high frequency generator excites the quartz crystal, which acts as a generating transducer. The resulting ultrasonic wave travels through the solution toward a movable reflecting metallic plate, which is kept parallel to the quartz crystal. The same quartz crystal also acts as a detecting transducer. As the reflecting plate is made to approach or recede from the vibrating crystal, a change in amplitude and phase occurs. The reflector can be made to pass through successive positions of maxima and minima of the ultrasonic wave in the solution. When the distance between the movable metallic plate and the quartz crystal is exactly a multiple of the ultrasonic wavelength, standing waves are formed in the solution. In other words the incident and reflected ultrasonic waves are superimposed in the solution. The aforementioned change in amplitude and phase of the vibrating crystal gives rise to piezoelectricity, and the current can be measured. The acoustic propagation produces an electrical reaction on the oscillator of the quartz crystal and the anode current of the generator is a maximum when standing waves are formed in solution. If the distance between the reflector plate and the crystal is increased or decreased, and the variation is a multiple of half of the wavelength ($\lambda/2$), the anode current is also at a maximum [30].

Calculation of Ultrasonic Speed

Measurements of the ultrasonic speed, which travels through the solution, are based chiefly on finding the wavelength (λ) of the ultrasonic wave in the solution. When a multiple of half the wavelength is achieved, by moving the movable metallic plate, a maximum in the observed current is obtained. Determination of the wavelength and subsequent measurement of the frequency, the speed can be easily calculated using the relation:

$$u = \nu\lambda \tag{3.35}$$

where u is the ultrasonic speed, ν is the frequency and λ is the wavelength. The wavelength is calculated as:

$$d = \frac{(n-1)\lambda}{2} \tag{3.36}$$

where d is the total distance moved by the micrometer screw for a maximum or minimum deflection, n is number of maxima (or minima, ≈ 20) of anode current for a distance d. Hence the ultrasonic speed can be calculated using:

$$u = \frac{2\nu d}{n} \tag{3.37}$$

Therefore ultrasonic interferometry is a simple experimental procedure for determining ultrasonic speed in solution, and hence (in this case) the isentropic compressibility and hydration number of solutes in solution. The accuracy in the measurements of the ultrasonic speed can be within ± 0.05%, even at elevated pressures and temperatures [30].

Isentropic Compressibility

Isentropic compressibility, κ_s, is very useful property of a system to study, because it depends upon the compressible volume of molecules around a solute; this volume is very sensitive to the intermolecular interactions in solutions. The relationship between isentropic compressibility and ultrasonic speed is given by the Newton-Laplace equation:

$$\kappa_s = \frac{1}{\rho u^2} \tag{3.38}$$

where u is the measured ultrasonic speed and ρ is the density of the solution.

Like isentropic compressibility, intermolecular free length, L_f (the distance between the surfaces of molecules in solution), is another useful property which also depends on the intermolecular interactions in solution. If a strong interaction occurs between the molecules in solution, the compressibility may be negative [40]. In general, κ_s values depend on two factors:

1. An increase in free length, defined by Jacobson [41], due to loss of dipolar association or breaking of hydrogen bonds and unfavourable fitting of component molecules into each others structures.
2. A decrease in free length as a result of dipole-dipole interactions, hydrogen bonding or complex formation between the component molecules.

The first effect decreases the ultrasonic speed in solution, which results in an increase in the isentropic compressibility, whereas the second effect increases the ultrasonic speed that causes a decrease in the isentropic compressibility. The observed κ_s values are a net result of the two factors.

Hydration Number

By measuring the ultrasonic speed in a solution, we are able to determine the isentropic compressibility, κ_s, and from this we can then determine the hydration number for the solute using the expression:

$$n_h = \frac{n_w}{n_s}\left(1 - \frac{\kappa_s}{\kappa_{s,o}}\right) \tag{3.39}$$

where $\kappa_{s,o}$ is the isentropic compressibility of the pure water, and n_w and n_s are the number of moles of water and solute, respectively. Equation 3.39 implicitly assumes that: (i) n_h is the number of water molecules in the first solvation shell surrounding the solute, and (ii) these water molecules are trapped so tightly that they can be considered as incompressible [42].

References

1. M. P. Allen, D. J. Tildesley. *Computer Simulations of Liquids*. Oxford University Press, London, **1987**, Chapter 2.
2. P. J. Rossky, M. Karplus. Solvation. *J. Am. Chem. Soc.* **1979**, *101(8)*, 1913.
3. M. M. Kuttel. "Developing Analytical Tools for Saccharides in Condensed Phases", MSc Thesis, University of Cape Town, **1999**, Chapter 3.
4. R. Best. "Combined NMR and Simulation Study of Carbohydrate Linkage Dynamics", MSc. Thesis, University of Cape Town, **2000**, Chapter 2.
5. R. K. Schmidt, M. Karplus, J. W. Brady. *J. Am. Chem. Soc.* **1996**, *118*, 541.
6. Q. Lui, J. W. Brady. *J. Phys. Chem. B.* **1997**, *101(8)*, 1817.
7. Q. Lui, R. K. Schmidt, B. Teo, M. Karplus, J. W. Brady. *J. Am. Chem. Soc.* **1997**, *119(33)*, 7851.
8. Q. Lui, J. W. Brady. *J. Am. Chem. Soc.* **1996**, *118*, 12276.
9. F. Fogolari, E. Moroni, M. Wojciechowski, M. Baginski, L. Ragona, H. Molinari. *Proteins: Structure, Function, and Bioinformatics*. **2005**, *59*, 91.
10. I. S. Moreira, P. A. Fernandes, M. J. Ramos. *Journal of Molecular Structure*. **2005**, *729*, 11.
11. S. R. Edinger, C. Cortis, P. S. Shenkin, R. A. Friesner. *J. Phys. Chem. B.* **1997**, *101*, 1190.
12. A. Shrake, J. A. Rupley. *J. Mol. Biol.* **1973**, *79(2)*, 351.
13. I. Andricioaeli, M. Karplus. *J. Chem. Phys.* **2001**, *115(14)*, 6289.
14. D. A. McQuarrie. *Statistical Mechanic.* **2000**, University Science Books, Sausalito, California, Ch. 21.
15. D. Chandler. *Introduction to Modern Statistical Mechanics.* **1987**, Oxford University Press, Oxford, 238-242.
16. R. W. Impey, P. A. Madden, I. R. McDonald. *J. Phys. Chem.* **1983**, *87*, 5071.
17. M. H. Levitt. *Spin Dynamics: Basics of Nuclear Magnetic Resonance.* **2001**, Wiley, Chapter 12.
18. A. E. Derome. *Modern NMR Techniques for Chemistry Research.* **1989**, Pergamon, Chapters 1-8.
19. G. Wagner. *Prog. NMR. Spectr.* **1990**, *22*, 101.

20. H. Friebolin. *Basic One- and Two-Dimensional NMR Spectroscopy.* **1991**, VCH Verlagsgesellschaft mbH, D-6940 Weinheim (Federal Republic of Germany).
21. R. J. Abraham, J. Fischer, P. Loftus. *Introduction to NMR Spectroscopy.* **1988**, John Wiley & Sons: Chichester.
22. M. Karplus. *J. Chem. Phys.* **1959**, *30(1)*, 11.
23. M. Karplus. *J. Am. Chem. Soc.* **1963**, *85(18)*, 2870.
24. I. Tvaroska, M Hricovini, E. Petrakova. *Carbohydr. Res.* **1989**, *189*, 359.
25. Q, Xu, A. Bush. *Carbohydr. Res.* **1998**, *306*, 335.
26. L.-F. Kao, M. Barfield. *J. Am. Chem. Soc.* **1985**, *107*, 2323.
27. F. Cloran, I. Carmichael, A. S. Serianni. *J. Am. Chem. Soc.* **1999**, *121*, 9843.
28. N. K. Batra, P. P. Delsanto, A. Romano, M. Scalerandi. "Acoustic Wave Interferometers." *The Wiley Encyclopedia of Electrical and Electronics Engineering*. **1999**, John Wiley & Sons, Vol. 1, 133-139.
29. C. M. Fortunko et al. *Rev. Sci. Instrum.* **1992**, *63*, 3477.
30. B. Lal. "Thermophysical and Sonochemical Behaviour of Some Organic Liquid Mixtures". PhD. Thesis, Jamia Millia Islamia (Central University), **2004**, Chapters 2 and 3.
31. W. G. Cady. *Proc. I. R. E.* **1992**, *10*, 183
32. D. W. Dye. *Proc. Phys. Soc.* **1923**, *38*, 399.
33. G. W. Pierce. *Proc. Nat. Am. Acad.* **1925**, *60*, 269.
34. E. B. Freyer, J. C. Hubbard, D. M. Andrews. *J. Am. Chem. Soc.* **1979**, *51*, 759.
35. J. C. Hubbard. *Phys. Rev.* **1923**, *41*, 523.
36. J. C. Hubbard, A. L. Loomis. *Phil Mag.* **1928**, *5*, 1178.
37. E. Klein, W. D. Mershberger. *Phys. Rev.* **1931**, *37*, 760.
38. J. C. Hubbard. *Phys. Rev.* **1930**, *35*, 1442.
39. J. C. Hubbard. *Phys. Rev.* **1930**, *36*, 1668.
40. R. J. Fort, W. R. Moore. *Trans Faraday Soc.* **1965**, *61*, 2102.
41. R. V. G. Rao, V. Venkataseshaih. *Z. Phys. Chem.* **1969**, *242*, 193.
42. E. Junquera, D. Olmos, E. Aicart. *Phys. Chem. Chem. Phys.* **2002**, *4*, 352.

Chapter 4

Glucose Orientation and Dynamics in α-, β- and γ-Cyclodextrins *

4.1 Introduction

Cyclodextrins (CDs) are cyclic α(1-4)-linked carbohydrate oligomers constructed from glucose units. Three of the most important unmodified cyclodextrins are the α-, β- and γ-CDs, see Figure 4.1, which comprise 6, 7 and 8 units of a glucopyranose monomer, respectively [1, 2].

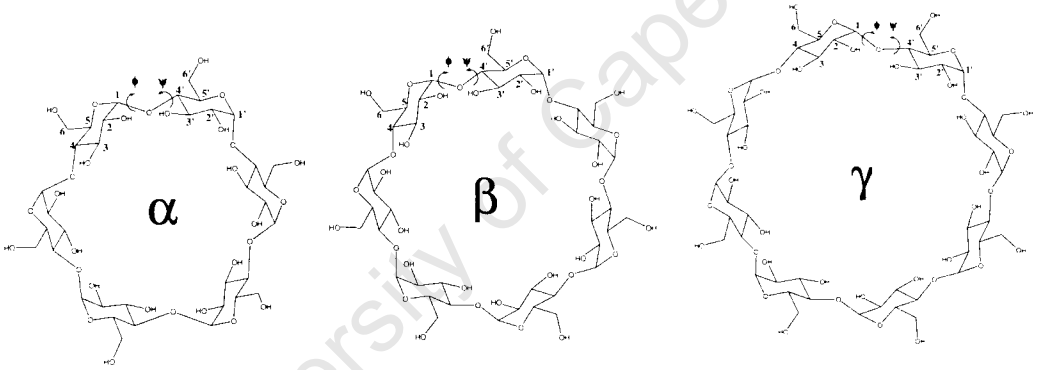

Figure 4.1: Cycloamylose (Cyclodextrins) showing 6 glucose monomers α-CD, 7 glucose monomers β-CD and 8 glucose monomers γ-CD. The dihedral angles are defined by the following atoms: ϕ using H1, C1, O4' and C4' and ψ using the C1, O4', C4'and H4'.

Cyclodextrins are produced from a variety of starch materials, and due to their unique structural features, these molecules are invaluable in a variety of industrial applications. The most important industrial sectors include pharmaceuticals, agricultural chemicals, cosmetics and foods. Almost all applications involve the ability of CDs to alter the physical, chemical and biological properties of guest molecules, through the formation of inclusion complexes. This effect is achieved by the encapsulation of guest molecules into the cones of the CDs. Small and larger aromatic substances complex best with β- and γ-CDs respectively, whereas aliphatic chains are more suited to α-CDs.

*** Submitted to the Journal of Physical Chemistry B**

The choice of CD used in drug delivery depends on requirements of the dosage form, the route of delivery, and the solubilising capacity required to handle the drug load. A vast amount of pharmaceuticals consist of low molecular weight aromatic moieties that are too large to fit into the α-CD cavity and while they do fit into the of γ-CD cavity it is often economically not viable. Therefore, β-CD has been traditionally identified as the preferred drug carrier. However, an inhibiting property of β-CD has been its relatively low solubility in water (18.8 mg/mL at 25 °C) compared with α-CD (145.0 mg/mL at 25 °C) and γ-CD (232.0 mg/mL at 25 °C) [3]. Consequently, although β-CD is best suited in most cases, due to its cavity size, its low intrinsic solubility in water makes it impractical to use. Because of commercial interest as well as the puzzling anomalous solubility trend of α-, β-, and γ-CDs, investigations into the underlying molecular reasons for these continue to attract substantial scientific attention [4, 5].

In general, CDs are only able to adopt a limited number of conformations in the aqueous medium due to the restrictions imposed on the rotation of sugar units about the inter-glycosidic bonds. The allowed conformations of CDs are also dependent on the intra- and inter-molecular forces. Thus, the most stable conformations of these cyclic oligomers will be those that balance these molecular interactions.

More than 20 years ago the Koehler et al. published a series of articles, using computational methods, addressing the phenomena of the unexpectedly low solubility of β-CD [6]. They investigated the characteristics of asymmetry motion of the macrocycles, mobility of the water molecules and the distribution of a variety of CD structural geometries [6]. More recently, other computational investigations of the hydration effects and conformational properties of the CDs were reported [4, 7]. Naidoo et al. reported earlier that the order of water structure around the β-CD is higher than in the case of the other two cycloamyloses and the differential solute solvent interactions of these three cyclic sugars may be due to differences in their conformational behaviour [8]. The conformations of CDs can be analysed on the molecular and monomeric length scales as are illustrated in Figure 4.2.

Studies of cyclodextrin conformation generally focuses only on the molecular length scales and the associated variables that govern the CD shapes. The most important of these are the rotations about the glycosidic bonds (the top row middle illustration in Figure 4.2) and the hydroxyl groups (top row last illustration in Figure 4.2). At the monomeric length scale the vertical and angular motion of the glucose

rings relative to the overall macrocyclic ring are the two fundamental variables that fully describe the monomer contributions to the overall CD conformations. Here we set out to address the underlying molecular and monomeric conformational reasons for the relative inflexibility of β-CD. The conformational behaviour of the three CDs in water was investigated using a 30 ns trajectory generated from a Molecular Dynamics (MD) computer simulation.

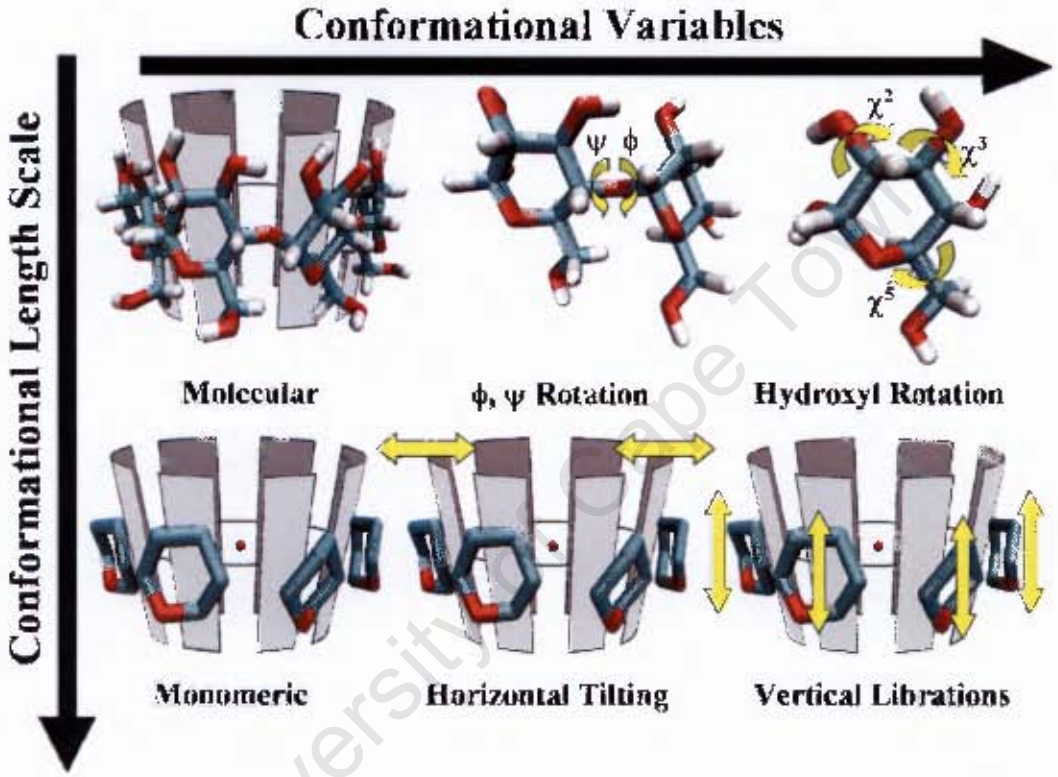

Figure 4.2: Conformational levels of abstraction. The top row shows the primary molecular rotors that affect the CD conformation. The bottom row shows the fundamental monomeric motion that most influences the CD macrocycle flexibility.

4.2 Computer Simulations

Atom names follow the convention for glucose, prefixed by the residue number. The torsion angles are defined by ϕ_{xy}: *x*H1-*x*C1-*x*O1-*y*C4 and ψ_{xy}: *x*C1-*x*O1-*y*C4-*y*H4 for the α(1-4)-linkages where x and y identify the atoms on adjacent glucose monomers respectively.

We follow the same computational procedure as previously reported [8]; consequently here we only summarise the main features of the MD simulations.

The program CHARMM was used for all the molecular mechanics computations [9] and a Carbohydrate Solution Force Field (CSFF) [10] was used to model the CDs. The SPCE water model [11] implemented with periodic boundary conditions was used to describe water interactions. The MD simulations were run for 500 ps to equilibrate the system and extended to 30 ns for data collection using a 1 fs integration time step with the bonds involving hydrogens fixed via the restraint algorithm SHAKE [12]. All solution simulations were run using an isothermal-isobaric ensemble (NPT) with P = 1 bar and T = 300 K. Configurations of the molecules were stored at intervals of 0.5 ps in all simulations and analysed over the entire 30 ns trajectory.

4.3 Results and Discussion

In this paper we assess the connection between the molecular and monomer degrees of freedom, illustrated in Figure 4.2, and the flexibilities of α-, β- and γ-CD. The principal molecular rotors affecting the internal motion of a cyclic system such as the CDs in an aqueous media are rotations about glycosidic bonds linking the glucose monomeric units and the rotation of the exo-cyclic hydroxyl groups. These motions determine the relative orientations of the glucopyranose monomer units. Since the exo-cyclic groups found on each CD are identical, the three CDs should exhibit analogous characteristics. Therefore, the drastic contrasts between the RMS fit amongst them, as shown by us previously [8], should be owing to their differences in conformational flexibility.

4.3.1 Glycosidic Linkage Analysis

In linear saccharides the major degrees of freedom are related to bond-rotations at the glycosidic torsion angles ϕ and ψ defined as H1'-C1'-O4-C4 and C1'-O4-C4-H4 that report the flexibility about the two bonds (C1'-O4 and O4-C4) connecting the component glucose monomers. These rotations are however significantly restricted in macrocyclic compounds such as the cyclodextrins. A powerful NMR technique for investigation of molecular conformations is the measurement of scalar spin-spin $^3J_{CH}$ couplings. The couplings related to the ϕ torsion angle, J_ϕ, are identical for the three molecules, whereas J_ψ values increase as a function of the macrocyclic size.

In contrast to the NMR experiments, computer simulations provide direct information about the average molecular conformations, i.e. ϕ and ψ torsion angles.

From the NMR coupling constants the CDs can be characterized by a single conformational state and no major conformational difference between the molecules can be observed. Both the averages and the standard deviations are similar for all three CDs [13]. The CD torsion angles can be translated to scalar $^3J_{CH}$ coupling constants using a Karplus-type relationship [14]

$$^3J_{COCH} = 7.49\cos^2\theta - 0.96\cos\theta + 0.15. \qquad (4.1)$$

where θ is ϕ or ψ. The couplings calculated from the trajectory are included in Table 4.1.

Molecule	Experimental		Simulation	
	J_ϕ	J_ψ	J_ϕ	J_ψ
α-CD	5.2	5.0	6.4	6.4
β-CD	5.2	5.2	6.6	6.6
γ-CD	5.2	5.4	6.6	6.6

Table 4.1: Long-range trans-glycosidic ^{1}H, ^{13}C coupling constants reported in Hz for the cyclodextrins where experimental values are taken from reference 13.

The agreement between the simulated and experimental values is satisfactory. In particular the trends for the three CDs from our computer simulations correlate with the NMR experimental results [13] indicating that the simulations are of sufficient length and the CSFF force field [10] is accurate enough to investigate the details of α-, β- and γ-CD conformational space.

We analyse the glycosidic linkages of the series of CDs in terms of the first primary molecular conformational variable (ϕ, ψ) as shown in Figure 4.2. The average ϕ and ψ torsion angles for all three CDs are close to zero degrees which significantly deviates from the average conformation of a maltose solution [15] (ϕ = -47.5°and ψ = -35.0°). However, the probability distribution $P(\phi,\psi)$ (Figs 4.3a-c) reveal much more about the conformational space explored by each CD and the extent of their relative conformational mobility.

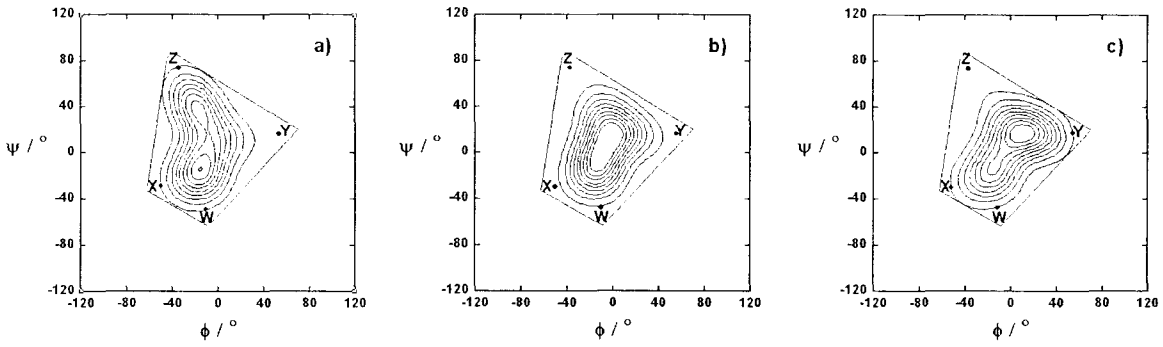

Figure 4.3: The probability density P(ϕ,ψ) for the CDs: (a) α, (b) β and (c) γ calculated for each dimer unit over the 30 ns simulation and normalised over number of dimers and frames. The probabilities are contoured at probability density intervals of 10 %, with the outer concentric indicating a 90 % probability density. The area bounded by the points W, X, Y and Z mark the maximum range of the probability density seen for γ-CD.

The entire $P(\phi, \psi)$ is shown in contours of 10% where the innermost concentric represents the smallest conformational deviation from the average ϕ and ψ torsion angles about zero degrees and is therefore the most probable population occurrence. It is immediately noticeable that β-CD samples a region of ϕ, ψ space that is centered about the (-8°, 4°) most probable conformation (Figure 4.3 b) compared with the two most probable conformations (-15°, -14°) and (-17°, 37°) for α-CD and (-16°, -13°) and (9°, 17°) for γ-CD (Figs 4.3 b and c respectively). These compare reasonably well with the average ϕ, ψ values observed in crystal structures of (-10.2, 7.6) for β-CD, (-10.8°, 8.8°) for α-CD and (-11.1°, 7.1) for γ-CD [16]. In Figure 4.3b, we note that the β-CD distribution exhibits only one minimum, whereas the other two (Figures 4.3a, c) indicate two states. This is in fact in agreement with the experimental distribution function derived from NMR parameters (NOEs, J- and dipolar couplings) for the α-CD [13].

This difference can be better contextualized when we describe the results in terms of a fundamental repeat unit of the CDs namely the disaccharide maltose (see Figure 4.2). The glycosidic linkage free energy landscape is a primary metric used to describe the conformational preferences of oligosaccharides. Previously Kuttel et al. calculated the free energy/potential of mean force (PMF) for maltose as a function of ϕ and ψ torsion angles [15]

$$W(\phi,\psi) = -kT \ln P(\phi,\psi). \tag{4.2}$$

The PMF surface for maltose in explicit aqueous solution was calculated using an iterative adaptive umbrella sampling method [17], as implemented by Naidoo et al. in CHARMM [18], combined with a weighted histogram analysis method [10, 19]. Here we show a section of the (ϕ, ψ - free energy surface, contoured at 2 kcal mol^{-1} (Figure 4.4), that is relevant to CD solution dynamics at ambient temperatures.

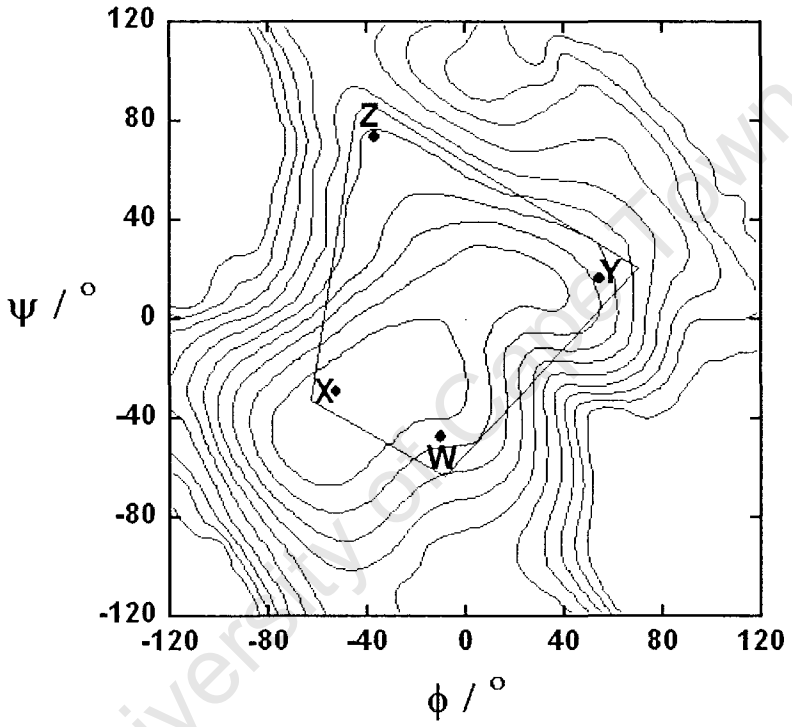

Figure 4.4: The maltose glycosidic linkage Potential of Mean Force as a function of the ϕ, ψ dihedral angles. The minimum is 0 kcal /mol and the maximum is 20 kcal/mol, and the plot is contoured at 2 kcal/mol. The area bounded by the points W, X, Y and Z mark the maximum range of the glycosidic ϕ, ψ torsion angles for γ-CD that is shown in Figure 3 (c). The free energy values are W = 4.5 kcal/mol, X = 0.4 kcal/mol, Y = 6.0 kcal/mol and Z = 9.2 kcal/mol.

We mark on the surface the furthest sampled points (W, X, Y and Z) seen in α- and γ-CD as they sample the largest ϕ, ψ surface. The β-CD samples a band of *syn* conformations along the X(0.4 kcal mol^{-1})↔Y(6.0 kcal mol^{-1}) leg and remains well within an energetic envelope of 6 kcal mol^{-1} about the global minimum for maltose. The glycosidic rotations in α-CD explores a similar area along the X↔Y line but in addition it expands along the Y(6.0 kcal mol^{-1})↔Z(9.2 kcal mol^{-1}) path almost reaching the conformational transition state conformation leading to the *anti*

conformation. The γ-CD also has low conformational strain on its component dimer glycosidic linkages since it samples conformations well within a 7 kcal mol^{-1} energetic envelope about the maltose Free Energy global minimum.

Notably, the variation in the highest probability distributions (inner contours) for the three CDs indicate that there is significantly more deviation in the ψ than the ϕ torsion angles. Although the CD molecules are strained, the macrocyclic restriction on the component maltose dimers still allows them to fluctuate on the periphery of their primary equilibrium energy well. The time scale for these fluctuations are defined by the time correlation functions (TCFs) calculated for the two torsion angles ψ and ϕ, and displayed in Figure 4.5.

All the TCFs exhibit very similar behaviour and the characteristic correlation times for these dynamical processes are of the order of 0.1-0.2 ns. We observe however that $C_\phi(t)$ decay in the order $\tau_\gamma > \tau_\alpha = \tau_\beta$ and the corresponding order for $C_\psi(t)$ is $\tau_\gamma > \tau_\alpha > \tau_\beta$. This could be the most important factor contributing to the relative internal dynamics of β-CD when compared to α- and γ-CD. The motion and distribution of the ϕ and ψ torsions therefore indicate that the relative flexibility of the CDs originates mostly from the ψ torsion angle and provides a partial explanation of the rms TCF trends describing the relative rigidity of the CD structures observed in our previous work [8].

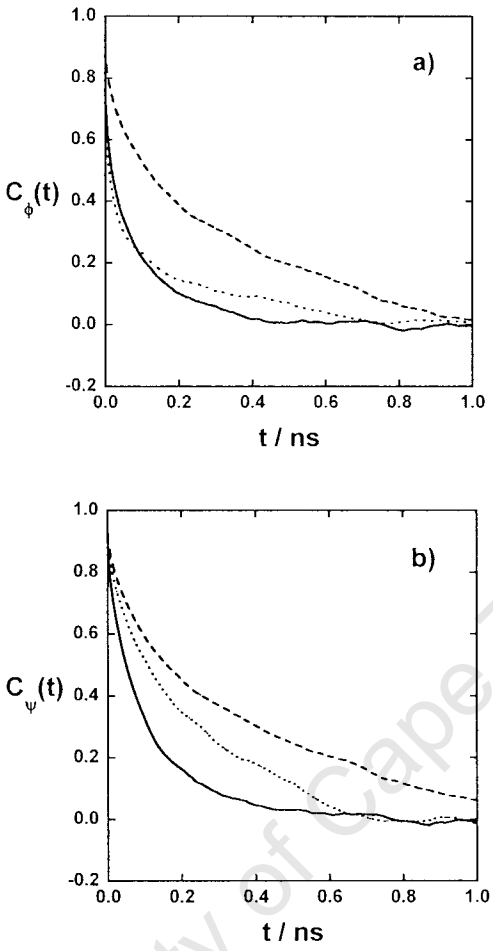

Figure 4.5: The time correlation function for the a) ϕ and b) ψ dihedral angles, where (·····), (——) and (----) denote α-, β- and γ-CD respectively.

4.3.2 Hydroxyl Rotation

Another parameter that strongly affects conformational transitions is the intra molecular hydrogen bond. For this we can investigate the relative orientations of the hydroxyl groups at C2, C3 and C6. The average O6---O6′ distances between adjacent primary alcohol oxygens observed in the 30 ns MD simulations for each CD are around 6 Å. This is beyond the Donor (D) --- Acceptor (A) limit of 3.5 Å required for intra-molecular hydrogen bonding.

However, we find that the hydroxyl groups at secondary carbon atoms present on the wider rim of the CD torus have average O3---O2′ distances that are well within the 3.5 Å D---A limit. Previously, results from chemical shift, coupling constant and temperature coefficient measurements have been used to establish that the strong presence of intra-molecular hydrogen bonding between OH-3 and OH-2′ exists [2]. In

these calculations we used the geometric criterion to define the hydrogen bond between a hydrogen bond D-H----A configuration for the OH(3)----OH(2) hydroxyl pair across the glycosidic linkage between two adjacent residues. Here, the D--A distance is 3.5 Å and the D-H---A angle is between 120° and 180°. We classified hydrogen bonds into types *A*, *B* and *C* as was done previously [2]. Type *A* hydrogen bonds have HO-2 as the hydrogen donor and O-3 as the acceptor, type *B* has HO-3 as the hydrogen donor and O-2 as the acceptor while type *C* ("flip-flop-type") hydrogen bonding is a rapid exchange hydrogen bond, where the hydrogen bond is flipping from type *A* to type *B* or vice versa. The type *A* and *B* hydrogen bonds are illustrated in Figure 4.6a along with numbers of times that these inter-residue hydrogen bonds occur in each of the CDs over the 30ns simulations. The β-CD exhibits the most frequent hydrogen bonding. In Figs 4.6b,c we show the frequency of hydrogen bonding between each pair of these hydroxyl groups for the three CDs. The hydrogen bond time series of both types are shown for the entire length of 30 ns trajectory in Figure 4.6b. However, because of the length of the trajectory the frequency of hydrogen bonds are obscured by the volume of the data. Therefore, we plot a segment of the hydrogen bond time series (Figure 4.6c) to illustrate typical inter residue hydrogen bonding patterns for the three CDs. From these time series plots (Figs 4.6b, c) and the distributions of hydrogen bonding (Figure 4.6a) it is evident that the type *B* hydrogen bonding is most dominant, which is in agreement with the chemical shift, coupling constant and temperature coefficient measurements in DMSO of Schneider et al [2].

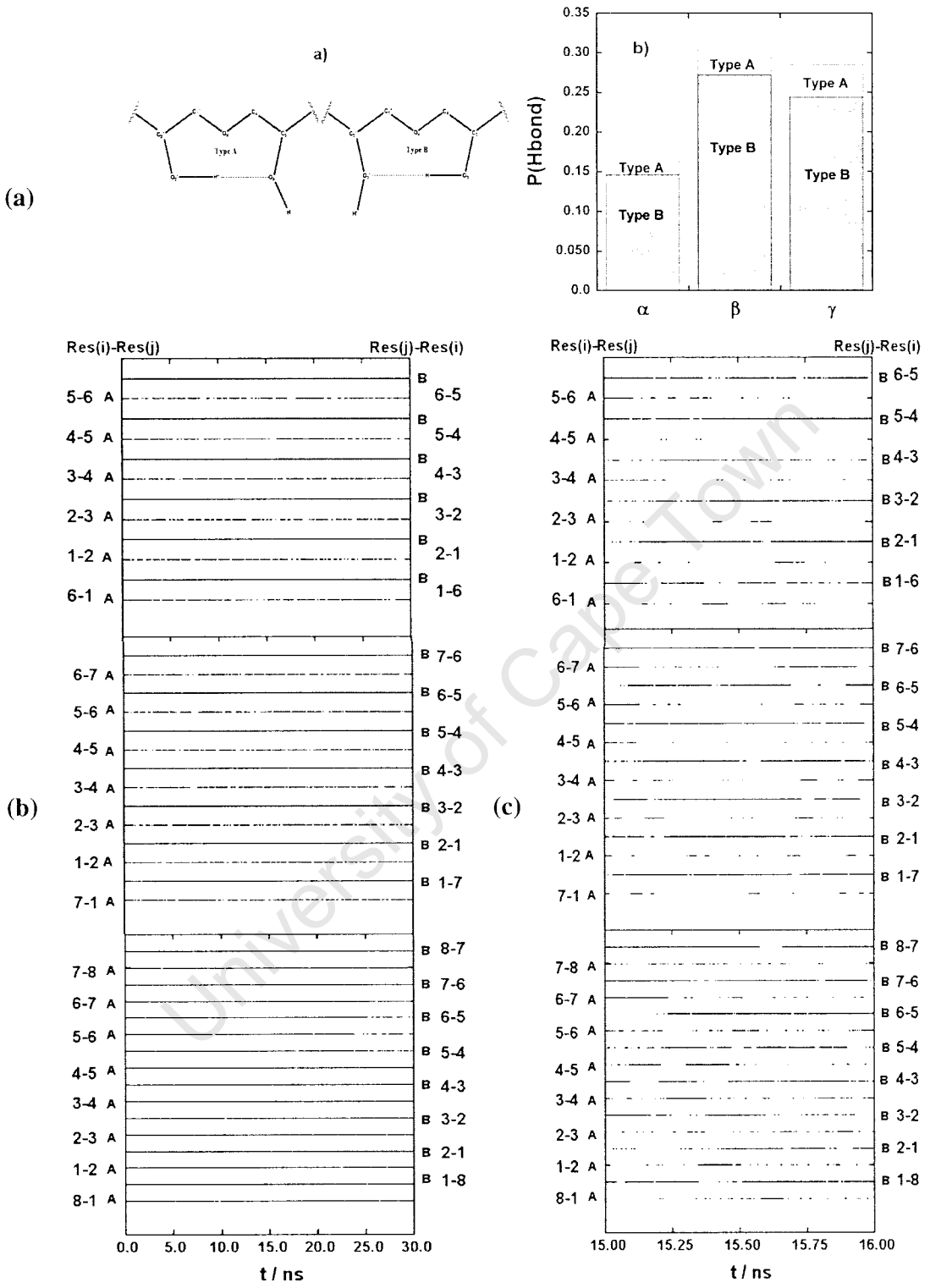

Figure 4.6: Inter-residue hydrogen bonding (OH(2)---OH(3)) (a) definition of hydrogen bonding types (type A and B) and the hydrogen bond distribution plot, (b) the time series of both types of hydrogen bonding for each CD over 30 ns and (c) an expansion of the time series between 15 and 16 ns.

The low aqueous solubility of β-CD was studied by Gillet *et al.* [20] and considered to be a result of a belt of hydrogen bonds occurring in the low rim of CD. They argued that since one of the glucopyranose units preferred to tilt in a distorted position, the hydrogen bond belt is broken. They reasoned further that under such circumstance only four out of all possible six hydrogen bonds can be established for α-CD and therefore concluded that such phenomenon in α-CD enhanced its solubility. These conclusions were drawn from solid state (static conformations) and experiments carried out in DMSO and do not provide direct evidence for the fundamental chemical and physical character leading to the anomalous solubility of β-CD in water. While β-CD makes more of these hydrogen bonds compared with α-CD and γ-CD (Figure 4.6a) the hydrogen bond frequency (Figure 4.6c) is not the same for each residue pair for γ-CD and so does not form a belt of hydrogen bonds on the wide rim. Whether these hydrogen bonds are the reason for the unusual solubility trend in the α-, β- and γ-CD series remains unsubstantiated. It has been shown for α(1-4)-linkages that the intra-molecular hydrogen bonding between these hydroxyl groups has a magnitude of -4.99 kcal/mol [21] while the competing inter-molecular hydrogen bonds that form between the hydroxyl groups at C2' and C3 and individual water molecules are at most -4.50 kcal/mol and -4.46 kcal/mol in strength respectively [22]. The difference in strength between the intra-molecular hydrogen bonds across the glycosidic linkages that would induce conformational rigidity in the CDs and the inter-molecular hydrogen bonds made with the waters that would promote conformational flexibility in solution is less than 0.5 kcal/mol. In a recent study [23] the ^{1}H spin-spin relaxation rate constant, R_2, of water was measured in aqueous solutions of α-, β- and γ-CDs to determine the relative difference between the CDs, which may arise from the degree to which the three CDs experience intra-molecular hydrogen bonding. Sabandini et. al. found that the rates that the water protons exchanged with the CD protons were the same in all three cases. Therefore water has similar access to these hydroxyl groups in each of the CDs indicating no significant difference in the extent of their intra-molecular and intermolecular hydrogen bonding [23]. This implies that it is unlikely that intra-molecular hydrogen bonding is the reason for the irregular solubility trend observed in cyclodextrins. A more likely reason may originate from the balance between the tension arising from the macrocyclic ring in β-CD and the deviation of the glycosidic conformations of its component dimers from the preferred *syn* dimer free energy well.

4.3.3 Glucose Monomer Tilting

The relative position and orientation of the glucose monomer plane with respect to the macrocyclic plane is the second metric that reveals the range conformations possible in the CDs. The centre of the monomer plane was calculated using atoms C1, C2, C4 and C5 and the centre of the CD macrocyclic ring was defined using the monomer centre of masses. We then calculated a vector between the centre of the CD and the monomer centre and a second vector from the monomer centre to the centre of the C5-O5 bond. We defined the monomer tilt angle (θ) in Figure 4.7a as the angle between these the two normalised vectors. A positive angle is obtained if the side of the glucose monomer, containing the O2 and O3 hydroxyls, tilts away from the cyclodextrin macrocyclic cavity and a negative value if it tilts in toward the cavity. The glucopyranose monomer unit tilt angle distributions, $P(\theta)$, for α-, β- and γ-CD are shown in Figure 4.7b, with averages of 13°±12°, 13°±13° and 9°±17°, respectively. The monomers for all three CDs appear to tilt both inward (toward) and outward (away) from the macrocyclic ring cavity. The cone shape where the O2------O3 lined rim has the larger circumference (more positive θ) compared with the O6 lined rim is more probable for β-CD compared with the other two CDs. In the case of α-CD the glucoses show a conformational preference to be oriented upright or very slightly tilted toward the cone opening. The tilting range of glucoses in the γ-CD case is greater than α- and β-CD and consequently a large variation in its CD conical shape can be expected. A comparison of these values to the observations made in the solid state where the average tilt angles of the glucopyranose monomer units for α-, β- and γ-CD are 13°±10°, 14°±10° and 19°±9° respectively [24] reveals a great similarity for α- and β-CD while there is a large difference for γ-CD, for which a very small number of crystal structures exists. However, the range of this motion and the average position does not significantly differ between the CDs.

While the $P(\theta)$, for the in-and-out of plane tilting reveals much about the relative shape of β-CD the associated dynamics of the glucose monomer tilting informs us of the relative flexibility of the CDs with respect to this shape. We calculate time correlation functions $C_\theta(t)$ averaged over the total number (n) of the tilt angles in the CD (Figure 4.7c). This tilt angle time correlation function is defined as:

$$C_\theta(t) = \frac{1}{n}\sum_i^n \left(\frac{\langle \theta_i(0)\theta_i(t) \rangle}{\langle \theta_i(0)^2 \rangle} \right) \tag{4.3}$$

where $\theta_i(t)$ is the tilt angle for glucose ring i at time interval t in the CD.

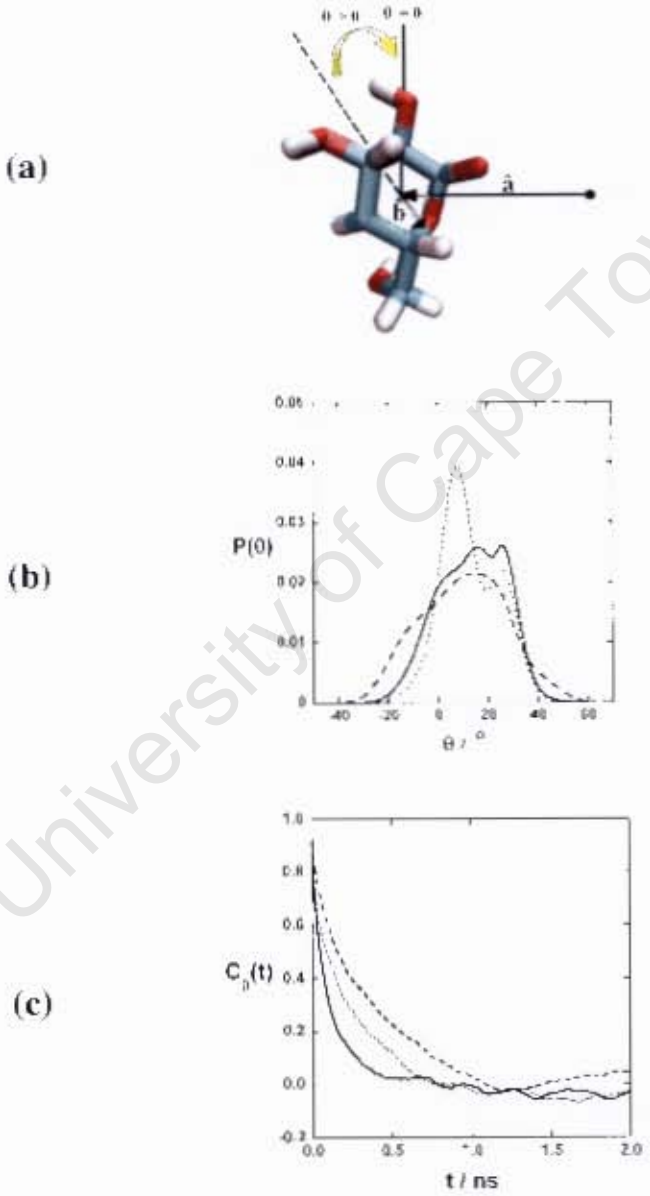

Figure 4.7: (a) Monomer Tilt angle θ (a) defined as the angle between the mean macrocyclic plane and a plane through the glucose monomer, $\theta = (\hat{a} \bullet \hat{b}) - 90$. The θ angle (b) distributions and (c) time correlation functions are shown, where (·····), (——) and (----) denote α-, β- and γ CD respectively.

Fitting the TCFs to single exponentials we obtain correlation times of 0.2, 0.1 and 0.4 ns for α-, β- and γ-CD. The relaxation time of β-CD for the glucose tilting motion is therefore shorter than that of α- and γ-CD, which is in agreement with the argument that it is conformationally the least flexible. Interestingly these TCFs decay in the same order for the three CDs as they do for the $C_{\psi}(t)$. We infer from this that the relative ease of rotation about the ψ torsion angle is the fundamental molecular cause for the differences in the CD glucose monomers tilting toward and away from their macrocyclic cones.

4.3.4 Glucose Monomer Vertical Librations

A general approach to investigate the internal motion of the cyclodextrins in solution is to calculate a TCF, say $C_{rms}(t) = \langle rms(0,t) \rangle$. Here, *rms(t1,t2)* denotes the best Root-Mean-Square fit obtained between structures at times t_1 and t_2 using a standard fitting procedure. Prior to calculation of the *rms* fit, the orientational and translational dynamics of the CDs should be removed. These TCFs were calculated for all three CDs in our previous communication [8].

Initially a perfect correlation between the molecular structures is observed, and therefore $C_{rms}(0) = \langle rms(0,0) \rangle = 0$. In our interpretation [8], the long time limit of the TCF corresponds to a generalised amplitude of the internal motion, whereas the time to reach the plateau reflects the relaxation process of the macrocyclic ring toward the equilibrium structure. The rms fit TCF for the β-CD indicates the most restricted macrocyclic ring motion and relaxes significantly faster than the other CDs. In that study we observed rms fits for each of macromolecule in its entirety that implied an inflexibility pattern, which follows the solubility trend for the three CDs.

We now consider trajectories of the glucose monomer centres of masses in relation to the mean macrocyclic ring plane in the laboratory frame (i.e., removing translation and orientation motion). Every structure 0.05 ps apart along the trajectory, for the full 30 ns, were superimposed onto a macrocyclic ring reference plane. This was constructed using the macrocyclic geometric centre (red dot in Figure 4.8a) and all the glucose centers of mass. The vertical displacement (δ) of the centre of mass of each glucose monomer in the CD ring relative to the mean macrocyclic ring plane is illustrated in Figure 4.8a. The δ for each glucose monomer was then calculated for the

whole trajectory. A positive value was assigned if a glucose monomer was found to have a positive z-coordinate and likewise a negative value for the opposite case.

We then calculated the Root Mean Square Deviation monomer displacement (δ) time correlation function $C_{rmsd(i)}(t)$ averaged over the total number (n) of monomer vertical librations in each CD. This is defined as:

$$C_{rmsd\delta}(t) = \frac{1}{n}\sum_{i}^{n}\left(\frac{\langle \delta_{rmsd(i)}(0)\delta_{rmsd(i)}(t)\rangle}{\langle \delta_{rmsd(i)}(0)^2\rangle}\right) \tag{4.4}$$

where $\delta_{rmsd(i)}(t)$ is the Root Mean Square Deviation of the center of mass displacement for glucose ring i at time interval t in the CD.

The RMSD TCFs for these vertical librations, $C_{rmsd(i)}(t)$, are displayed in Figure 4.8b. These TCFs approximate an exponential functional form where at the start there is perfect correlation between the molecular structures with $C_{rmsd}(0,0)=0$. After this the difference between the structures, increase and reach a plateau in the long time limit. Two types of information can be derived from the RMSD TCFs: *i*) a value of $C_{rmsd(i)}(t)$ corresponding to the plateau, and *ii*) a characteristic (relaxation) time required to reach the plateau. The value of the displacement read from the plateau can be interpreted as a generalised amplitude of the monomer motion, whereas the time to reach the plateau reflects the relaxation process of the glucose monomer unit toward the average distance from the mean macrocyclic ring plane. The average levels of the plateau are 2.2, 1.6 and 2.4 Å for α-, β- and γ-CDs, while the estimated relaxation times are 1.4, 1.0 and 1.7 ns for α-, β- and γ-CDs, respectively. We see from this that β-CD undergoes the most restricted macrocyclic ring motion and relaxes significantly faster than the other CDs. Furthermore there are much less fluctuations in $C_{rmsd(i)}(t)$ compared with α- and γ-CD. This leads us to conclude that the monomers in β-CD undergo small amplitude fast vertical librations when reaching their librational equilibrium.

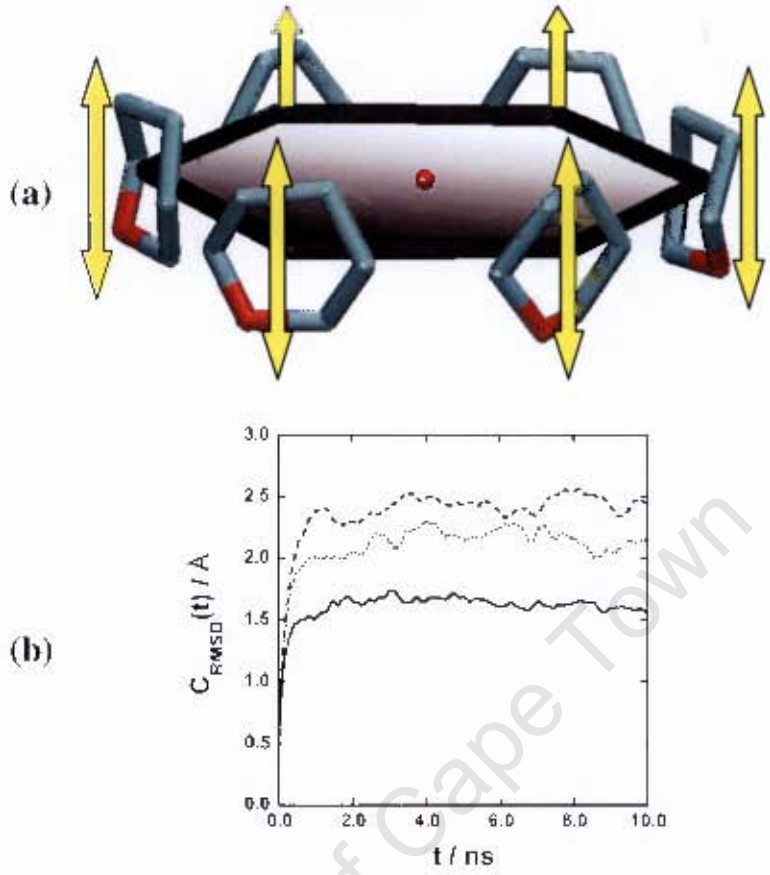

Figure 4.8: (a) Monomer vertical librations defined and (''), (——) and (----) denote α-, β- and γ-CD, respectively, for the (b) for the averaged (over all the glucose centres) RMSD time correlation functions.

4.4 Conclusions

We have studied aqueous solutions of α-, β-, and γ-CDs using MD computer simulations. The agreement between experimental and calculated spin-spin couplings is satisfactory and we therefore conclude that the conformational behaviour of the CDs is adequately described in our computer simulations. The general structural behaviour, monitored by the time correlation function, $C_{rmsd}(t)$, indicated that β-CD is more rigid than the other CDs. A possible molecular origin for this observation is the differences in their glycosidic conformational space $P(\phi,\psi)$, which was evaluated from the trajectory. The conformational space is a bit more limited in the β-CD case compared with the other two CDs.

The ranges of the angular tilting (θ) of the glucose monomers toward and away from the macrocyclic centre are relatively similar for the three CDs. However, the time taken to relax into an equilibrium state follows the pattern where β-CD relaxes

fastest followed by α-CD and then γ-CD. Similarly the generalised amplitude of the monomer motion relative to the mean macrocyclic plane have magnitudes in the order γ-CD > α-CD > β-CD. These monomer librations relax fastest for β-CD followed by α-CD with γ-CD being the slowest to reach the mean displacement from the macrocyclic plane. This finding confirms that at the monomeric length scale β-CD is the least flexible of the three CDs.

We indicated in Figure 4.2 that the most influential molecular parameters that could affect oligosaccharide conformations are the torsion rotations about the glycosidic bonds and the orientations of the hydroxyl groups that may form intermolecular hydrogen bonds. In this study we find that the molecular parameters that track the solubility trend are the distribution of the ψ torsions and the associated relaxation times with it. This leads us to conclude that the range and rate of rotation of the ψ torsions in the CDs are the basis of their glucose monomer units' orientational and librational motion and dynamics. This in turn underlies the order of macrocyclic inflexibility.

We have investigated 30 ns solution simulations of α-, β-, and γ-CDs, which resulted in a detailed comparative description of their molecular and monomeric conformations. This provides a description of the solution conformational behavior of these three macrocyclic compounds that are consistent with their relative solubility. We speculate that the reason for the rigidity of the β-CD is simply due to the inability of the uneven number of constituent maltose dimers making up the β-CD ring, to traverse the favoured *syn* free energy ϕ,ψ well as easily as even numbered α- and γ-CDs. This difference is most apparent for transitions that require glycosidic linkage rotations about the ψ torsion angle.

References

1. H. Dodziuk. *Cyclodextrins and Their Complexes: Chemistry, Analytical Methods, Applications.*; Wiley: London, 2006; Lipkowitz, K. B. *Chem. Rev* **1998**, *98*, 1829.
2. H. J. Schneider, F. Hacket, V. Rudiger, H. Ikeda. *Chem. Rev* **1998**, *98*, 1755.
3. D. French. *Adv. Carbohydr. Chem.* **1957**, *12*, 189.
4. M. J. Jozwiakowski, K. A. Connors. *Carbohydr. Res* **1985**, *143*, 51; E. B. Starikov, K. Brasicke, E. W. Knapp, W. Saenger. *Chem. Phys. Lett.* **2001**, *336*, 504.
5. W. Linert, P. Marg, F. Renz. *Chem. Phys.* **1992**, *161*, 327; F. A. Momany, J. L. Willet. *Carbohydr. Res.* **2000**, *326*, 210; J. Varady, X. W. Wu, S. M. Wang. *J. Phys. Chem. B.* **2002**, *106*, 4863; N. Funasaki, S. Ishikawa, S. Neya. *J. Phys. Chem. B.* **2002**, *106*, 6431; T. Heine, H. F. Dos Santos, S. Patchkovskii, H. A. Duarte. *J. Phys. Chem. A* **2007**, *111*, 5648; W. Snor, E. Liedl, P. Weiss-Greiler, A. Karpfen, H. Viernstein, P. Wolschann. *Chem. Phys. Lett.* **2007**, *441*, 159; A. Karpfen, E. Liedl, W. Snor, H. Viernstein, P. Weiss-Greiler, P. Wolschann. *Monatshefte Fur Chemie* **2008**, *139*, 363.
6. J. E. H. Koehler, W. Saenger, W. F. van Gunsteren. *J. Mol. Biol.* **1988**, *203*, 241; J. E. H. Koehler, W. Saenger, W. F. van Gunsteren. *Euro. Biophys. J.* **1988**, *16*, 153; J. E. H. Koehler, W. Saenger, W. F. van Gunsteren. *Euro. Biophys. J.* **1987**, *15*, 197; J. E. H. Koehler, W. Saenger, W. F. van Gunsteren. *Euro. Biophys. J.* **1987**, *15*, 211.
7. P. Bonnet, C. Jaime, L. Morin-Allory. *J. Org. Chem.* **2002**, *67*, 8602; C. S. Pereira, A. E. de Moura, L. C. G. Freitas, R. A. Lins. *J. Braz. Chem. Soc.* **2007**, *18*, 951.
8. K. J. Naidoo, Y. -J. Chen, J. L. M. Jansson, G. Widmalm, A. Maliniak. *J. Phys. Chem. B* **2004**, *108*, 4236.
9. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus. *J. Comput. Chem.* **1983**, *4(2)*, 187.
10. M. M. Kuttel, J. W. Brady, K. J. Naidoo. *J. Comput. Chem.* **2002**, *23*, 1236.
11. H. J. C. Berendsen, R. J. Grigera, T. P. Straatsma. *J. Phys. Chem.* **1987**, *91*, 6269.
12. W. F. van Gunsteren, H. J. C. Berendsen. *Mol. Phys.* **1977**, *34*, 1311.

13. J. Thaning, B. Stevensson, J. Östervall, K. J. Naidoo, G. Widmalm, A. Maliniak. *J. Phys. Chem.* **2008**, *in press*.
14. F. Cloran, I. Carmichael, A. S. Serianni. *J. Am. Chem. Soc.* **1999**, *121*, 9843.
15. M. M. Kuttel, K. J. Naidoo. *J. Phys. Chem. B.* **2005**, *109*, 7468.
16. W. R. Saenger, J. Jacob, K. Gessler, T. Steiner, D. Hoffmann, H. Sanbe, K. Koizumi, S. M. Smith, T. Takaha. *Chem. Rev.* **1998**, *98*, 1787.
17. R. K. Schmidt, J. W. Brady, B. Teo. *J. Phys. Chem.* **1995**, *99*, 11339; C. Bartels, M. Karplus. *J. Comput. Chem.* **1997**, *18*, 1450.
18. K. J. Naidoo, J. W. Brady. *J. Am. Chem. Soc.* **1999**, *121*, 2244.
19. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg. *J. Comput. Chem.* **1992**, *13*, 1011.
20. B. Gillet, D. J. Nicole, J. J. Delpueche. *Tetrahedron letters* **1982**, *23*, 65.
21. Y. -J. Chen, K. J. Naidoo. *J. Phys. Chem. B* **2003**, *107*, 9558.
22. K. J. Naidoo, Y. -J. Chen. *Mol. Phys.* **2003**, *101*, 2687.
23. E. Sabadini, F. D. Egidio, F. Y. Fujiwara, T. Cosgrove. *J. Phys. Chem. B.* **2008**, *112*, 3328.
24. K. Harata. *Chem. Rev.* **1998**, *98*, 1803.

Chapter 5

Relative Aqueous Solubility of α-, β- and γ-Cyclodextrins

5.1 Introduction

In the previous chapter we showed that the macrocyclic ring conformations and dynamical motions were more restricted for β-CD than the other two CDs. In this chapter we examine the solution properties of the CDs in solution; we employ both experimental and computational methods to complete this task. We use ultrasonic interferometry to ascertain the hydration number, i.e. the number of waters in the first solvation shell, and the computational studies afforded the solvent structuring, around the CDs, residence time of the waters in the first solvation shell and energetics of the solvation layer.

Since most applications of CDs have been developed in aqueous media, the investigation and understanding of the hydration structure and dynamics of CDs in aqueous media is considerably important for developing new, more advanced CD applications [1]. However, there have only been a few reports so far on the hydration number and dynamics of water molecules associated with CDs [2]. There are many definitions for waters of "hydration", which depend on the probing technique and the energetic or structural criteria applied [3]. Spectroscopy (including NMR, IR and other techniques) is available to determine the hydration number, but they do not provide accurate and precise hydration numbers of solute molecules dissolved in dilute solutions in general [4].

Among the experimental methods very well suited for investigating structure of liquids and hydration number, ultrasonic interferometry is one of most promising and has showed extensive possibilities in the field of solution chemistry, particularly for the study of structural properties of carbohydrates such as their compressibility and volumetric parameters [5]. It is also possible to estimate the hydration numbers of biopolymers from their compressibility values at infinite dilution. Researchers have done extensive studies on carbohydrates, including CDs and their derivatives, to

determine the hydration behaviour in various aqueous and non-aqueous environments at different temperatures covering a wide range of compositions – by employing ultrasonic speed and density measurement [6].

Hydration number

The assumption of the equation 5.2 (section 5.3.1, page 106), derived by Pasynski, is very simple: hydration number means that after the addition of a solute there are molecules of water which become non-compressible because they are now bonded to the guest molecule; remember that bonding means, to chemists, a shortening of the distance between interacting moieties and a rigid link between them [7]. Thus, a number of water molecules become excluded from the total compressibility of the solution. Originally, Pasynski applied this idea to electrolytic solutions, supposing that water molecules undergo electrostrictive compression in electrostatic field of ions, which acts like very high static pressure and makes them less compressible (or even non-compressible at all). Is it reasonable to treat diluted aqueous solutions of non-electrolytes different than those of electrolytes? The interactions between water and the solute leads, in both cases, to the same final result: independent on the fact if electrostriction caused by electrostatic forces occurs or if hydrogen or other bonds are created, the local structure should become enhanced and more rigid, always leading to lowered compressibility.

The hydration number for a series of dilute non-electrolyte organic compounds was studied by Burakowski and Glinski [7] in aqueous solution using ultrasonic intereferometry. They tested the Pasynski method of determination of hydration numbers for non-electrolytic aqueous solutions. It was shown for dilute aqueous solutions of few homologous series of organic non-electrolytes that the hydration numbers are strongly and directly depending on the length of the solute molecules and the groups which are their constituents. The hydration number versus the number of carbon atoms in the alkyl chains of the solutes for the homologous series were investigated, Table 5.1. Note that the dependencies are linear or close to linear, at least in the limitations of the experimental error, Figure 5.1.

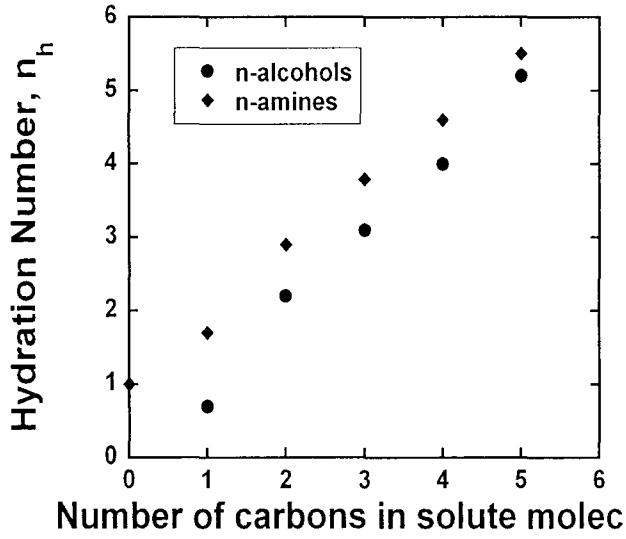

Figure 5.1: The hydration number of non-electrolyte organic, alcohol and amine, compounds as the aliphatic chain is increased.

The data obtained by Burakowski and Glinsk suggested that the addition of a –CH2– group to the chain increases the total hydration number of the molecule by approximately one. Thus, the results suggest that hydration is an additive parameter although its mechanisms are different.

Solute (name)	Solute (formula)	Hydration Number, n_h
Methanol	CH_3-OH	0.7
Ethanol	CH_3-CH_2-OH	2.2
n-Propanol	CH_3-CH_2-CH_2-OH	3.1
n-Butanol	CH_3-CH_2-CH_2-CH_2-OH	4.0
n-Pentanol	CH_3-CH_2-CH_2-CH_2-CH_2-OH	5.2
Ammonia	NH_3	1.0
Methylamine	CH_3-NH_2	1.7
Ethylamine	CH_3-CH_2-NH_2	2.9
n-Propylamine	CH_3-CH_2-CH_2-NH_2	3.8
n-Butylamine	CH_3-CH_2-CH_2-CH_2-NH_2	4.6
n-Pentylamine	CH_3-CH_2-CH_2-CH_2-CH_2-NH_2	5.5

Table 5.1: The experimental results obtained at 298.15 K; the hydration number, calculated from equation 5.2, of alcohols and amines with increasing alkyl chain. Taken from Burakowski and Glinski [7].

When a non-electrolyte (for instance an organic non-electrolyte molecule) interacts with the water network, three mechanisms of hydration seem to be dominating: (i) formation of hydrogen bonds, (ii) compression of neighbouring water molecules by electrostatic forces if there are charged parts to the solute molecule and (iii) formation of cages around the hydrophobic parts of the molecules. The first mechanism only affects molecules that are engaged in the bonding, making them almost completely incompressible, and corresponds to hydrophilic hydration; it seems that such building into the original water network does not affect it at a longer distance. The second mechanism is supposed to be much weaker than that observed in the electrolytic solutions because of low charge density of the charged fragments for typical, polar non-electrolyte molecules. Finally, the last effect is hydrophobic hydration, which should increase the rigidity of the water network, again leading to a decrease of the total compressibility of the system [7].

5.2 Experimental and Computational Methods

Materials

The α-, β- and γ-CDs were purchased from Sigma Aldrich Inc. St Luis, 3050 Spruce Street, MO 63103 USA. All the substances, with purities higher than 99%, were used without further purification. All the solutions were freshly prepared with distilled and deionized water (taken from a Millipore Super-Q System, with a conductivity lower than 18 $\mu\Omega^{-1}cm^{-1}$). The CD solutions were prepared by weight, with the utmost care, and stored in special airtight bottles to avoid moisture contamination and evaporation. The weighings were done on an Afcoset-ER-120A electronic balance with a precision of ± 0.01 mg. The homogeneity of the initial solutions was assured by sonicating them for 2 hours in an ultrasonic bath. The same CD solutions, of various concentrations, were used for the density and compressibility measurements.

Instrumentation

The ultrasonic speeds through all of the CD solutions were measured using a single-crystal variable-path interferometer operating at 3 MHz by the method

described elsewhere [8]. The ultrasonic speed data were reproducible to within ± 0.05 m/s. Density measurements were carried out at 298.15K by using an Anton Paar digital Model DMA 34. The reliability of experimental measurements of u and ρ were ascertained by comparing the experimental data of the glucose with the corresponding reported values at the studied temperatures [9]. Our ultrasonic experiments for glucose was able to reproduce the results obtained by the authors of ref. 9, thus proving that our ultrasonic technique is reliable. The temperature of the test liquids during the measurements was maintained to an accuracy of ±0.02 K in an electronically controlled thermostatic water bath (JULABO, Model-MD, Germany).

Molecular Dynamics

We follow the same computational procedure as previously reported in chapter 4.

5.3 Results and Discussion

5.3.1 Ultrasonic Interferometry Results

A thermodynamic study of the CD solutions have been carried out at 298.15K, by measuring the speed of sound, u, and the density, ρ. Given that ultrasonic absorption is negligible under the experimental conditions used herein, the isentropic compressibility, κ_s, can be determined from speed of sound and density by the Newton–Laplace equation:

$$\kappa_s = \frac{1}{\rho u^2} \tag{5.1}$$

There can be found in the literature several methods to determine hydration numbers, i.e., from ionic mobilities, entropy values, apparent molar properties at infinite dilution, ultrasonic measurements, etc. The latter allows us to obtain the primary hydration numbers, n_h, through the expression

$$n_h = \frac{n_w}{n_s}\left(1 - \frac{k_s}{k_{s,o}}\right) \tag{5.2}$$

Where $\kappa_{s,o}$ and κ_s are the isentropic compressibility of pure water and experimental solution and, n_w and n_s, are the number of moles of water and CDs, respectively. Equation 5.2 implicitly assumes that: (i) n_h is the number of water molecules of the first layer surrounding the solute, and (ii) these water molecules are trapped so tightly that they can be considered as incompressible.

The values of ρ u, κ_s and n_h as functions of molality of CDs at 298.15K are reported in Table 5.2. The experimental values of ρ and u were used to calculate isentropic compressibility and hydration number. The thermoacoustic results reveal that ultrasonic velocity increases with an increase in the concentration of α-, β- and γ-CD solutions. The linear increase of velocity with concentration indicates weakening of solute-solvent interactions and dominating solute-solute interaction. It is observed that the values of ultrasonic speed increases, while those of isentropic compressibility, κ_s, decreases with molality of CDs for all three systems under study, Figure 5.2a. The observed isentropic compressibility values depend on several factors, which are of a physical and/or chemical nature [10].

CDs	Fit Coefficients	u / m.s^{-1}	ρ / g.cm^{-3}	k_s / Pa^{-1}	n_h
α-CD	a0	1496 (0.02)	0.998 (1.7)	447.7 (0.02)	16.49 (0.22)
	a1	167.9 (4.5)	0.062 (0.002)	-137.64 (3.55)	-43.45 (36.85)
β-CD	a0	1495.2 ± 0.01	0.9991 ± 1.86	447.69 ± 0.01	35.14 ± 0.30
	a1	409.89 ± 2.17	0.0921±0.003	-286.28 ± 1.90	37.53 ± 49.84
γ-CD	a0	1495.1 ± 0.005	0.9991 ± 1.4	447.71 ± 0.004	8.25 ± 0.09
	a1	46.39 ± 0.80	0.0916 ± 0.002	-68.91 ± 0.77	19.12 ± 15.55

Table 5.2: Fit coefficients of the variation *of* u,ρ, k_s and n_h as a function of the molarity of solutions, at 298.15K, for the CD systems.

Here, the increase in the density, ρ, does not imply a change in the hydration shell of the substrates, since the number of water molecules in this hydration shell, n_h, remains basically constant with the molality of the solution. From the experimental properties, ρ and u, it can be noted that the speed of sound is more sensitive than the density to both the conformation and arrangement of the hydroxyl groups of the glucopyranose units of the CDs. As expected isentropic compressibility, κ_s, decreases with an increase of molar concentration of CD solutions, suggesting strong solute-

solvent interactions, through dipole-dipole interactions of OH groups of the CDs with the surrounding water molecules. The following solute-solvent interaction trend is observed as γ-CD > α-CD > β-CD in our present thermoacoustic study. It is also interesting to note that isentropic compressibility values of the α- and γ- CD solutions are lower than in the β-CD solution. This clearly indicates substantial interaction between water with α- and γ-CD as compare to β-CD. From thermoacoustic experiment, the hydration number for the CDs (α-CD: 16.4, γ-CD: 8.3, β-CD: 35.1, at infinite dilution) clearly indicates that there are more water molecules around β-CD as compared to the other two CDs, Figure 5.2b.

It is worth considering two main aspects: (a) from the behavior of the properties studied herein, it can be concluded that an increase in sound of speed and density, involves a decrease in isentropic compressibility and an increase in hydration number; (b) the values of n_h, are directly related to the behavior of the solvation shell of the molecule in the aqueous solution and particularly the way in which this molecule fits into the three-dimensional hydrogen-bonded structure of the water solvent. The higher the hydration value, the lower the isentropic compressibility and the higher the perturbation on the water structure when introducing the solute molecule.

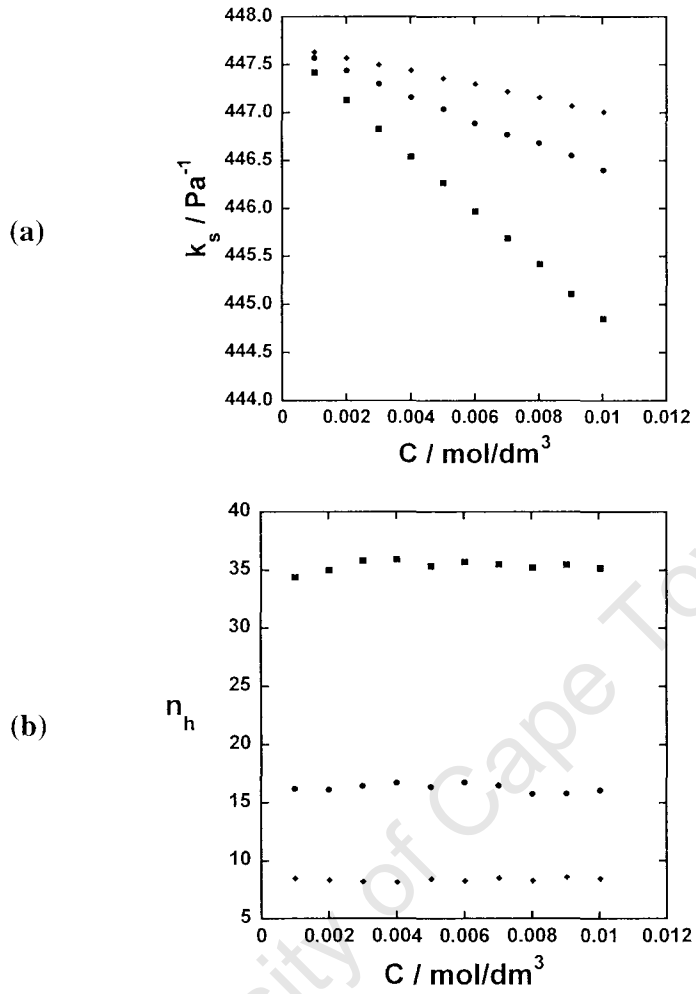

Figure 5.2: (a) Isentropic compressibility and (b) hydration number studied as a function of CD concentration. The circles, squares and diamonds denote α-, β- and γ-CD respectively.

5.3.2 Molecular Dynamics Results

Pair and Spatial Distribution Functions

The pair distribution function (PDF) shows the probability of finding a water molecule a certain distance away from the atom of interest in the solute molecule and is independent of orientation, while the spatial distribution function shows the spatial arrangement of the waters around the entire solute.

The function $g_{ij}(r)$ describes how the density of surrounding matter varies as a function of the distance from a distinguished point, known as the site-site PDF:

$$g_{ij} = \frac{1}{4\pi r_{ij}^{\,2}} \frac{dN(r_{ij})}{dr_{ij}} \tag{5.3}$$

The centre of mass (CM) of the CD was first determined using the centre of masses for each glucose unit in each CD. The pair distribution function was then calculated from this central site to the oxygen, of the waters, in the simulation. First thing to note is that the heights of the peaks are in the order: β-CD > α-CD > γ-CD. There are two definite peaks which appear in the PDF; indicating the probability of finding the waters in the inside cavity of the CDs and the first solvation shell outside of the CD. We found that the size of the cavity can be determined from the minimum of the first peak in PDF. The cavity diameters, from the first minima of the PDF, where found to be approximately 5.8 Å for α-CD, 6.2 Å for β-CD and 6.8 Å γ-CD, Figure 5.3, which is comparable to the literature values (Table 1.1, page 9).

The spatial distribution functions (or the water probability density) for the three CDs, are shown in Figure 5.3. The chosen contours represent the probability of finding water molecules around the CD that is 50% higher compared with that in the bulk solvent. Hence, the SDFs reflect the local water structuring around the CD and can be related to the tendency of the CDs to limit the configurational entropy of the water as compared to the bulk region. We note that the least soluble, β-CD, induces the strongest structuring of the surrounding waters. However, the most soluble γ-CD structures the surrounding water to a lesser extent compared to the two other CDs. The α-CD is an intermediate case. Thus we observe a strong correlation between the local water structuring and the solubility of the CDs.

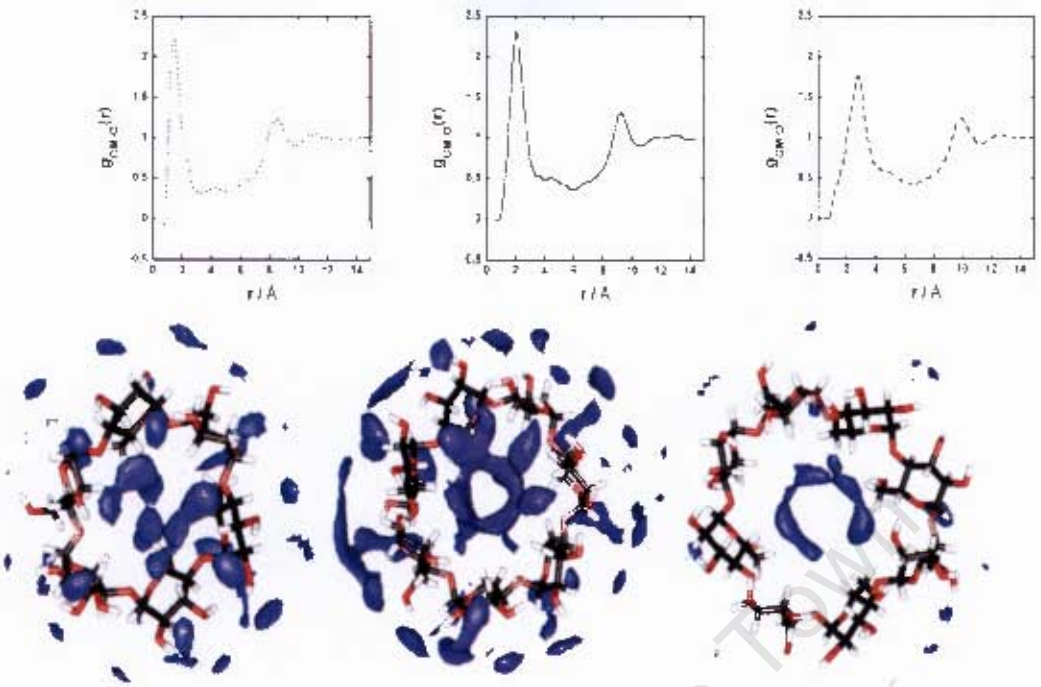

Figure 5.3: The pair (top) and spatial (bottom) distribution functions. (·····), (——) and (----) denote α-, β- and γ-CD, respectively. The pair distribution functions were calculated from the centre of masses of the CD.

Residence Time

The time that a water molecule spends in the first solvation shell of the solute before diffusing back out into the bulk solution is known as the residence time. The residence time is calculated from a number correlation function (equation 5.4) is similar to a time correlation function, except that the dynamical variable has a value of either 0 or 1. The correlation time, τ, calculated from the time correlation function is the average residence time that water molecule will spend in the first solvation shell. A spherical cutoff is used in the calculation; the cutoff radius extends slightly more than the distance of the first solvation shell from the centre of the solute molecule, which is obtained from the second peak in the PDF, Figure 5.3.

$$C_{res}(t) = \frac{1}{N_t}\sum_{n=1}^{N}\sum_{j} C_j(t_n, t; t^*) \tag{5.4}$$

To obtain a precise definition for the residence time, we introduce a number correlation function $C_j(t_n, t; t^*)$. This is the property of the water molecule j and is equal to either 1 or 0. Water molecule j takes a value of 1 if the water molecule lies

within the first solvation shell of the solute at both time steps t_n and $t + t_n$, and in the interim does not leave the solvation shell for any continuous period longer than t^*. A value of 0 is assigned to water molecule j if it is outside the first solvation shell or leaves the solvation shell for a period longer than t^*. A time period, called t^*, is used to take account of water molecules that leave the first solvation shell only temporarily and return to it without ever having properly entered the bulk solution. A typical value for t^* is 2 ps, since this is the residence time for pure water [9].

The residence times were calculated for the waters in the first hydration sphere (both inside the cavity and outside the CD), with a cutoff radius of 15 Å used. A t^* value of 2 ps was used, the time of diffusion of water in the bulk solution. The residence times were found to fit to one exponential curve of the form

$$f(x) = m1\exp\left(-\frac{x}{m2}\right) \tag{5.5}$$

where m1 and m2 are exponential coefficients and m2 is the residence time.

The residence times, of water in the first solvation shell, were found to be 11.7, 21.9 and 10.2 ps for α-, β- and γ-CD respectively, which goes according the solubility trend, Figure 5.4. Thus the waters in the first solvation shell for β-CD spend a longer time around the β-CD when compared to the other two CDs, thus explaining the higher water probability distribution around β-CD.

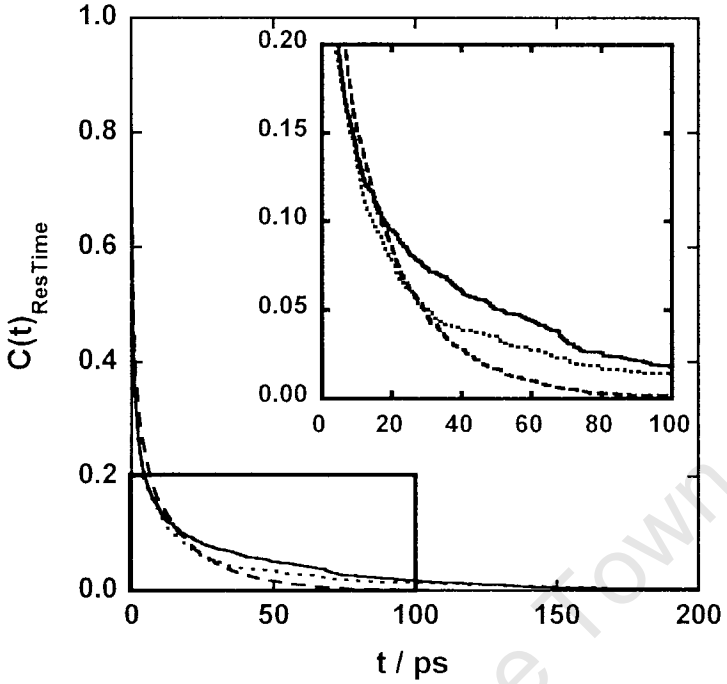

Figure 5.4: The residence times (time correlation functions) are plotted. It can be seen in the insert that the waters spend a longer time around β-CD when compared to α- and γ-CD. (·····), (——) and (----) denote α-, β- and γ-CD, respectively.

Free Energy of Interaction between CD and Water

The "free energy" of interaction between the CDs and water was calculated using all the configurations of water about the CD in the 30 ns trajectory; with the CDs recentered in the centre of the simulation box. The interaction energy is simply the sum of the of the van der Waals and Electrostatic energies between the water and the CD. The calculation was performed from the centre of mass of each CD, out toward 15 Å. The way in which these interaction energies where computed are akin to the method used for the pair distribution function. The volume within and around the CD where divided up into spherical shells, with a shell size of 0.5 Å, and the water-CD interaction energy was then calculated within each shell. The averaging of the energy was done using the total number of frames in the trajectory.

$$\Delta G = \frac{1}{N_T}\sum_{1}^{T}\sum_{1}^{N}\delta(R)(\Delta V_{VdW} + \Delta V_{Elec}) \tag{5.6}$$

where T is the total number of atoms, N is the total number of waters and R is the distance between the water and the centre of mass of the CD. The distances at which

the peaks of the "free energy" curve and the pair distribution function occur are similar, Figure 5.5.

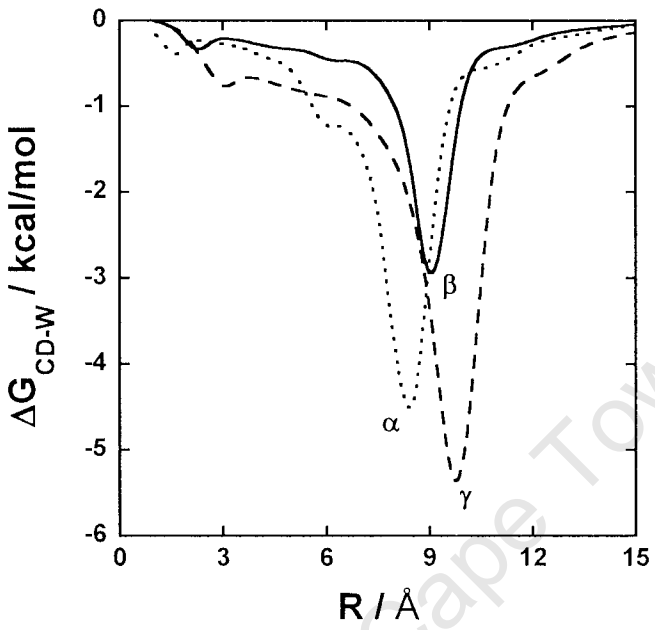

Figure 5.5: The "free energy" of interaction curves are plotted as a function of the distance between the water and the centre of masses of the CDs.

The peak heights are in the following order from smallest to largest: β-, α- and γ-CD, which follows the solubility trend. The water on the inside of the CD cavity have weaker interaction energy than the water on the outside of the CD, indicating the strong hydrophobic nature of the cavity. The "free energy" of interaction between the CD and water was found to be -0.6, -0.4 and -0.8 kcal/mol for the water molecules inside the cavity and -58.8, -56.1 and -59.2 kcal/mol for the water molecules outside the CD, for α-, β- and γ-CD respectively. In chapter 4 we showed that β-CD is conformationally less flexible than α- and γ-CD and as a result the local water structuring probability of β-CD is greatest as shown by Naidoo et al [1]. This conformational inflexibility can be used to explain the reason why the "free energy" of interaction between of β-CD and water is the lowest of the three CDs. The high water probability regions, as shown by the spatial distribution functions (Figure 5.3), shows that the water has a very high probability of being located in the hydrophobic interior of the β-CD, compared to the other two CDs. Due to the relative inflexibility of β-CD water molecules are allowed to form cage structures around the hydrophobic groups of the CD and hence we get a weaker "free energy" of interaction energy,

which results from interaction of the water with the hydrophobic portions of β-CD. It was shown by Burakowski and Glinski that as we add a –CH2– to an alkyl chain, then the hydration number increases by one water molecule. We deduce from this that the relative inflexibility of β-CD results in it having a higher water probability distribution around its hydrophobic sites, resulting in a "free energy" of interaction that is significantly lower than α- and γ-CD.

The isentropic compressibility gives a measure how strongly the solute and the solvent, in the first solvation shell, interact; a higher isentropic compressibility indicates a weak interaction and vice versa. The results obtained show that β-CD has the highest isentropic compressibility and thus has the lowest interaction with the surrounding solvent. This is further emphasized by the "free energy" of interaction between the water and the CDs, where β-CD also showed the least favourable interaction with the water molecules in the first solvation shell. The hydration number was found to be the highest for β-CD, α-CD was the intermediate and γ-CD had the lowest hydration number. This further confirms the result obtained by Naidoo et al. [1] that β-CD has a higher water probability distribution around it, as shown by the SDFs, i.e. more water around the solute will result in a higher probability distribution. A higher water probability distribution should also imply that the waters remain around β-CD for longer; this statement is then justified by the residence time correlation function, which shows that the residence for water, around β-CD, is the longest.

5.4 Conclusions

Ultrasonic measurements have been made to ascertain the hydration number of each of the three most commonly used CDs. The results show that an increase in the ultrasonic speed and the density, involves a decrease in isentropic compressibility, with an increase in the hydration number for all three CDs. The computational studies showed that the water probability distributions had a strong correlation with the solubility trend, with β-CD having the highest water probability distribution. Also in agreement with the SDFs, the residence time of water in the first solvation shell shows that the water molecules, immediately around β-CD, reside there longer. The "free energy" of interaction shows that the interaction between relatively rigid macrocyclic

ring of β-CD and water has the lowest energy due it having a higher water probability distribution around the hydrophobic sites. This is confirmed by the isentropic compressibility data from the ultrasonic results, which shows that isentropic compressibility values of the α- and γ-CD solutions are lower than in the β-CD solution. Thus the following solute-solvent interaction trend is observed as α-CD > γ-CD > β-CD in the thermoacoustic study. The agreement between the ultrasonic computational experiments shows that there is a greater solvation sphere around β-CD and that water molecules reside around the hydrophobic sites, which results in the lowest "free energy" of interaction for β-CD, whereas a higher solute-solvent interaction is observed for the other two CDs. Further studies about the hydration around hydrophobic sites and residence times are currently being conducted. Results from these studies evidently show that the water molecules in the β-CD solution occupy spaces around the hydrophobic sites, resulting in a lower solute-solvent interaction, and as hence tend to reside there longer.

Another study has also shown that the water probability distribution around β-CD is greater compared to α- and γ-CD [11]. They also computed the free energy of solvation, which showed that it increased with the number glucose units and hence the anomalous solubility of β-CD cannot be explained by its solvation free energy alone. Perez-Miron et al. [12] have studied the methylated β-CDs in solution using molecular dynamics. They showed that as the degree of methylation increases, the solvation free energy and water probability distribution decreases. The overall effect of methylation increases the flexibility of the macrocyclic ring conformation of β-CD, which results in a lower water probability distribution around the methylated β-CDs. Therefore the conclusion that can be drawn is that flexibility, as shown in the previous chapter, and water probability distribution is the key factors for the anomalous solubility of β-CD. However, the underlying chemical nature for relative rigidity of the β-CD macrocyclic ring has to still be investigated.

References

1. (a) K. J. Naidoo, J. Y-J. Chen, J. L. M. Jansson, G. Widmalm, A. Maliniak. *J. Phys. Chem. B.* **2004**, *108*, 4236. (b) T. Shikata, R. Takahashi, Y. Satokawa. *J. Phys. Chem. B.* **2007**, *111*, 12239.
2. H. Uedaira, M. Ikura, H. Uedaira. *Bull. Chem. Soc. Jpn* **1990**, *63*, 3376.
3. D. F. Liu, T. Wyttenbach, P. E. Barran, M. T. Bowers. *J. Am. Chem. Soc.* **2003**, *125*, 8458.
4. T. Shikata, R. Takahashi, A. Sakamoto. *J. Phys. Chem. B.* **2006**, *110*, 8941.
5. (a) U. Kaatze, F. Eggers, K. Lautscham. *Meas. Sci. Technol.* **2000**, *11*, 1432. (b) A. P. Sarvazyan. *Ann. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 321.
6. (a) H. Hailand, L. H. Hald, O. J. Kvammen. *J. Sol. Chem.* **1981**, *10*, 775. (b) E. Junquera, D. Dania Olmos, E. Aicart. *Phys. Chem. Chem. Phys.* **2002**, *4*, 352. (c) S. Kristine, H. Hailand. *J. Sol. Chem.* **2003**, *31*, 149. (d) H. Hailand, G. M. Ferrland. *J. Phys. Chem.* **1993**, *97*, 6885. (e) P. S. Nikam, H. R. Ansari, H. Hasan. *J. Mol. Liq.* **2000**, *84*, 169. (f) P. S. Nikam, H. R. Ansari, H. Hasan. *J. Mol. Liq.* **2000**, *87*, 97. (g) D. E. Smith, W. C. Winder. *J. Food Sci.* **1983**, *48*, 1822.
7. A. Burakowski, J. Glinski. *Chem. Phys.* **2007**, *332*, 336.
8. (a) A. Ali, A. K. Nain, B. Lal, D. Chand. *Phys. Chem. Liq.* **2005**, *43*, 1. (b) A. Ali, A. K. Nain, D. Chand, R. Ahmad. *Bull. Chem. Soc. Jpn.* **2006**, *79*, 702. (c) A. K. Nain. *Fluid Phase Equilibria.* **2007**, *259*, 218.
9. S. A. Galema, H. Hailand. *J. Phys. Chem.* **1991**, *95*, 5321.
10. L. Pauling. *The Nature of Chemical Bond.* Cornell University Press, Ithaca, New York, 1948.
11. W. Cai, T. Sun, X. Shao, C. Chipot. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3236.
12. J. Perez-Miron, C. Jaime, P. M. Ivanov. *Chirality.* **2008**, *20*, 1127.

Chapter 6

Association and Solution Properties of "Capped" β-Cyclodextrin

6.1 Introduction

In the previous two chapters we discussed the conformational, dynamical motions and solutions properties of the three main CDs (namely α, β and γ). We now focus on a modified β-CD, 3-*O*-(2-methylnaphtyl)-β-CD or "Capped" β-CD, and study its dimer associative and binding characteristics. Experimental results have shown that this type of modified β-CD can form three distinct types of dimer conformations, discussed later in this chapter. These dimer conformations are important from an inclusion capability and stochiometric point of view.

Modified cyclodextrins (MCDs) are known to behave better than natural cyclodextrins (CDs) for molecular recognition purposes, and MCDs are also more soluble than CDs in solvents other than water [1]. Several such modifications have been published [2]. The modifications introduced on a CD are normally of two types: i) substitutions of hydroxyls by other functional groups, partial or total (persubstitution), and ii) addition of regiospecific linkage fragments onto one of the CD rims to obtain mono or multiple CD-derivatives. The most popular functional groups introduced are ethers and ester. Here ethers such as methyl, ethyl, (2-hydroxy) propyl and pentyl ethers are widely used for industrial applications amongst others [3], while the ester derivatives are used as precursors for the linkage of other organic fragments [2].

Fragments located over the CD cavity are called "Capped" CDs [4]; the first of these had substituents linked to only one glucose ring and were referred to as flexible, Figure 6.1a. Only one year later, the first compound was prepared which had a single fragment bonding two opposite glucoses (forming an intramolecular bridge) and are referred to as rigid, Figure 6.1b [5]. Capped CD's have larger cavities compared with the naturally occurring CDs and are therefore able to complex more ligands. CDs are intrinsically chiral, inducing a circular dichromism on the guest even if it is achiral,

which makes it possible to measure the complexation constant from changes in the induced circular dichromism [6].

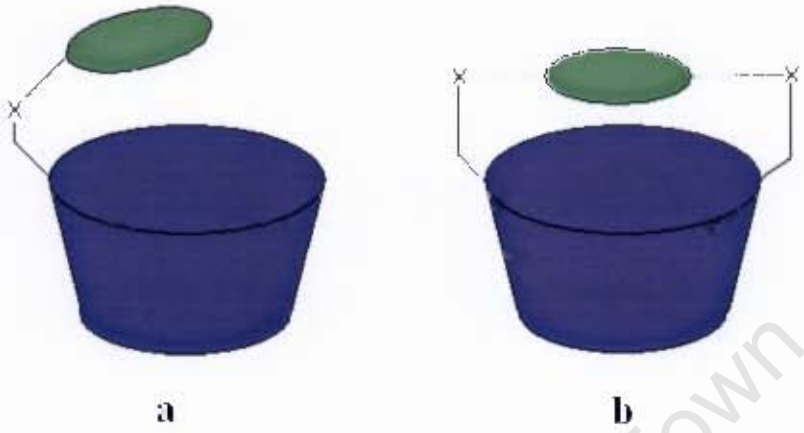

Figure 6.1: Schematic drawing representing the two types of "Capped" CDs: a) flexible and (b) rigid.

Investigations into the energetic balance in the complexation processes, where the hosts were natural CDs, were researched [7-14]. In this chapter, we present a study of the conformational associative range and energetics of dimeric 3-*O*-(2-methylnaphtyl)-β-CD (hereinafter called Cap-βCD). The structural details of Cap-βCD have remained uncertain from its synthesis [15]. The authors postulated first a temperature-dependent inside-outside isomerism of the naphtyl group. Later, aggregate structures with the naphtyl group inside the cavity of another CD were proposed based on spectroscopic studies [16]. However, ^{1}H-NMR spectra over a wide concentration range strongly suggested that Cap-βCD form dimers at concentrations higher than $1.6 \cdot 10^{-4}$ M; with the appearance of monomers at lower concentrations. Circular dichromism analysis, revealed the presence of a) monomers displaying a "cap-like" structure with the naphtyl group only partially included and b) three different dimers with the possible arrangements of two head-head and one head-tail configurations, Figure 6.2. The two head-head dimers differ in the position of the naphtyl groups; in the one they are inside the cavities and in the other they are located outside and parallel (hereinafter we refer to them as HHi and HHp, respectively).

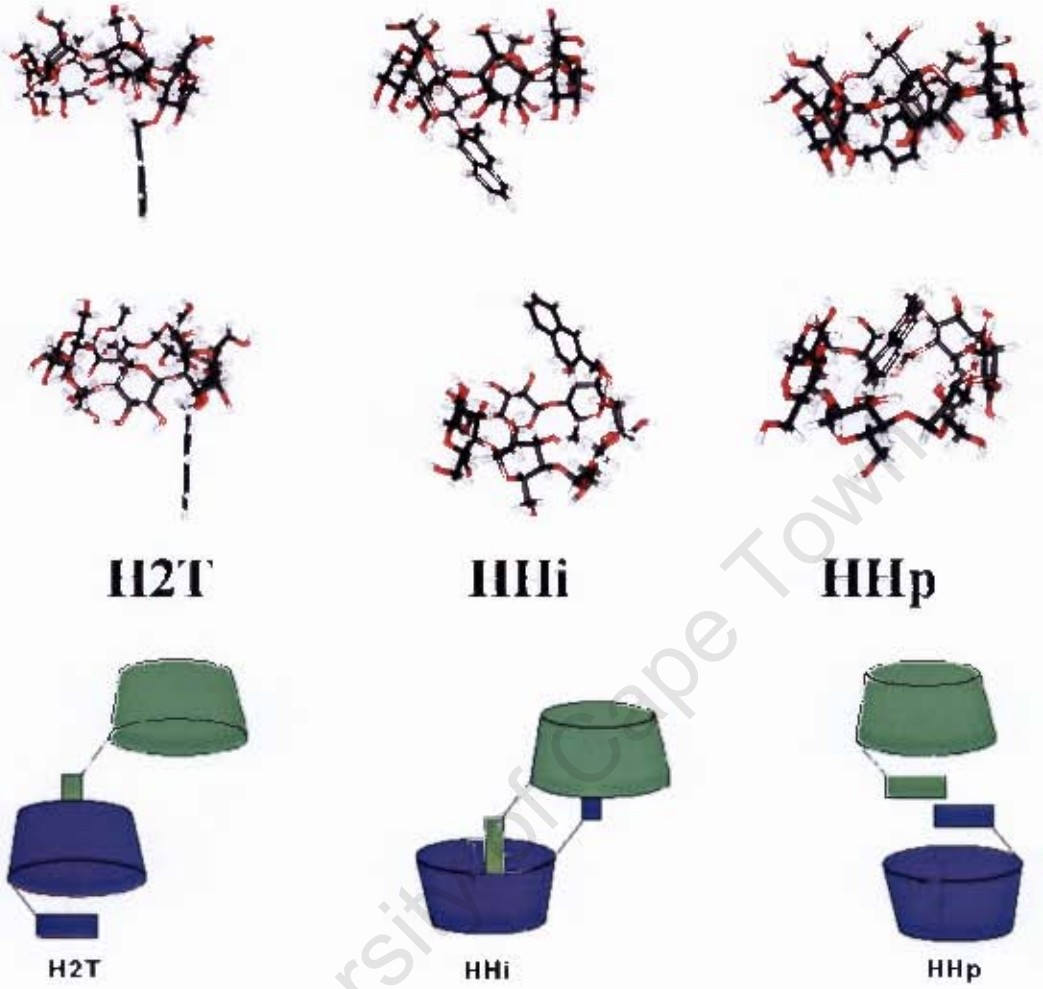

Figure 6.2: The three proposed possible dimers for Cap-βCD: H2T, HHi and HHp. The green monomer is B CD and the blue monomer is BCD1.

The interglycosidic linkage dihedral angles are defined as previously discussed in chapter 4, Figure 6.3a. The dihedral angles between the glucose and the methyl naphthalene group are defined as: ϕ_C = C3'-C2'-C1'-O3, ψ_C = C2'-C1'-O3-C3 and ω_C = C1'-O3-C3-C2, Figure 6.3b. The ϕ_C dihedral angle represents the rotations of the aromatic ring around itself; while ψ_C dihedral angle describes the location of the aromatic ring – being either inside the CD cavity or perpendicular to the CD plane. Finally the ω_C dihedral angle describes the position in relation to the CD cavity, i.e. whether the naphtyl aromatic ring takes on a "cap-like" conformation or if it lies outside the CD and is parallel to the CD plane.

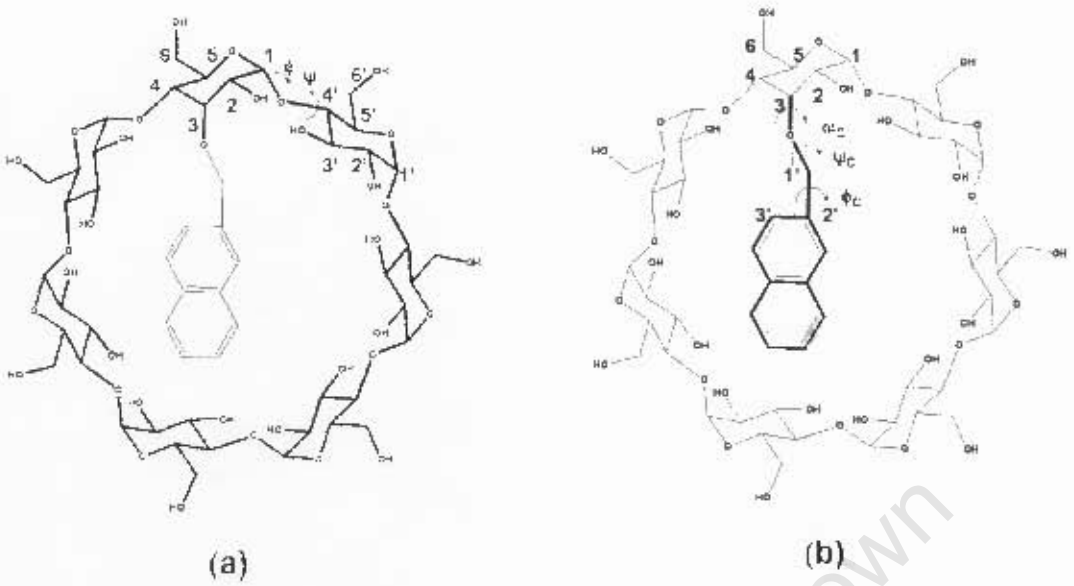

Figure 6.3: (a) Depiction of the ϕ and ψ dihedral angles, the prime represents the adjacent glucose residue, and (b) the o_C, ψ_C and ω_C, here the prime are used to denote the atoms of the methylnaphthalene group. The atoms involved in the dihedral angles are explained in the text.

6.2 Methods

6.2.1 Computational methodology

MD simulations were performed in water using the modelling program CHARMM [17]. Before running the MD simulations, the structures were fully optimized using the steepest descent method followed by conjugate gradient with constraints placed on the glucose intra-annular torsions to prevent transitions to the boat conformers. Non bonded interactions, within a minimum radius of 10 Å, were used. Both the van der Waals and Electrostatic interactions were calculated on a group by group basis using a switching function.

The simulations included a heating phase from 50 to 300 K over 200ps and an equilibration phase at 300 K over further 1ns before production dynamics. The simulations were run for 20ns, using a microcanonical ensemble with the leapfrog Verlet algorithm [18] and bonds involving hydrogens were fixed via the restraint algorithm SHAKE [19]. Simulations were run using an explicit TIP3P water model [20] in a sphere of a 30.1 Å radius under stochastic boundary conditions [21]. The buffer boundary was set at between 25 Å, where the atoms were propagated under Langevin dynamics [22] in this region. A heat bath of temperature 300 K coupled to

the buffer atoms kept the system at thermal equilibrium while a spherical boundary potential was employed to maintain the correct average distribution of water molecules and to prevent water from escaping into the gas phase. A modification was made to the CSFF [23] force field to take into account the correct energy profile describing the rotation between the sugar and aromatic rings.

6.2.2 Angle and Distance between Monomers

The relative conformational association of the Cap-βCD monomers to each other were done as a function of inter CD ring distance. Here the centre of mass for each CD has been calculated using the centres of all the glucose rings (excluding the naphthalene aromatic group). The inter CD ring distance is defined as the separation between CD centres of mass (see R in Figure 6.4). The orientation of the Cap-βCD monomers, relative to each other, were calculated as a function of the angle, θ between the CD ring planes. This angle is defined as the dot product of the normals $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ of each CD plane as shown in Figure 6.4.

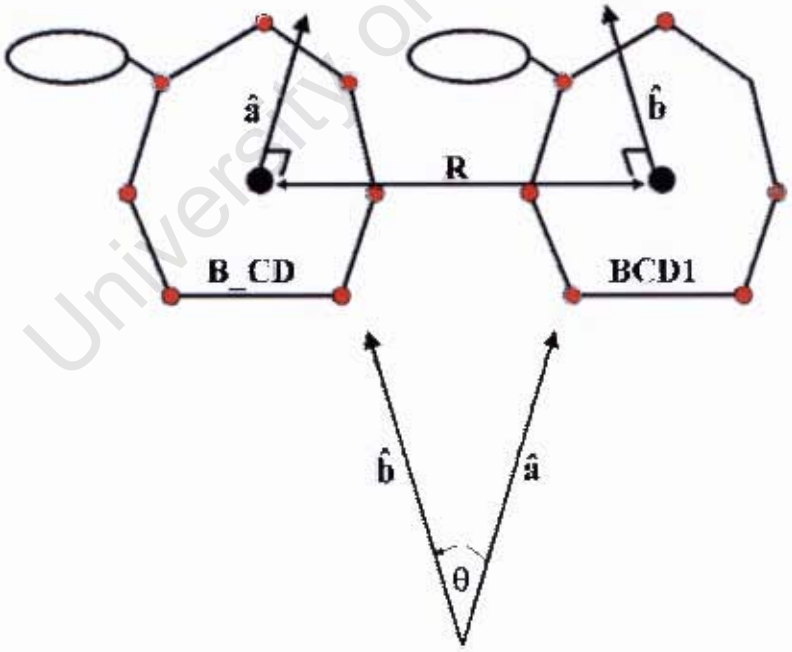

Figure 6.4: Schematic diagram for the calculation of the monomer-monomer angle and distance. The solid black circles represent the centre of mass for the CDs and the red circles represent the glucose residues. The distance between the monomers are calculated between the centres of masses and the normal vectors, $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, are used to calculate the angle between the monomers, using the dot product.

6.2.3 MM-PBSA Methodology

The relative binding free energies were estimated using the Molecular Mechanics-Poisson-Boltzmann Solvent Accessible area (MM-PBSA) methodology, which has become the most popular and commonly used method for estimating the relative binding free energies between monomers. The method requires the internal energy derived from molecular mechanics, the non-polar and the polar contributions to the free energy of binding. The non-polar portion is estimated by using the solvent accessible surface area of the molecules, while the polar portion is estimated using the Poisson-Boltzmann equation for solving the electrostatic contribution to the solvation energy.

$$\Delta G_{Binding} = E_{MM} - \Delta G_{water}(AB) - (\Delta G_{water}(A) + \Delta G_{water}(B)) - T\Delta S \qquad (6.2)$$

$$\Delta G_{water} = \Delta G_{non\text{-}polar} + \Delta G_{polar} \qquad (6.2.1)$$

6.2.4 Hydrogen Bonding

The intermolecular (between the Cap-βCD monomers) hydrogen bonding analysis was used to determine the percentage of hydrogen bonding between all the possible hydrogen bond donors and acceptors between the Cap-βCD monomers. Possible hydrogen bonds can occur between the O2, O3, O4, O5 and O6 atoms; we used the usual geometric limits to assess the possibility of hydrogen bonding, viz. a donor-acceptor distance of less than 3.5 Å and donor-hydrogen-acceptor angle of greater than 135°.

6.3 Results and Discussion

In this paper we assess the dimer conformations and binding free energies of Cap-βCD in solution. We investigate the three possible dimer conformations namely: the H2T, HHi and HHp conformations, Figure 6.2. We examine the relative conformations by considering the distance and angle between the Cap-βCD monomers. The differences in these conformations determine their stability in solution. The interactions between the monomers can be easily described as favourable (hydrogen bonding, Cap-Cap and Cap-Cavity hydrophobic interactions)

and unfavourable (Cap-Water and Cavity-Water interactions) contributions to the overall stability of the dimer conformations, Table 1.

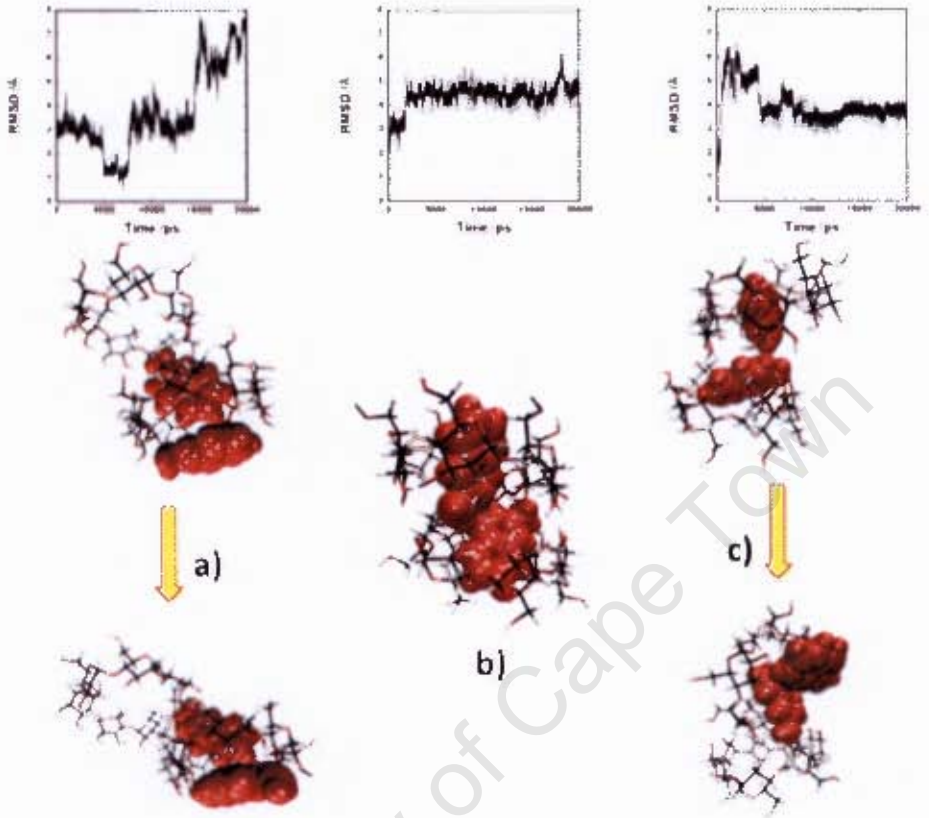

Figure 6.5: The RMSD for the dimer conformations of a) H2T, b) HHi and b) HHp. The conformations of the dimers before and after the transition are also shown for H2T and HHp.

		Before Transition		After Transition	
	HHi	**H2T**	**HHp**	**H2T**	**HHp**
CD-CD HB (%)	66.4	26.4	17.8	16.6	42.5
Cap-Cap Hydrophobic	n	n	y	n	y
Cap-Cavity Hydrophobic	2	1	0	1	1
Cap-Water Interactions	0	1	2	1	1
Cavity-Water Interactions	0	1	2	1	1
Distance (Å)	8.01 (0.32)	8.20 (0.46)	11.77 (1.56)	11.39 (1.06)	9.11 (0.31)
Angle (°)	162.32 (2.32)	31.07 (8.17)	128.43 (23.10)	31.07 (8.17)	124.10 (8.62)
E_{int} (kcal/mol)	-59.77 (4.91)	-40.36 (5.30)	-26.98 (11.09)	-28.36 (5.72)	-45.61 (6.16)

Table 6.1: The interactions between the CD monomers and solvent before and after a transition have occurred. The numbers in parentheses are the standard deviations.

6.3.1 Conformation and Interaction Energy

The root mean square deviation (RMSD) was used to ascertain by how much the dimer conformations changed with respect to the initial starting conformation, Figure 6.2, for each of the three dimer conformations, Figure 6.5. The HHi dimer conformation, Figure 6.5b, showed a stable RMS deviation from its initial conformation. However the H2T and HHp dimer conformations showed significant deviation from their initial conformations, Figure 6.5a and c, which can be seen as large changes in the RMS deviation during the simulation. For the H2T, a change in the cap conformation is observed, as the RMSD decreases after 5ns. The cap of the B_CD monomer forms a perpendicular conformation, with respect to the CD plane. The RMSD decreases as the staring conformation had this cap in a perpendicular conformation. A larger transition in the conformation is observed after 8ns when the entire dimer conformation becomes unstable. The HHp dimer conformation immediately undergoes a transition, because the parallel conformation of the cap groups is not favourable. After approximately 7ns the dimer conformation reaches a stable conformation as can be seen by a stabilisation in the RMSD. A more detailed analysis of the conformation and interaction energy is shown in Figure 6.6.

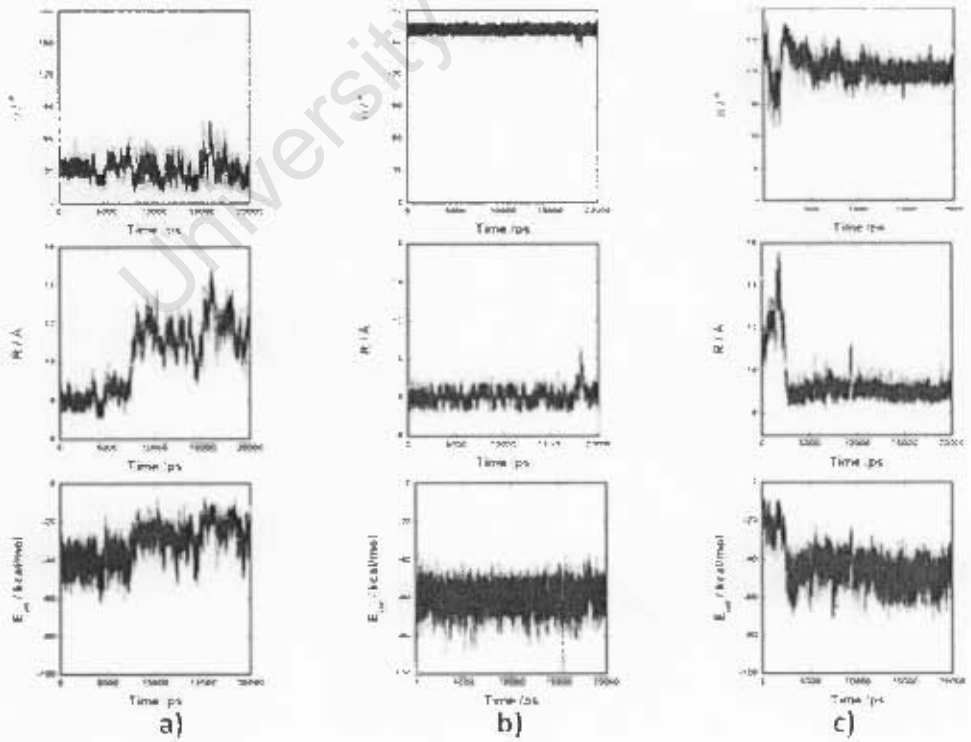

Figure 6.6: The monomer monomer angle, distance and the interaction energy between the monomers for the a) H2T, b) HHi and c) HHp dimer conformations.

The distance and angle between the monomers are defined in Figure 6.4 – perfect head-to-tail and perfect head-in-head conformations are defined as having an angle of 0 and 180° respectively. The interaction energy is calculated as the sum of the non-bonded (van der Waals and electrostatic) interactions between the monomers. From the distance and angle between the monomers it can be seen that the HHi, Figure 6.6b, dimer conformation does not undergo any major conformational changes and remains close to its initial parallel conformation (with respect to the CD planes) throughout the simulation, Figure 6.6b, hence a stable interaction energy is observed; whereas the H2T and HHp, Figure 6.6a and c, have significant changes in the distance between the monomers and interaction energy (indicative of a conformational change). The HHi has the highest percent of hydrogen bonding compared to the other two dimer conformations, Table 6.1, which results in it having a more favourable interaction throughout the simulation.

The H2T conformation is initially stable before the transition at 7ns, with a dimer conformation where the monomers are nearly on top of each other, Figure 6.5a. Significant intermolecular hydrogen bonding is observed during the period before the transition and together with a short distance between the monomers, a very favourable interaction energy is observed. However, this interaction energy is almost 20 kcal/mol less favourable than the HHi dimer; this is a result of the significantly lower hydrogen bonding percentage, the B_CD cap and the BCD1 cavity being occupied by water, Figure 6.8a. After the transition, at 7ns, the distance and interaction energy change dramatically as the conformation changes significantly from where the CDs where on top of each to being almost adjacent to each other, Figure 6.5a. This conformation is unstable and as a result of the change the amount of hydrogen bonding decreases dramatically.

The HHp conformation is unstable from the start of the simulation, before the transition that occurs at 3ns; immediately the distance and angle between the monomers change significantly, Figure 6.6c. The interaction is highly unfavourable, due to the caps and cavities being exposed to the water and there is also a very low hydrogen bonding percentage. This low hydrogen bonding percentage can be explained by observing that the CDs move further apart, during this period, and hence the possibility for forming intermolecular hydrogen bonding significantly decreases. During this transition phase in the HHp simulation, the cap groups go from capping their CDs cavity to forming a perpendicular conformation (with respect to their CD

plane). The distance clearly illustrates this point, because there is a large change in the distance of approximately 6 Å – which is roughly equal to the length of the naphthalene rings. This occurs so that the cap groups can include into the cavity of the other monomer. However, only the BCD1 cap group includes into the B_CD cavity and the B_CD cap misses being included in the BCD1 cavity, but it does cap its CDs cavity, Figure 6.6c, allowing for possible π-stacking with the BCD1 cap. After the transition the conformation is very stable with respect to its conformation (distance and angle) and its interaction energy, Table 6.1. The interaction energy profile looks similar to the HHi conformations interaction energy but is at least 14 kcal/mol less favourable, which is mainly due to hydrophobic B_CD cap group not being included in the BCD1 cavity and also this cavity is also exposed to the solvent. The interaction energy is, however, more favourable than the H2T dimer due to the increase in the percentage of hydrogen bonding observed after the transition, Table 1.

6.3.2 Binding Free Energy and Further…

The MM-PBSA methodology was used to estimate the relative binding free energy between the monomers. The binding free energy was calculated for each nanosecond of the simulation, rather than the traditional average for the entire simulation, Figure 6.7. From this we expect to be able to ascertain what effect the transitions, of the conformations, have on the binding free energy. For all three dimer conformations the binding free energy was found to favourable, hence these conformations are plausible in solution. The binding free energy for the HHi dimer was found to be stable throughout the simulation, with an average value of -66.72±1.08 kcal/mol, Figure 6.7, which is due to it having only favourable interactions as described in Table 6.1. The H2T had a stable free energy, before the transition, of -54.32±1.60 kcal/mol; after the transition the energy become less favourable and larger fluctuations in the value of the energy was observed (-45.76±5.52 kcal/mol). This is due to the large movements experienced by the B_CD monomer as a result of the swinging motions of the BCD1 cap, which is included in the BCD1 cavity. The HHp dimer's free energy was found to be less favourable (-54.68±6.34 kcal/mol) before the transition, which was expected as the monomers moved apart to allow for the outward motion of the caps, and the subsequent inclusion of the BCD1 cap in the B_CD cavity. After the transition, the energy was stable with an average value of -56.37±2.64 kcal/mol, Figure 6.7.

We expect there to the H2T and HHp dimer conformations to make a final transition to the IIIIi form; this would require large simulation times to observe these transitions. The conformational analysis, interaction energy and binding free energy for the H2T all support that it has large conformational fluctuations and is not very stable. It has too many unfavourable interactions, Table 6.1, to remain in the final conformation we observe from our simulations. On the other hand, the HHp dimer conformation seems to be stable after the 20ns simulation. However, there is still the unfavourable Cap-Water and Cavity-Water interactions; and given a long enough simulation with sufficient internal energy, the system will minimize its energy by forming the IIIIi conformation.

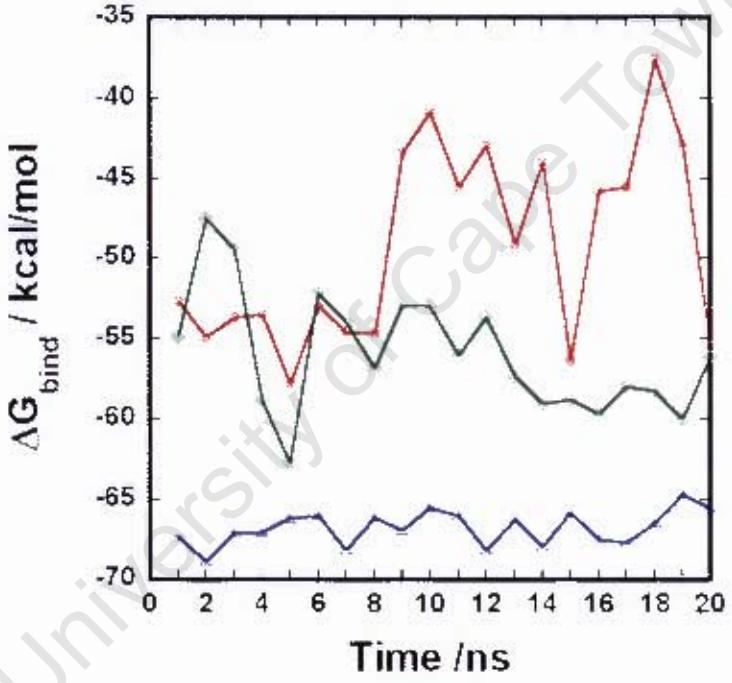

Figure 6.7: The relative free energy of binding curves, plotted at 1ns increments, for the H2T (red curve), HHi (blue curve) and HHp (green curve) dimer conformations.

6.3.3 Solvent Distribution Functions

The pair distribution functions (PDFs) were calculated before and after transitions (in the case of the H2T and HHp dimer) and for the entire trajectory for the HHi dimer conformation. The centre of mass for each CD was used to calculate the site-site PDF. This was achieved by first calculating the centre of massed for glucose monomers using atoms C1, C2, C4 and C5 and the centre of CD macrocyclic ring was defined using the glucose monomer centre of masses. This type of analysis gives a good indication of how the solvent structuring changes as the aforementioned dimer

conformations make a transition. While PDFs are a radially averaged picture of the solvent structuring around a solute molecule, spatial distribution functions (SDFs) gives a more spatial description of solvent structuring around the solute. The translation and rotational motions were removed by fitting the structures to an initial structure using the dihedral angle atoms for the fitting procedure. Again SDFs are calculated before and after the transition seen for the H2T and HHp dimer conformations.

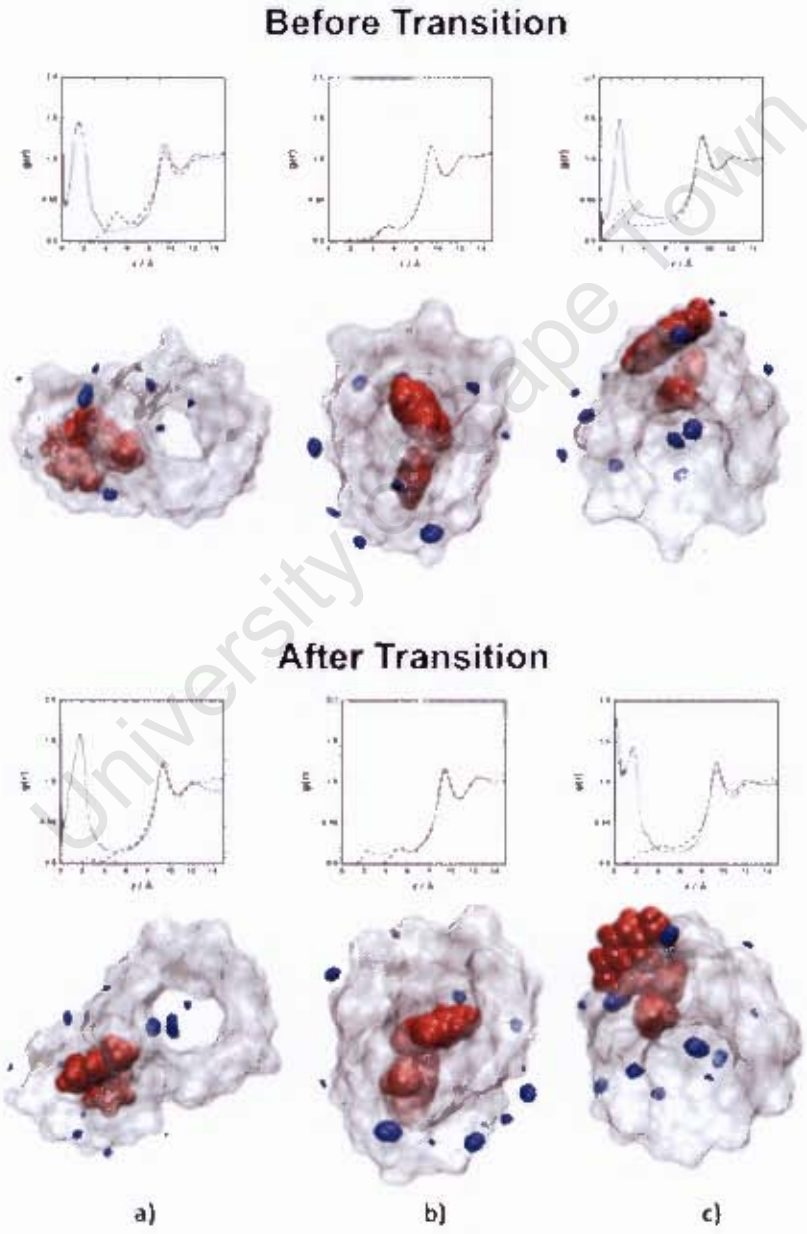

Figure 6.8: The PDFs and SDFs for the a) H2T, b) HHt and c) HHp dimer conformations. In the PDFs, the solid line is the B_CD monomer and the dashed line is the BCD1 monomer. The SDFs are all plotted at 85% above bulk density.

The pair distribution function for the HHi dimer remains the same throughout the simulation for both monomers, showing no water structuring within the cavities of the CDs, as expected. However a small peak at approximately 5 Å, Figure 6.8b, shows that there is a slight probability of finding waters above the glucose units. The PDFs for the H2T dimer shows a slightly different solvent structuring before and after the transition, Figure 6.8a. For the BCD1 monomer (which has its cap included in the cavity of B_CD) the solvent structuring inside its cavity remains the fairly constant. The difference for this monomer is that it has a stronger second solvation shell than before the transition. The B_CD monomer, however, loses its water probability above the glucose units after the transition, but a stronger first solvation shell is observed, as the aforementioned water probability is more dispersed further out into the first solvation shell. The HHp BCD1 monomer has a change in its solvent structuring inside the cavity, Figure 6.8c. The distribution becomes broader after the transition, indicating that the cavity has become larger to allow for more waters to found there. However, the B_CD monomer has lost a large water probability inside the cavity, but there is now a slightly increased probability of finding waters above the glucose units.

The SDFs display a more visual picture of how the structuring changes after the transition, in the case of the H2T and HHp (Figure 6.8a and c) dimer conformations, and how it remains the same for the HHi dimer conformation (Figure 6.8b). All SDFs were plotted at 85% above bulk density. We can clearly see that the probability distribution above the glucose units of the BCD1 monomer, in H2T conformation, disappears after the transition. It's also clear to see the probability decrease in the cavity of the BCD1 monomer. The HHi dimer's probability distribution remains the same throughout the trajectory, Figure 6.8b. The HHp SDFs shows a more constricted distribution inside the cavity of the BCD1 monomer, while some probability density is seen nearby the cap of BCD1 inside the B_CD cavity, hence resulting an increase in density inside the B_CD cavity (as seen from the PDF).

6.4 Conclusions

From the angle and distance between the monomers, interaction energy, distribution functions and MM-PBSA binding free energy calculations the HHi dimer conformation was found to be the most stable. The HHp dimer conformation initially

had an unfavourable conformation but quickly stabilized by forming a conformation comparable to that of the HHi conformation – with the exception of the B_CD cap not being included in the BCD1 cavity. However, the H2T conformation was found to be the least favourable of the three; initially it showed a stable conformation but after the transition, where the BCD1 monomer was no longer directly above the B_CD monomer but almost adjacent to it, it became less stable with large fluctuations in its conformation and binding free energy. The pair and spatial distribution functions showed a different water probability density around the H2T and HHp dimer before and after the conformational transitions. The water probability around the HHi dimer remained the similar throughout the simulation. We therefore conclude that of the three dimer conformations studied, namely the H2T, HHi and HHp dimer conformations – the HHi dimer conformation is the most stable dimer conformation. We can speculate from the conformational analysis and binding free energy that both the HHp and H2T conformations will make a final transition to form the HHi conformation. We therefore speculate that the transition from the final H2T and HHp dimer conformations, we observe, to the minimum energy HHi dimer conformation will occur given that the simulation time and energy, in the system, are large enough. A more accurate method would be to perform a free energy calculation, such as a potential of mean force, to understand the full chemical nature of each dimer conformation and the free energy associated with making the aforementioned transitions to the HHi dimer conformation.

References

1. As an example see: K. Uekama, F. Hirayama, T. Iire. *Chem. Rev.* **1998**, *98*, 2045.
2. a) Y. Liu, B. H. Han, S. X. Sun, T. Wada, Y. Inoue. *J. Org. Chem.* **1999**, *64*, 1487. b) R. Breslow, B. Zhang. *J. Am. Chem. Soc.* **1996**, *118*, 8495. c) R. Breslow, N. Greenspoon, T. Guo, R. Zarzycki. *J. Am. Chem. Soc.* **1998**, *111*, 8296. d) A. R. Khan, P. Forgo, K. J. Stine, V. T. D'Souza. *Chem. Rev.* **1998**, *98*, 1977.
3. J. Szejtli. *J. Mater. Chem.* **1997**, *7(4)*, 575.
4. J. Emert, R. Breslow. *J. Am. Chem. Soc.*, **1975**, *97(3)*, 670.
5. I. Tabushi, K. Shimokawa, N. Shimizu, H. Shirakata, K. Fujita. *J. Am. Chem. Soc.* **1976**, *98(24)*, 7855.
6. A. Ueno, H. Yoshimura, R. Saka, T. Osa. *J. Am. Chem. Soc.* **1979**, *101(10)*, 2779.
7. a) J. Redondo, C. Jaime, A. Virgili, F. Sánchez-Ferrando. *J. Molec. Struct.* **1991**, *248*, 6521; b) X. Sánchez-Ruiz, A. Alvarez-Larena, C. Jaime, J. F. Piniella, J. Redondo, A. Virgili, F. Sánchez-Ferrando, G. Germain, F. Baert. *Supramol. Chem.* **1999**, *10*, 219; c) C. Jaime, J. Redondo, F. Sánchez-Ferrando, A. Virgili. *J. Org. Chem.* **1990**, *55*, 4773; d) P. Ivanov, C. Jaime. *Anales de Química, Int. Ed.* **1996**, *92*, 13; e) P. M. Ivanov, C. Jaime. *J. Molec. Struc.* **1996**, *377*, 137; f) P. Ivanov, D. Salvatierra, C. Jaime. *J. Org. Chem.* **1996**, *61*, 7012; g) D. Salvatierra, C. Jaime, A. Virgili, F. Sánchez-Ferrando. *J. Org. Chem.* **1996**, *61*, 9578; h) M. Zubiaur, C. Jaime. *J. Org. Chem.* **2000**, *65*, 8139.
8. a) F. Fotiadu, M. Fathallah, C. Jaime. *J. Org. Chem.* **1994**, *59*, 1288; b) F. Pérez, C. Jaime, X. Sánchez-Ruiz. *J. Org. Chem.* **1995**, *60*, 3840.
9. a) E. Cervelló, C. Jaime. *J. Molec. Struct. (THEOCHEM).* **1998**, *428*, 195.; b) E. Cervelló, C. Jaime. *Anales de Quím. It. Ed.* **1998**, *4-5*, 244; c) E. Cervelló, F. Mazzucchi, C. Jaime. *J. Molec. Struct. (THEOCHEM).* **2000**, *530*, 155; c) X. Sánchez-Ruiz, M. Ramos, C. Jaime. *J. Molec. Struct.* **1998**, *442*, 93.

10. a) D. Salvatierra, X. Sánchez-Ruiz, R. Garduño, E. Cervelló, C. Jaime, A. Virgili, F. Sánchez-Ferrando. *Tetrahedron.* **2000**, *56*, 3035; b) I. Beà, C. Jaime, J. Redondo, P. Bonnet, A. Torrens, J. Frigola. *Supramol. Chem.*, **2001**, *14*, 33.
11. a) A. Entrena, C. Jaime. *J. Org. Chem.* **1997**, *62*, 5923; b) M. A. Fernández, R. H. de Rossi, E. Cervelló, C. Jaime. *J. Org. Chem.* **1991**, *66*, 4399.
12. a) P. Bonnet, C. Jaime, L. Morin-Allory. *J. Org. Chem.* **2001**, *66*, 689; b) P. Bonnet, C. Jaime, L. Morin-Allory. *J. Org. Chem.* **2002**, *67*, 8602; c) P. Bonnet, I. Beà, C. Jaime, L. Morin-Allory. *Supramol. Chem.* **2003**, *15*, 251-260; d) A. C. S. Lino, Y. Takahata, C. Jaime. *J. Molec. Struct. (THEOCHEM).* **2002**, *594*, 207.
13. a) I. Beà, C. Jaime, P. A. Kollman. *Theor. Chem. Acc.*, **1992**, *108*, 286; b) I. Beà, E. Cervelló, P. A. Kollman, C. Jaime. *Comb. Chem and High Throughput Screening.* **2001**, *4*, 605.
14. P. M. Ivanov, C. Jaime. *J. Phys. Chem. B.* **2004**, *108(20)*, 6261-6274.
15. S. R. Mc. Alpine, M. A. Garcia-Garibay. *J. Am. Chem. Soc.* **1996**, *118*, 2750.
16. S. R. Mc. Alpine, M. A. Garcia-Garibay. *J. Am. Chem. Soc.* **1998**, *120*, 4269.
17. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. J. Karplus. *J. Comput. Chem.* **1983**, *4*, 187.
18. a) L. Verlet. *Phys. Rev.* **1967**, *159*, 98; b) L. Verlet. *Phys. Rev.* **1967**, *165*, 201.
19. W. F. van Gunsteren, H. J. C. Berendsen. *Mol. Phys.* **1977**, *34*, 1311.
20. H. J. C. Berendsen, R. J. Grigera, T. P. Straatsma. *J. Phys. Chem.* **1987**, *91*, 6269.
21. C. L. Brooks III, M. Karplus. *J. Chem. Phys.* **1983**, *79(12)*, 6312.
22. S. E. Feller, Y. Zhang, R. W. Pastor, B. R. Brooks. *J. Chem. Phys.* **1995**, *103(11)*, 4613.
23. M. M. Kuttel, J. W. Brady, K. J. Naidoo. *J. Comp. Chem.* **2002**, *23*, 6269.

Chapter 7

Conclusions and Future Work

The primary objective of this thesis was to investigate the anomalous solubility of β-CD in aqueous solution. The secondary objective was to analyse the associative and binding capabilities of dimeric "Capped" β-CD in solution. These objectives have been achieved and some interesting results have been obtained.

The conformational and dynamical analyses performed on the three unmodified CDs revealed that β-CD had the least flexible macrocyclic ring when compared to α- and γ-CD. This was demonstrated by using a combination of distribution and dynamics sets of analyses. The dihedral angle distribution showed the narrowest distribution for β-CD, followed by α-CD and γ-CD had the largest distribution. The dynamic motions for the ψ dihedral angle showed that β-CD motion relaxed the quickest with α-CD slower and γ-CD slowest. The hydrogen boding distributions, across the glycosidic linkage, indicated that there was no co-operative hydrogen bonding around any of the CDs macrocyclic rings, i.e. no hydrogen bonding belt was observed for any of the CDs. However, we observed that β-CD formed more hydrogen bonds. The average tilt angle values compared well with the experimental single X-ray data. The ranges (or distributions) for the CDs were found to be very similar, but the time taken for relaxation to occur followed the pattern where β-CD relaxed the fastest followed by α-CD and then γ-CD. The generalised amplitude for the monomer motion relative to the macrocyclic plane, analysed for each glucose unit, was found to be the greatest for γ-CD and β-CD had the lowest displacement from the macrocyclic plane, α-CD was the intermediate case. The monomer librations relaxed the quickest for β-CD followed by α-CD with γ-CD being the slowest to reach its mean displacement from the macrocyclic plane.

The solution properties of the CDs were investigated using ultrasonic interferometry and molecular dynamics simulations. The results showed, for increasing molality, that an increase in the ultrasonic speed and the density, involves a decrease in isentropic compressibility and an increase in hydration number for all three CDs. The hydration number trend was shown to be consistent with the solubility

trend, i.e. the hydration number was found to be greatest for β-CD, followed by α-CD and γ-CD showed the lowest hydration number. The pair distribution function showed that there was a high probability of finding waters inside the cavities of the CDs, and also showed a high probability outside the CD. This was confirmed by the spatial distribution function, it also showed that β-CD had the highest water probability density around it, followed by α-CD and then γ-CD, which follows the solubility trend. The residence time indicated that the waters, in the first solvation shell, spent a longer time around β-CD when compared to the other two CDs. The "free energy" of interaction pointed out that the interaction energy followed the solubility trend, i.e. γ-CD > α-CD > β-CD. The reason for β-CD having the lowest interaction energy was that the relatively rigid macrocyclic ring of β-CD exposed a rigid hydrophobic surface to the solvent, around which the waters would sit. This was confirmed by isentropic compressibility, which gives a measure of the solute-solvent interactions, following the same trend where β-CD had the highest isentropic compressibility and lowest solute-solvent interaction. However, current research shows that the hydration free energy increases with the number glucose units, and hence the anomalous solubility of β-CD cannot be explained by its hydration free energy alone. Other researchers have studied the methylated β-CDs, in solution, using molecular dynamics. They showed that as the degree of methylation increases, the solvation free energy and water probability distribution decreases. The overall effect of methylation increases the flexibility of the macrocyclic ring conformation of β-CD, which results in a lower water probability distribution around the methylated β-CDs. This supports our argument that the high water probability distribution for β-CD is a key aspect for its anomalous solubility.

The Cap-βCD conformational analysis showed that the HHi dimer conformation was the most stable, with minimal fluctuations in its monomer-monomer angle, distance and interaction energy, when compared to the other two dimer conformations. This was further confirmed by the relative binding free energy, as calculated from the MM-PBSA methodology, showed that the HHi dimer conformation was the most stable, and showed minimal fluctuations in this energy throughout the simulation. The HHp dimer conformation initial showed a large fluctuation in its binding free energy, but became stable after a transition after 7ns; although its final dimer conformation was similar to that of HHi, it had a significantly higher energy. This was due to the BCD1 cap not being included in the B_CD cavity. The H2t dimer also showed very

large fluctuations in its relative binding free energy, with no stabilization evident throughout the simulation. This was due to one of monomers cavities being exposed to the solvent, as well as one of the caps was also exposed to the solvent. The solvent structuring (pair and spatial) distribution functions indicated that water structures strongly around the outside of these dimer conformations, but the monomers which have no cap group included in their cavity had a high water probability distribution in them. Therefore of the three possible ways for these CDs to associate, but the HHi is most likely to be the favoured dimer conformation, based on the conformational and binding free energy analyses.

We have not yet been able to determine the actual chemical reason as to why the macrocyclic ring of β-CD is less flexible and hence least soluble. Thus, future work would include performing PMF calculations about the interglycosidic linkage dihedral angles, on the monomer tilting and vertical librations; then we should have a clear chemical understanding for the anomalous solubility of β-CD. Future work on the Cap-βCD would be to perform free energy calculations, such as a potential of mean force, to understand the full chemical nature of each dimer conformation and the free energy associated with the H2T and HHp making transitions to the HHi dimer conformation, which would prove our postulation that HHi is the most favourable dimer conformation.