# CONFORMATIONAL ANALYSIS OF DIGLYCOSYL DISULPHIDES CONTAINING $\beta$S(1$\rightarrow$1')S INTERGLYCOSIDIC LINKAGES

Dissertation presented to the

**UNIVERSITY OF CAPE TOWN**

In fulfilment of the requirements for the degree of

**MASTER OF SCIENCE**

By

**RICHARD P. MATTHEWS**

**BSc Hons**

Supervisor: Assoc. Professor Kevin J. Naidoo

Department of Chemistry

University of Cape Town

2007

# ACKNOWLEDGEMENTS

# ABSTRACT

The conformation around the glycosidic linkage has been shown to be the single most important factor in determining the molecular shape of oligosaccharides. This property is of fundamental importance in influencing biological activity such as binding to enzymes or receptors. Therefore, knowledge of the conformational preference associated with glycosidic linkages is required.

In this thesis the conformational preferences of the $\beta$S(1→1')S glycosidic linkage within asymmetrical dissacharide mimetics have been studied using computational methods. The sulphur-sulphur torsion angle parameters contained in the CHARMM22 force field were refined based on density functional theory (DFT) calculations of dimethyl disulphide, which is a representative fragment containing the sulphur-sulphur torsion angle. The refined parameters were then used in molecular dynamics (MD) simulations of three disaccharide mimetics in vacuum, water and *dimethyl sulphoxide* (DMSO).

The resulting conformational analysis reveals that in the case of disaccharides containing unsubstituted glycosyl rings, good agreement with experimental *nuclear overhauser effects* (NOEs) and spin-spin coupling constants is obtained. However, for disaccharides containing substituted glycosyl rings poorer agreement with experimental data is obtained, which may suggest that further refinement of the force field is required.

In addition, a natural bond orbital analysis was conducted on the 1, 1' – disulphanediyl diethanol fragment to investigate the origin of the conformational preference for the +g and –g conformers of the sulphur-sulphur torsion angle. This phenomenon is shown to be as a result of a two-electron stabilising $n_0 \rightarrow \sigma^*_{C-S}$ delocalisation from the lone pair of the oxygen atoms to the carbon-sulphur anti-bonding orbitals.

# ABBREVIATIONS

| | |
|---|---|
| Å | angstrom |
| AIM | Atoms In Molecules |
| CHARMM | Chemistry at Harvard Macromolecular Mechanics |
| CSFF | Carbohydrate Solution Force Field |
| DFT | Density Functional Theory |
| DMDS | Dimethyl Disulphide |
| DMSO | Dimethyl Sulphoxide |
| DSDE | 1, 1'- disulphanediyl diethanol |
| ESAs | Electronic Structure Approaches |
| HF | Hartree-Fock |
| MD | Molecular Dynamics |
| MM | Molecular Mechanics |
| NBO | Natural Bond Orbital |
| NMR | Nuclear Magnetic Resonance |
| NOE | *Nuclear overhauser effect* |
| NPT | Isothermal-isobaric ensemble |
| ns | Nanosecond |
| PDF | Pair Distribution Function |
| ps | Picosecond |
| QM | Quantum Mechanics |
| SCF | Self Consistent Field |
| SDF | Solvent Distribution Function |
| $T_1$ | Longitudinal Spin Relaxation Time |
| $T_2$ | Transverse Spin Relaxation Time |

# CONTENTS

# Chapter 1

# Carbohydrates and Their Importance

## 1.1  Introduction

Carbohydrates are by far the most plentiful of all natural products and their variation in structure is best reflected in the vast biological roles they play. Although commonly known for their application as structural supports (cellulose) and as storage molecules (cyclodextrins). The involvement of oligosaccharides and glycoconjugates in important biological phenomena such as cell recognition, cell-cell adhesion, trafficking and immune response has come to the fore in the last decade.[1-3]

From a pharmacological point of view, carbohydrate research helps in the design of drugs that can disrupt the interaction of viruses, bacteria and parasites on a cell surface. This is done by binding preferentially to the various carbohydrate-binding proteins (lectins).[2, 4]

For these reasons the study of carbohydrates is of great importance and interest to the scientific community; from the design of the simplest biodegradable polymers to that of drug inhibitors. The common aim is the determination of the complex conformational preferences and the roles they play in the macroscopic properties of the overall molecule.[5, 6]
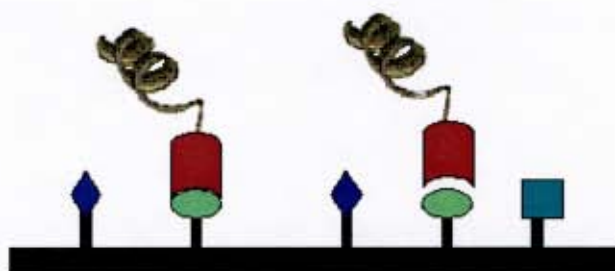


Figure 1.1 A simplified representation of cell recognition.

## 1.2  Carbohydrate Classification

Carbohydrates or saccharides are molecules primarily made up of carbon, hydrogen and oxygen atoms, with the general formula $C_xH_{2x}O_x$. Their classification is based on the ability of a particular saccharide to be hydrolysed, as well as on the number of saccharides linked together. This basic description leads to two groupings; simple and complex carbohydrates. Simple carbohydrates are monosaccharides that cannot be broken down to smaller carbohydrate units via any process, particularly hydrolysis. Complex carbohydrates on the other hand are larger and are usually found as polymers comprised of various monosaccharides. These carbohydrate polymers can readily undergo hydrolysis, as is required for their classification as complex carbohydrates.[5, 7]

### 1.2.1 Simple Carbohydrates

As stated previously, simple carbohydrates (monosaccharides) are the building blocks in the production of complex carbohydrates. Monosaccharides are subdivided into groups according to the number of carbon atoms they contain, as well as the position of the hydroxyl groups.

Conformationally, monosaccharides exist in equilibrium between their straight chain form and their closed-ring molecular structure with the equilibrium generally favouring the ring structure, as is the case of $\beta$-D-glucose in the aqueous phase, in which the cyclic form is found to exist at >99%.[8]

The straight chain form of the monosaccharides allows for greater flexibility due to single bond rotations between the carbon atoms. Upon this rotation the free aldehyde- or keto- group attached at carbon 1 (C1) undergoes a spontaneous hemiacetal formation with a hydroxyl group, leading to the ring structure, as can be seen in Figure 1.2.

The rings formed as a result usually consist of either five or six heavy atoms, depending on the point of attack. During this ring closure process two stereoisomers are formed as a result of the formation of a chiral centre at C1, called the anomeric carbon. These stereoisomers are designated as the α- and β-anomers.[7]

2
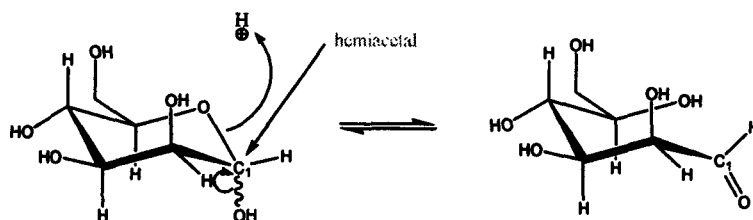
Figure 1.2 Spontaneous hemiacetal formation leading to the inter-conversion of a straight chained monosaccharide into the ring structure.

In addition, the ring closure of monosaccharides is an easily reversible process and for this reason inter-conversion of the different forms is found to occur at room temperature in an aqueous medium, as can be seen in Figure 1.3 which gives the possible inter-conversions for mannose.
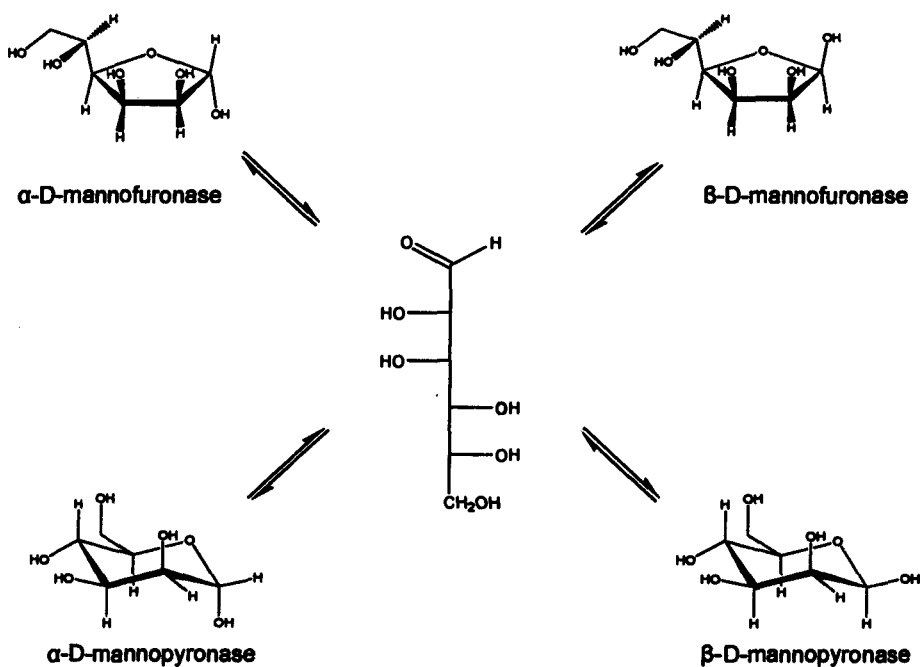


Figure 1.3 Inter-conversion of the α- and β-anomers of mannofuronase and mannopyronase via an open chain mannose monosaccharide.

3

## 1.2.2 Complex Carbohydrates

By definition the smallest complex carbohydrate is a disaccharide, which is formed by joining two monosaccharides via a condensation reaction. A new bond is created between the anomeric carbon of the first monosaccharide and any of the hydroxyl groups of the second monosaccharide. For example, a linkage from the $\beta$-anomeric C1 of a monosaccharide can connect to the hydroxyl groups on the C1, C2, C3, C4 or C6 of the second monosaccharide and in this manner five possible disaccharide linkages can be formed. These linkage for the above example would be denoted by the symbols: $\beta$(1 → X), where $\beta$ represents the orientation of the anomeric carbon, the '1' represents the carbon from which the linkage originated and the 'X' represents the carbon atom to which the anomeric centre C1 has become linked.

The linking of several monosaccharides leads to polysaccharides. Polysaccharides which consist of only a few monosaccharides (between six and ten) are known as oligosaccharides. Polysaccharides may either consist of linear or branched systems, in which the monosaccharides can occur in periodic sequences as in the case of starch and cellulose, or in interrupted sequences and aperiodic sequences.
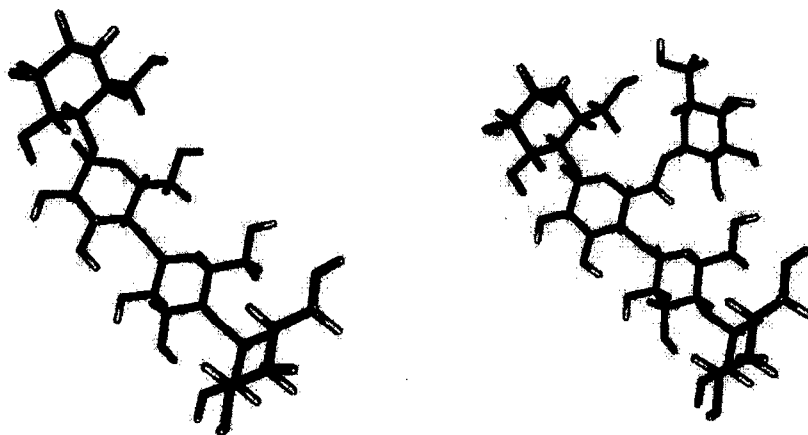


Figure 1.4 A segment for a linear (left) and branched (right) complex carbohydrate.

4

# 1.3 Three Bond Interglycosidic Linkages

Generally disaccharides, oligosaccharides and polysaccharides contain two-bond linkages (C-X-C). Where X = O, this linkage is more commonly known as an interglycosidic linkage. Introducing an extra atom into the above linkage to form a three-bond linkage (C-X-Y-C) is less common, and until recently the most noticeable of these three-bond linkages (where X,Y = O, C, N,), was found to occur in natural oligosaccharides which are incorporated in the antibiotics tunicamycin,[9] namenamicin[10] and sialyl $T_N$.[11]

Interest in these and other three-bond interglycosidic linkages has risen in recent times due to the potential use of carbohydrates containing these linkages in significant biological interactions, such as in cell recognition, cell proliferation and in carbohydrate metabolism.[12, 13]

Therefore, in order to expand the library of potential drug candidates, studies of novel carbohydrate mimetics are required, such as the study of the recently synthesised disulphide disaccharides,[14] where the three-bond linkage is made uniform by making X = Y = S, thus creating the disulphide linkage.

The disulphide linkage is primarily known for its role in the stabilisation of tertiary structures of proteins by linking two cysteine amino acid units. It is also prominent in the formation of cyclopeptides. The recent introduction of the disulphide unit as an interglycosidic linkage has resulted in glycosyl disulphides being identified as efficient glycosyl donors and as potential glycomimetics.[15]

The main characteristics of the three bond linkages that play a part in the above interactions are:

1. The altered electronic and steric properties of the linker atoms, due to the use of di-heteroatomic atoms.

2. An increased distance between glycosyl components in terms of the number of connecting bonds resulting in an extension of the available conformational space and the introduction of additional motional freedom.

3. The inherent chirality of the di-heteroatomic bond in sulphur based linkages.

As a result, the following disaccharide mimics were selected to be studied, they are: D-GlcNAc$p$-$\beta$S(1→1')S$\beta$-D-Glc$p$ (1), D-Man$p$-$\beta$S(1→1')S$\beta$-D-Glc$p$ (2) and D-Gal$p$-$\beta$S(1→1')S$\beta$-D-Glc$p$ (3) and are depicted below:
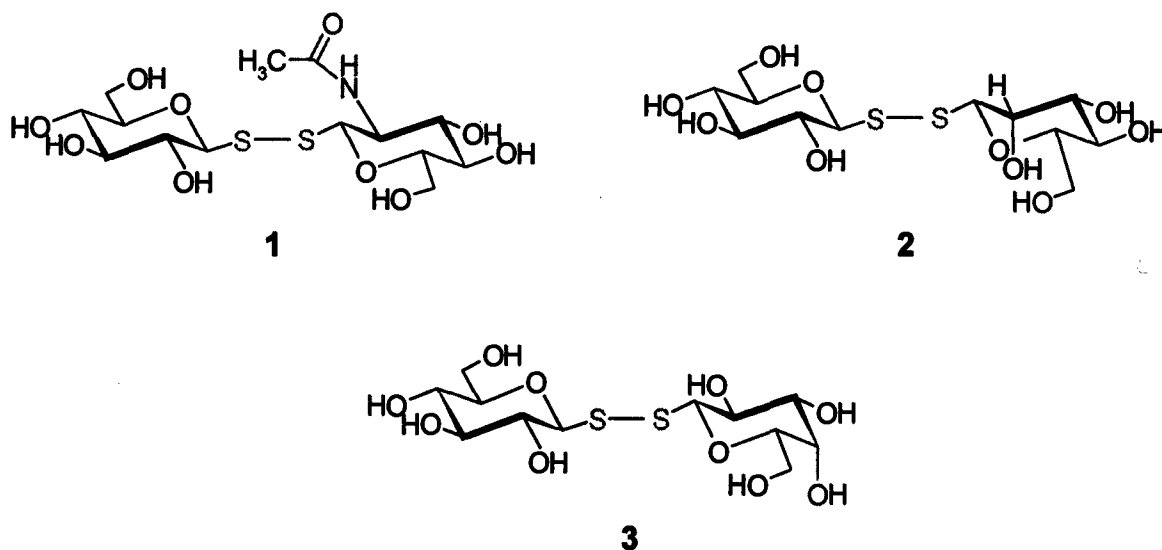


Figure1.5 Line diagrams of D-GlcNAc$p$-$\beta$S(1→1')S$\beta$-D-Glc$p$ (1), D-Man$p$-$\beta$S(1→1')S$\beta$ D-Glc$p$ (2) and D-Gal$p$-$\beta$S(1→1')S$\beta$-D-Glc$p$ (3)

The monosaccharide units contained in the above asymmetrical disulphide derivatives are common constituents of the carbohydrate chains in many glycoproteins and/or biologically active oligosaccharides. The different configurations and/or groups in one of the monosaccharide moieties in each of these disaccharides (axial OH groups in position 2 and 4, respectively, in 2 and 3 as well as the presence of a 2-N-acetyl group in 1) allow for the exploration of the influence of structural effects on the conformational preferences. For the purpose of simplification the disaccharide mimics have been renamed as ITZ55, ITZ56 and ITZ57 for mimics 1, 2 and 3 respectively.

6

# 1.4  Conformational Descriptors

Presented below is an outline of the conformational descriptors used for describing the conformations of disaccharides containing the $\beta S(1\rightarrow1')S$ linkage.

### 1.4.1 Interglycosidic Torsion Angles

It is commonly acknowledged that there is a direct correlation between the flexibility of a carbohydrate and the nature of interglycosidic linkages within these molecules. In solution the $^4C_1$ hexopyranose ring is found to be extremely rigid,[16] and therefore the main conformations of disaccharides and polysaccharides containing pyranose units are primarily based upon the interglycosidic linkages, as these constitute the majority of the principle degrees of freedom in the molecules. It is therefore imperative that the three torsion angles involved in the $\beta S(1\rightarrow1')S$ linkage are described here. They are; $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\Omega}$. A figurative representation is given in Figure 1.6.

$$\boldsymbol{\Phi} = H1 - C1 - S1 - S1'$$
$$\boldsymbol{\Psi} = C1 - S1 - S1' - C1'$$
$$\boldsymbol{\Omega} = S1 - S1' - C1' - H1'$$

These descriptors have independently been used by Andre et al. in a recent study of the lectin-binding properties of glycosyl disulphides.[17] However, the use of the oxygen atom or carbon atom linked to atom C1, (leading to the dihedrals O5 (C5) - C1 - S1 - S1' and S1 - S1' - C1' - O5' (C5')) rather than a hydrogen atom is permissible and may lead to a smaller standard deviation due to reduced mobility.
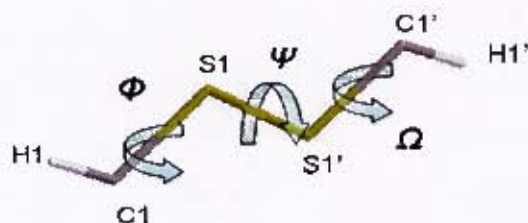
Figure 1.6 Definition of the torsion angles used to specify the $\beta S(1\rightarrow1')S$ interglycosidic linkage

### 1.4.2 Primary Alcohol Orientations

For a complete conformational analysis the introduction of a further degree of freedom is necessary. The orientation of the primary alcohol (the 6- hydroxymethyl group) in relation to the monosaccharide ring is thus introduced. This extra orientation aspect is defined by two dihedral angles:

$$\theta_1 = O5 - C5 - C6 - O6,$$
$$\theta_2 = C4 - C5 - C6 - O6$$

These dihedrals should be able to adopt any value between $-180°$ and $180°$, but in practice only one of three staggered orientations occur. These orientations are denoted by *tg*, *gg* and gt respectively, where *t* refers to trans and *g* refers to gauche.
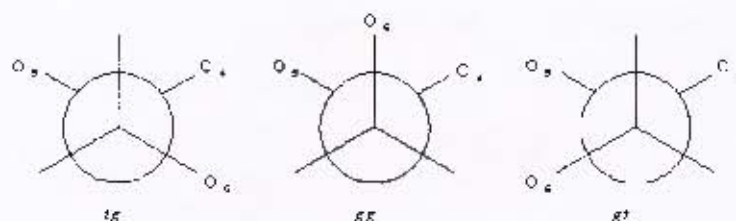


Figure 1.7    Primary alcohol positions in a monosaccharide ring

8

The orientation of *tg, gg* and *gt* are defined as follows: the first letter refers to the orientation of the first dihedral, $\theta_1$, and similarly, the second refers to the orientation of the second dihedral, $\theta_2$.

## 1.5 Environmental effects

In a study of the biophysical properties and functions of a molecule it is imperative that one obtains its molecular structure and properties in its natural environment. It is just as important to understand how alterations in this environment will affect the structure and properties of the molecule. These points are of particular importance in the study of carbohydrates, as slight alterations in environment, such as temperature changes, can cause large structural changes as a result of the increased mobility of the interglycosidic linkages.

The main environment in which carbohydrates exist and have been studied over the years is the aqueous environment, and it is therefore essential that we study carbohydrates in this environment in order to comprehensively understand their behaviour. Over the previous two decades it has been shown that interactions between carbohydrates and water, via the hydroxyl groups of the carbohydrate, play a vital role in structural transformations of the carbohydrates, thus resulting in numerous biological activities.[18] Carbohydrates have also been found to impart a structuring effect upon the solvent molecules around them by disrupting the hydrogen bonding pattern of the solvent, such as in the case with water.[19] Therefore a detailed understanding of the effect of solvents on the stability of the various conformations found to exist in solution is vital in the analysis of the structural properties of a carbohydrate system.

However, obtaining molecular structure data for carbohydrates in water via the experimental tool of choice, *Nuclear Magnetic Resonance* (NMR), is very difficult. This is primarily due to the phenomenon known as fast proton transfer, which occurs between the hydroxyl groups on the carbohydrates and the water molecules, leading to little structural data being obtained. This problem may be overcome by introducing a solvent which does not interfere with the NMR experiment, such as the aprotic *Dimethyl Sulphoxide* (DMSO) solvent system.[20] The drawback with DMSO is that no guarantee

can be given that the molecular structure in the DMSO solvent system corresponds with the structure in the water system, as has been shown for polysaccharides, via optical rotation experiments.[21]

# 1.6 Experimental Studies

There have been numerous attempts to study the structures of carbohydrates experimentally via methods such as NMR spectroscopy and X-ray diffraction. Below is an outline of the information obtained from these methods.

### 1.6.1 NMR Spectroscopy

NMR spectroscopy is regarded as one of the most valuable methods in obtaining experimental data regarding the conformations of molecules in solution. In the liquid phase the most useful NMR parameters directly related to molecular geometry are the *nuclear overhauser effect* (NOE) and scalar spin-spin coupling constants, reflecting the internuclear distances and torsion angles, respectively. As discussed in section 1.4.1, glycopyranosyl rings in disaccharides are found to remain in the $^4C_1$ chair configuration when in solution.

Therefore, conformational studies are focussed on the determination of the interglycosidic dihedral angles. In regular two-bond interglycosidic linked oligosaccharides the approach is based on the measurement of three-bond proton-carbon coupling constants across the glycosidic linkage (H1'- C1' – X – Cn and/or C1' – X – Cn – Hn; X usually being O, S or N) and estimating the torsion angles via a Karplus-type relationship.[22, 23] However, this approach cannot be applied in the case of disulphide three bond interglycosidic linkages because the analogous *four-bond* couplings ($^4J_{CH}$) tend to be very small.

An alternative approach to determining interglycosidic torsion angles in principle is to measure the three-bond carbon-carbon couplings across the S-S-linkage (C1-S-S'-C1'). Such measurements are technically difficult at natural $^{13}C$ abundance and although Karplus-type dependence of these couplings has been demonstrated both theoretically

and experimentally,[24] application to carbohydrates is still problematic because appropriate parameters are not yet available.[25, 26]

Conformational preferences of the hydroxyl groups as well as intra- and interresidue hydrogen bonding interactions constitute further important aspects of carbohydrate structures. Moreover, temperature dependences of the chemical shifts of OH and amide NH protons are indicative of the degree of solvent exposure, and hence of hydrogen bonding interactions. Rotamer populations of OH-groups are another source of information related to hydrogen bonding. This data can be estimated from measurements of the relevant proton-proton ($^3J_{OH, H}$) and proton-carbon ($^3J_{OH, C}$) scalar coupling constants.

### 1.6.2 X-Ray Diffraction

X-ray diffraction is the principal method used to analyse polysaccharide structures in the solid state. There are, however, two major drawbacks to this method. The first is as a result of the flexibility of non-cyclic and branched oligosaccharides, resulting in various conformations that can be adopted by these molecules, and leading to problems with crystallisation.[22, 27]

The second problem arises due to a single conformation being the result of an x-ray diffraction study. This leads to a lack of dynamic information regarding the motion of molecules being studied and it can not be guaranteed that the solid state structure is the same as the conformations adopted in the solution phase.

## 1.7 Computational Studies

In the last few decades advances in computational techniques have allowed for an extended dynamical study of carbohydrates.[28] This has allowed one to overcome the limitations inherent in the conventional experimental methods used to determine the solution and solid phase structures of carbohydrates.

Of the numerous computational techniques available the one most frequently used for the study of carbohydrates has been Molecular Dynamic (MD) simulations, which has

been used to corroborate data obtained from NMR experiments.[29] Allowing for an insight into the conformational preferences of numerous saccharides and their glycosidic linkages, such as the study of trimannosides[30] and the modeling of the $\alpha(1\rightarrow6)$ branch point of Amylopectin.[31]

Quantum Mechanic (QM) studies of saccharides have mainly been limited to monosaccharides such as glucose and fructose.[32, 33] This is mainly due to limited computational power and therefore an increase in the time required for the study of larger systems of structural complexity such as oligosaccharides. More recently the structure and energies of numerous disaccharide conformers are being studied using QM methods.[34]

The recent introduction of the Quantum Mechanics/Molecular Mechanics (QM/MM) hybrid [35] has brought about a method that can be used to analyse monosaccharides[36] and disaccharides,[37] in which their electrostatic potentials are expressed by the QM section of the hybrid and therefore allows for the study of properties that have not previously been studied via MM studies, such as polarisation.[37]

# 1.8 Objectives

The principal objective of this thesis is to obtain an understanding of the conformations and dynamics of several disaccharide mimetics containing a sulphur-sulphur $(\beta S(1\rightarrow1')S)$ three bond interglycosidic linkage.

To obtain this information it is necessary to first determine the parameters for the $\beta S(1\rightarrow1')S$ linkage by performing *ab initio* and density functional theory (DFT) calculations on several fragments containing an unconstrained sulphur-sulphur linkage and progressing to a complete disaccharide. These parameters are to be appended to the Carbohydrate Solution Force Field (CSFF)[38] to allow for MD simulations to be run in vacuum, with explicit inclusion of water and with explicit inclusion of DMSO.

For the MD simulations both the time evolution and relaxation behaviour of the interglycosidic torsional angles (principal degrees of freedom) can be calculated. This is to be complemented with the experimental NMR data in order to obtain a representative

comparison of the structural preferences of diglycosyl disulphides and conclusions are to be drawn regarding the effect of solvation in both water and DMSO.

A study of the stereoelectronic effects associated with the sulphur-sulphur linkage is to be conducted using the Atoms in Molecule theory developed by Bader.[39]

# References

1.  Bertozzi, C. R.; Kiessling L.L. *Science* **2001**, *291*, 2357.

2.  Lis, H.; Sharon, N. *Chem. Rev.* **1998**, *98*, 637.

3.  Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683.

4.  Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321.

5.  French, A. D. *Carbohydrate*, In *Encyclopedia of Computational Chemistry*; 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, 1998; pp 209.

6.  Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. *J. Am. Chem. Soc.* **2001**, *123*, 12238.

7.  Mathews, C. K.; van Holde, K. E. *Biochemistry*; The Benjamin & Cummings Pub. Co.; Redwood City, 1991.

8.  Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. *Organic Chemistry*, Oxford University Press, Great Clarendon street, Oxford OX2 6DP, 2001; pp 146.

9.  Eckhardt, K. *J. Nat. Prod.*, **1983**, *46*, 544.

10. Weinstein, D. S.; Nicolaou, K.C. *J. Chem. Soc., Perkin Trans. 1* **1999**, 545.

11. Teo, C.-F.; Hwang, T.-S.; Chen, P.-H.; Hung, C.-H; Gao, H.-S.; Chang, L.-S.; Lin, C.-H. *Adv. Synth. Catal.*, **2005**, *347*, 967.

12. Macindoe, W. M.; van Oijen, A. H.; Boons, G. J. *J. Chem. Soc. Commun.* **1998**, 847.

13. Davis, B.G.; Lloyd, R. C.; Jones, L. B. *J. Org. Chem.* **1998**, *63*, 9614.

14. Szilágyi, L.; Illyés, T.-Z., Herczegh, P. *Tetrahedron Letters* **2001**, *42*, 3901.

15. Sears, P.; Wong, C.-H. *Angew. Chem., Int. Ed.* **1999**, *38*, 2300.

16. Rees, D. A. *Polysaccharide Shapes*. Outline Studies in Biology. Chapman and Hall Ltd., 11 New Fetter Lane, London EC4P 4EE, 1977.

17. André, S.; Pei, Z.; Siebert, H.-C.; Ramström, O.; Gabius, H.-J. *Bioorganic & Medicinal Chemistry* **2006**, *14*, 6314.

18. Brady, J. W. Solvation: Carbohydrates. In *Encyclopedia of Computational Chemistry*; 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, 1998; pp 2609.

19. Liu, Q.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 12276.

20. Casu, B.; Reggiani, M.; Gallo, G. G.; Vigevani, A. *Tetrahedron*, **1966**, *22*, 3061.

21. Rees, D. A; Thorn, D. *J. C. S. Perkin II*, **1977**, 191.

22. Wormald, M. R.; Petrescu, A. J.; Pao, Y. -L. P.; Glithero, A.; Elliot, T.; Dwek, R. A. *Chem. Rev.*, **2002**, *102*, 371.

23. Tvaroska, I.; Taravel, F.R. *Adv. Carbohydr. Chem. Biochem.*, **1995**, *51*, 15.

24. Krivdin, L. B.; Della, E. W. *Progr. NMR Spectrosc.*, **1991**, *23*, 364.

25. Xu, Q.; Bush, C. A. *Carbohydr. Res.*, **1998**, *306*, 335.

26. Thibaudeau, C.; Stenutz, R.; Hertz, B.; Klepach, T.; Zhao, S.; Wu, Q.; Carmichael, I.; Serianni, A. S., *J. Am. Chem. Soc.*, **2004**, *126(48)*, 15668.

27. Jeffrey, G. A.; Takagi, S. *Accounts of Chemical Research*, **1978**, 264.

28. Short, P. L.; Mullin, R.; Tullo, A. H.; Wilson, E. K. *Chemical and Engineering News*, **2004**, *82(17)*, 36.

29. Dyekjaer, J. D.; Rasmussen, K. *Mini Reviews in Medicinal Chemistry*, **2003**, *3*, 713.

30. Sayers, E. W.; Prestegard, J. H. *Biophysical Journal*, **2000**, *79*, 3313.

31. Best, R. B.; Jackson, G. E.; Naidoo, K. *J. J. Phys. Chem. B*, **2002**, *106*, 5091.

32. Cramer, C. J.; Truhler, D. G. *J. Am. Chem. Soc.*, **1993**, *115*, 5745.

33. Csonka, G. I. *J. Mol. Structure (Theochem)*, **2002**, *584*, 1.

34. Momany, F. A.; Willet, J. L. *J. Comp. Chem.*, **2000**, *21*, 1204.

35. Field, M.J.; Bash, P. A.; Karplus, M. *J. Comp. Chem.*, **1990**, *11*, 700.

36. Marilia, T. C.; Martins, C. *Carbohydr. Res.*, **2005**, *340*, 2185.

37. Muslim, A.-M.; Bryce, R. A. *Chem. Phys. Lett.*, **2004**, *388*, 473.

38. Kuttel, M.M; Brady, J. W.; Naidoo, K. J. *J. Comp. Chem.*, **2002**, *23*, 1236.

39. Bader, R. F. W. *Acc. Chem. Res.*, **1985**, *18*, 9.

# Chapter 2

# Computer Simulations: Quantum Methods

## 2.1 General Overview

Computational chemistry is a set of methodologies employed to study chemical problems with the aid of a computer. These chemical problems range from the basic (e.g. analysis of the conformation of a molecule) to the more intricate (e.g. the interaction of a substrate with an enzyme). Over time, the need to solve these varied problems has lead to the development of numerous methods. These methods can generally be grouped into five broad classes[1] and are briefly discussed below.

The first of the five groups is Molecular Mechanics (MM). In this model the atoms are represented by balls and the bonds by springs. Reference bond angles' lengths are set up from experimental data and any deviations from these reference values are translated into energy penalties using a mathematical representation. From these energy penalties an overall energy for a molecule (system) can be calculated, enabling the lowest energy conformation (geometry optimisation) to be found.

MM is usually used in conjunction with the second group, Molecular Dynamics (MD), whereby Newtonian laws of motion are applied to molecules. This allows one to simulate the motion of a molecule. These two categories are also known as empirical force field methods[1,2,4] and shall be discussed further in section 2.3.

The three remaining categories are more commonly known as the electronic structure approaches (ESAs).[1-4] Two of the categories, *ab initio* and semi-empirical (SE) method, are based on solving the wavefunction for the 'electronic' Schrödinger equation, which describes how the electrons behave in a molecule. Theoretically, these methods can be used to calculate anything about the molecule, for any molecule.

The remaining category is known as density functional calculations. This model is better known as density functional theory (DFT) calculations. DFT is similar to *ab initio* and semi-empirical methods. It too is based on solving Schrödinger's equation. Unlike

the previous two models, DFT derives the electron distribution (electron density) directly, rather than calculating the wavefunction.

The five groups above are tabulated below in Table 2.1. This is done in order to categorise them in terms of classical and quantum mechanics, enabling the models to be viewed together in terms of their general applications and speeds, in order to give an idea of which model is viable depending on the molecule or system that is to be studied.

| | Classical | Quantum Mechanics | | |
| --- | --- | --- | --- | --- |
| | MM + MD | *ab initio* | SE | DFT |
| Theory | 1. Uses a ball and spring model<br>2. Uses Newtonian laws of motion | 1.Solves H $\Psi$=E $\Psi$<br>2. Uses 1st principles | 1.Solves H $\Psi$=E $\Psi$<br>2. Theory and experimental | 1.Solves H$\Psi$=E$\Psi$ |
| Calculates | Energy from change in geometry | $\Psi$, which is used to calculate the electron distribution | $\Psi$, which is used to calculate the electron distribution | Calculates electron distribution, without 1st calculating $\Psi$ |
| Time, relative to MM[1,4] | 1 | 10000-100000 | 100 | 100-10000 |
| Uses[1,4] | Large molecules, such as proteins and DNA, etc. | Novel molecules and active sites of enzymes (as part of QM/MM study) | Moderately large systems | Novel molecules |
| Limitations[1,4] | Parameterisations only reliable for particular molecule | Computationally intensive | Parameterisations only reliable for particular molecule | Computationally intensive and hence limits system size |

Table 2.1 Summary of Classical and Quantum Mechanical Simulation Methods

## 2.2 Quantum Chemistry Methods

Quantum chemistry methods or ESAs are methods based upon calculating the electron distribution attributed to a fixed set of nuclear positions (as determined by the Born-Oppenheimer (BO) or "clamped nuclei" approximation). This is done by solving the Schrödinger equation with the aid of necessary simplifications and approximations. With the wavefunction $\Psi$ (from Schrödinger equation) obtained, it is possible to calculate the energy and other properties of a molecular system, by performing operations upon the wavefunction.

In this thesis the primary ESAs applied has been the DFT method, which requires the assistance of a limited number of physical constants, such as the speed of light, $c$, and Planks constant, $h$.[3]

## 2.2.1 The Born-Oppenheimer Approximation

Prior to solving Schrödinger's equation, the Born-Oppenheimer (BO) approximation must be applied. This approximation allows the separate treatment of electron and nuclei motions and can be validated in terms of the electrons moving faster than the nuclei with respect to the atom's centre of mass. This results in an average electron distribution being realised by the nuclei, whereas the electrons can react at once to any change in the nuclear arrangement.

From the BO approximation it is shown that the electronic wavefunction depends not upon the momenta of the nuclei, but only upon the relative nuclei positions. [1-4] The total wavefunction for a molecule can be written as follows:

$$\Psi_{tot}(\text{nuclei, electrons}) = \Psi(\text{electrons})\Psi(\text{nuclei}) \qquad (2.1)$$

Looking at the total energy of the above system simplistically, it is seen that the energy can be broken up into two terms; namely the nuclear energy term representing electrostatic repulsion between positively charged nuclei, and the electronic energy term consisting of the kinetic and potential energies of the electrons moving in the electrostatic field of the nuclei, coupled with the electron-electron repulsion. The total energy can now be represented by:

$$E_{tot} = E\,(\text{electrons}) + E\,(\text{nuclei}) \qquad (2.2)$$

Thus, applying the BO approximation to the above energy sum, with the nuclei fixed, the emphasis is placed on electronic motions only, thereby simplifying the calculation of the Schrödinger equation.[1-4]

18

## 2.2.2 The Schrödinger Equation

Any introduction to quantum mechanics must surely lie with the Schrödinger equation. This equation allows for the derivation of the energy of a system as a function of its electronic properties.[1,3] The mathematical representation of the time-dependent form of the Schrödinger equation, as denoted in equation 2.3, gives reference to a single particle with mass, $m$, moving through space, $p$, and time, $t$, with the particle under the influence of an external field $V$.

$$\left\{ \frac{-\hbar^2}{8\pi^2 m} \nabla^2 + V \right\} \psi(p,t) = \frac{i\hbar}{2\pi} \frac{\partial \psi(p,t)}{\partial t} \tag{2.3}$$

The wavefunction $\Psi$ characterises a particle's motion and by performing an operation upon it via an operator we can obtain various properties for the particle.[1-4]

If we now look at the situation where $V$ is independent of time, it is possible to rewrite the Schrödinger equation such that:

$$\left\{ \frac{-\hbar^2}{8\pi^2 m} \nabla^2 + V \right\} \psi(p) = \frac{i\hbar}{2\pi} \frac{\partial \psi(p)}{\partial t} \tag{2.4}$$

This simplification is achieved by writing the wave function as the product of its spatial and time components:

$$\psi(p,t) = \psi(p)\tau(t) \tag{2.5}$$

The time-independent equation can be further simplified to its more familiar form:

$$\hat{H}\psi(p) = E\psi(p) \tag{2.6}$$

By using the Hamiltonian operator $\hat{H}$, where $\hat{H}$ is derived from equation 2.7:

19

$$\hat{H} = \frac{-\hbar}{2m} \nabla^2 + V \qquad (2.7)$$

In solving the Schrödinger equation it is necessary to ascertain values of E and functions $\Psi$ as to return multiples of the wavefunction and the energy, when the wavefunction is operated on by the Hamiltonian.

Difficulties arising in solving the Schrödinger have necessitated simplifications and as such several approximations have been introduced. These approximations can be subdivided into two groups; the first being the choice of electron functions to adequately represent the electronic wavefunction and the second being the choice of basis functions to represent the one-electron functions (molecular orbitals).[1-4]

### 2.2.3 The Hartree Self Consistent Field (SCF) Theory

The Hartree-Fock (HF) calculation is seen as the most basic type of *ab initio* calculation.[1] The HF calculation is based on the HF approximation, which allows for a practical study of molecular systems via QM calculations. This is achieved by removing the many electron problem and replacing it with a one electron problem in which electron-electron repulsions are treated in an average manner by making use of the self-consistent-field (SCF) method in solving the HF equation.[5] Upon solving the HF equation a Hartree wavefunction is obtained. The Hartree wavefunction, $\Psi_{HP}$, can be defined as the product of all the individual electronic functions associated with the orbitals, [1, 3] as in equation 2.8, and is known as the Hartree product.

$$\Psi_{HP}(r_1, r_2, \ldots, r_n) = \Phi_1(r_1)\Phi_2(r_2)\ldots\ldots\Phi_n(r_n) \qquad (2.8)$$

The SCF method is an iterative method in which an initial guess for the spin orbitals is made, from which the average field seen by an electron can be calculated, as well as the corresponding eigenvalue equation. This equation is then used to calculate a new set of spin orbitals. This process is repeated until the spin orbitals from the previous iteration used to construct the Fock operator are the same as the resulting eigenfunctions.

In 1930, Fock and Slater indicated that $\Psi_{HP}$ was incorrect and corrected it such that $\Psi$ is asymmetric to electron switching, thus obeying the Pauli principle.[6, 7]

## 2.2.4 Molecular Orbital Theory

In QM, the behaviour of a single electron is mathematically represented by a three-dimensional atomic orbital.[1, 3] Molecular Orbital (MO) theory is the process whereby these three-dimensional atomic orbitals are linearly combined to make up the system's wavefunction, $\Psi_i$, such that:

$$\Psi_i = \sum_{\mu} x_{\mu}(r)C_{\mu i} \tag{2.9}$$

This linear combination of atomic orbitals (LCAOs) is more commonly referred to as a basis function or basis set. When implemented, these basis sets will treat each electron independently and assign to each electron its own electronic function.[3] Basis sets are discussed further in section 2.2.7.

## 2.2.5 Slater Determinants to the Roothaan-Hall Equations

As described above, the Hartree wavefunction is a product of orbitals (spatial orbitals, which are a function of the x, y, z space coordinates) and, when taking electron switching into account, results in a symmetric wavefunction and non-compliance with Pauli's principle, which states that "no two electrons may have the same set of four quantum numbers". As such, an electron pair must consist of one electron in the spin up (alpha,$\alpha$) state and one electron in the spin down (beta,$\beta$) state.[1, 3]

Fock showed that an anti-symmetric Hartree product was obtainable by means of adding and subtracting the permutations of the Hartree product. Slater on the other hand introduced a Slater wavefunction consisting of a spin orbital, $\Psi$ (spin), which is the product of a spatial orbital and a spin function ($\alpha$ or$\beta$), such that;

$$\Psi(\text{spin}\alpha) = \Psi(\text{spatial})\alpha \qquad (2.10)$$

$$\Psi(\text{spin}\beta) = \Psi(\text{spatial})\beta \qquad (2.11)$$

The Slater wavefunction also differs from the Hartree function in that it is a determinant matrix and not the simple product of one electron functions.[1] This matrix is more commonly known as the Slater matrix determinant.

After determination of a wave function it is important to note that the variation principle must be taken into account. The variation theorem states that the energy calculated by equation 2.12 using an approximate wavefunction must always be greater than or equal to the real ground-state energy of the molecule.[1]

$$E = \left\langle \Psi \middle| \hat{H} \middle| \Psi \right\rangle \qquad (2.12)$$

Thus, if you calculate the HF energy of a system, the energy calculated is dependent on how correct the wavefunction is and the more correct the wavefunction the closer the calculated energy will be to the system.

As they stand, the HF equations are of little use for molecular calculations as they are not mathematically viable in guessing the initial MO wavefunction (which is required to start an iterative process) and as a result of the complexity of the wavefunctions to be calculated.

Roothaan and Hall independently solved these shortcomings of the HF equations by proposing that the molecular orbitals be represented by linear combinations of their basis functions (Atomic orbitals),[8, 9] which leads to the following coefficient expansion equation:

$$FC = SCe \qquad (2.13)$$

where F is the Fock matrix,[10] C is the molecular orbital expansion coefficient related to the LCAOs and e is the diagonal matrix of the orbital energies.

## 2.2.6 Closed and Open Shell Methods

Open shell molecules are made up of atoms with unpaired electrons, such as free radicals. The first approach in handling open shell systems is known as the spin-restricted Hartree-Fock (RHF) theory (closed shell), in which combinations of molecular orbitals have either singly or doubly occupied orbitals. This method is thus restrictive and leads to inaccurate results.

The more commonly method used is the application of the spin-unrestricted Hartree-Fock (UHF) theory (open shell) developed by Pople and Nesbet in 1958.[10] UHF theory utilises two sets of molecular orbitals; one for alpha-spin electrons and the other for beta-spin electrons (both spin sets are derived from the Slater determinants), thus solving eigenvalue equations for both the alpha- and beta-spin electrons and this is found to reproduce open-shell attributes more closely than RHF theory.

Other methods that can be applied to open shell molecules and other problems are the restricted open-shell Hartree-Fock (ROHF) method and the generalised valence bond[11] (GVB) method, which are both considered open shell. The ROHF method combines attributes of both the RHF and UHF methods in that it uses the same orbitals for different spin states, but uses different occupations, whereas the GVB method uses doubly and singly occupied orbitals.

## 2.2.7 Electron Correlation Methods

The main failure of HF theory is the inadequate manner in which it represents electron correlation. Using the SCF method, electrons are assumed to be moving in an average potential of all the other electrons, which leads to the neglect of the effects the neighbouring electrons have on the instantaneous positions.[3]

In reality electrons try to avoid interactions between each other and from this it is expected that lower energies would result. In order to overcome this problem of electron correlation representation various methods are employed, which differ from open shell methods in the approximations made and the representation of the wavefunction.

The more common of these methods are Møller-Plesset (MP) perturbation theory, Configuration Interaction (CI), Coupled Clusters (CC) and DFT.[1-4]

### 2.2.7.1 Møller-Plesset Perturbation Theory

The Møller-Plesset (MP) method[12] is one of the more popular electron correlation methods, in which the electron correlation is treated as a perturbation on the HF problem, such that the Hamiltonian operator $\hat{H}$ is expressed as the sum of a zeroth order Hamiltonian, $H_0$, and a perturbation, V:

$$\hat{H} = H_0 + \lambda V \qquad (2.14)$$

This treatment of the HF problem thus allows for a description of both open- and closed-shell systems based on an unrestricted HF wavefunction.[13, 14]

The theory is based on the proposal that if one has acceptable solutions to problems, answers to similar problems can be calculated. To make this easier, the simplest $H_0$ values must be used and if the results are not acceptable the differences between the Hamiltonians must be solved.

In the MP setup, the energy and the wavefunction are expanded in a power series of the perturbation and since the HF energy is correct to the first order, the perturbation energies only start contributing from the second order.[15] This gives rise to the MP$n$ notation for energies to the given order, $n$.[15] Even though MP theory over corrects for HF, the appealing part of the theory over both CI and CC methods is the low computational cost involved.[16] Of the common MP perturbations the second order MP2 is most commonly used when compared to its higher order 'brothers', MP3 and MP4, due to its low computational cost and relatively accurate results.[16]

## 2.2.7.2 Density Functional Theory

A method initially developed by Hohnberg and Kohn for the study of solids and later extended by Kohn and Sham, DFT is a post HF method, but with one significant difference.[15, 16] In DFT the main emphasis is placed upon solving the electronic charge density, $\rho$ directly rather than the many-electron wavefunction.

This is achieved by writing the total energy as a functional of the density, $\rho$, and breaking up the electronic energy into several terms:

$$E_{DFT}[\rho] = E^T[\rho] + E^V[\rho] + E^J[\rho] + E^{XC}[\rho] \qquad (2.15)$$

Where $E^T$ is the kinetic energy term, derived from the motion of the electrons. $E^V$ contains terms for the description of the nuclear-electron attraction, as well as the repulsion between nuclei pairs. $E^J$ is the electron-electron repulsion term and $E^{XC}$ is the exchange-correlation term that includes the remaining section of the electron-electron interactions. Of all the terms above only the nuclear-nuclear repulsion is not a functional of $\rho$. In DFT the exchange-correlation energy functional has to be approximated, as the exact mathematical form is not known. This can lead to inaccuracies in predicting certain properties of a system, such as the binding energies.[16, 17]

To overcome this problem, numerous exchange-correlation functionals have been developed. One of the most frequently used functionals for molecular systems and the one used in this thesis is the Becke-three-Lee-Yang-Parr (B3LYP) hybrid functional.[18] These particular functionals fall under the category of *generalised gradient approximations* (GGAs), which has been shown to give better accuracy in predicting chemical properties and computing the density fluctuations via the gradient of the densities.[5]

## 2.2.8 Basis Sets

A basis set is a mathematical description of the LCAO's making up a system. When the orbitals are combined, they give an approximation of the total electronic wavefunction. The purpose of a basis set is to define the possible positions of electrons in space and the larger the basis set, the more accurate it is found to be.

### 2.2.8.1 Minimal Basis Sets

These basis sets contain the smallest number of basis functions required for each atom and make use of fixed-size atomic orbitals. The Slater-type orbitals (STOs), which represent the electron density in the valence region, are the smallest possible basis sets available. The increased speed given by STOs in the evaluation of orbital integrals is offset by the inability to obtain accurate results when dealing with complex systems.

To overcome this problem, Gaussian-type orbital functions are introduced. This inclusion leads to the STO-nG basis sets and increases accuracy in calculations. The only limit is that the maximum number of Gaussian-type orbitals that may be added is six.

The additional auxiliary functions used to improve the accuracy in the solutions to a simulation can be expressed as:

$$g(\mu, p) = cx^i y^j z^k e^{-\mu p^2} \tag{2.16}$$

where $p$ is composed of x, y and z coordinates and $\mu$ is a constant, used to determine the radial extent of the function.

### 2.2.8.2 Split-Valence Basis Sets

The initial step in improving a basis set is to make it larger. This is achieved by increasing the number of basis functions per atom. The split-valence basis set adjusts the minimal basis set by introducing two orbitals of different sizes, one large component and one small component, respectively.

The most common of the above basis sets is the Gaussian-type orbital functions, 3-21G and 6-21G. Similarly, the double zeta basis sets, such as those devised by Dunning and Huzinaga (D95), introduce the same effect. By introducing the split-valence basis sets one can obtain more accurate fits to the molecular condition.[2, 3]

### 2.2.8.3 Polarised Basis Sets

In order to remove the limitation of an orbital's shape, the polarised basis set adds orbitals with angular momentum beyond that which is required for a description of the groundstate of each atom.[3]

This addition allows for the study of molecules containing heavy atoms, by distorting the orbital shapes by mixing orbitals with higher valence shell symmetry. Polarisation functions, as with split-valence functions, are readily added to Gaussian-type orbitals and can yield greater accuracy with calculations on charged systems.

The common basis sets containing polarisation are the 6-31G (d) and 6-31G (d, p), which have added d and p functions, respectively.[3]

### 2.2.8.4 Diffuse Functions

Diffuse functions introduce oversized s- and p-type functions, allowing orbitals to occupy a larger region of space. The introduction of diffuse functions allows for added accuracy in the study of systems, in which the electrons are relatively far from the nucleus, e.g. molecules that contain lone pairs, molecular systems that contain negative charges and molecular systems with low ionisation potentials.[2, 3]

### 2.2.9 Geometry Optimisation

Geometry optimisation is the process used to find both local and global minima on a potential energy surface (PES) plot. The PES plot is a figurative representation of the energies of a molecular system, in which each point gives a particular energy corresponding to a structure or orientation of the molecular system.[1-4]

27

In order to study the properties of a molecular system a 'stable' wavefunction is required. ESAs obtain this 'stable' wavefunction by utilising a numerical minimisation algorithm and the QM theory, resulting in a 'optimised' geometry.

The word 'optimised' is used with some apprehension, as the 'optimised' geometry obtained is not necessarily a minima on the PES. Performing a frequency calculation on the 'optimised' geometry and comparing the frequencies to experimental data can ascertain this information. There is a further possibility of calculating imaginary frequencies, which also give non-minima.

The existence of negative frequencies is also possible and indicates that a re-optimisation should be performed using different convergence criteria.[3]
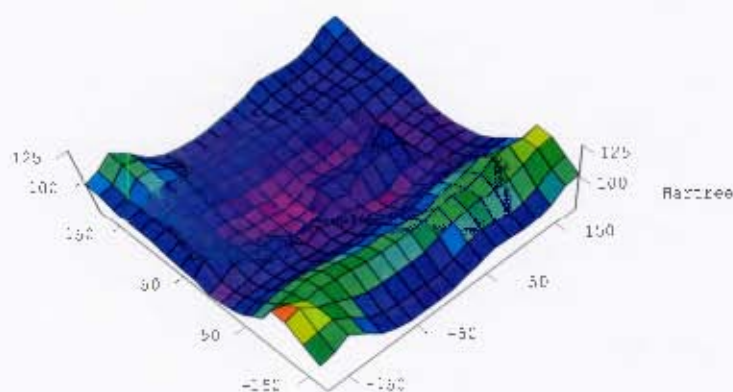


Figure 2.1 A potential energy surface (PES)

## 2.2.10 Atoms In Molecules Theory

The Atoms In Molecules (AIM) method of R.F. W. Bader[19] was developed based upon the quantum mechanical postulate that every property of a chemical system is contained in a wavefunction, $\Psi$. This implies that more than just the energies and the corresponding geometries of a molecular system are obtainable, but that $\Psi$ contains the information regarding the atoms in a molecule and their respective connectivity, along with the electronic charge distribution (electron density) of a molecule.[19]

The AIM method is designed to allow for the analysis of the topology (minima, maxima and saddle points) of the electron density as a function of three spatial coordinates, by dividing a molecular volume into atomic subspaces.[20] Where the electron density is $\Psi^2$ integrated over n-1 coordinates, as shown in equation 2.17 below

$$\rho(r_1) = \int |\Psi(r_2, r_3, ..., r_n)|^2 dr_2 dr_3 ... dr_n \qquad (2.17)$$

Maxima in the electron density generally indicate positively charged nuclei and are referred to as nuclear attractors, designated by (3,-3) as they draw electron density towards themselves, as can be seen by the direction of the gradient of the electron densities towards the strongest attractor. Dividing the physical space into atomic subspace is achieved by taking infinitesimal steps in the gradient direction leading to the various attractors. Thus forming atomic basins associated with the atomic nuclei.

Bond critical points (3,-1) are determined by searching between the various nuclear attractors (3,-3). Bond paths are then constructed by following the eigenvectors corresponding to the negative eigenvalues at specific bond critical points. The collection of these eigenvectors form a line of maximum increase in electron density starting from the (3,-1) critical points leading to (3,-3) nuclear attractors.

Constructing contour plots of the Laplacian of the electron density, $\nabla^2 \rho(r)$ gives an indication of the areas of local charge concentration ($\nabla^2 \rho(r) < 0$) and local charge depletion ($\nabla^2 \rho(r) > 0$). From these it is possible to determine the sharing of the electron density, as required in this work.

## 2.2.11 Natural Bond Orbital Theory

Natural Bond Orbital (NBO) analysis which is based on local block eigenvectors of one-particle density matrices, was originally designed for the study of the covalency and hybridisation effects in polyatomic wavefunctions.[21]

NBO analysis is made up of sequential transformations from the starting basis set assigned to a molecule to numerous localised basis sets (natural atomic orbitals (NAOs)),

to natural hybrid orbitals (NHOs), natural bond orbitals (NBOs) and natural localised orbitals (NLMOs). From which the NLMOs can be further transformed to delocalised natural orbitals (DNOs) or canonical molecular orbitals (CMOs).[22-23]

$$AOs \rightarrow NAOs \rightarrow NHOs \rightarrow NBOs \rightarrow NLMOs \rightarrow DNOs \rightarrow CMOs \qquad (2.18)$$

In this thesis the primary objective behind using the NBO analysis method developed by Weinhold et al.[22] is to obtain the atomic orbitals shapes for a pseudo $\beta S(1 \rightarrow 1)$'S interglycosidic linkage within its molecular environment and thus derive the molecular bonds present.
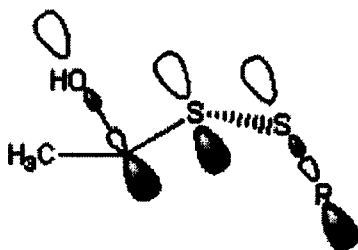


Figure 2.2 An illustration of potential stabilising orbital interactions present within the –g conformation of the pseudo $\beta S(1 \rightarrow 1)$'S linkage.

# References

1. Lewars, E. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Kluwer Academic Publishers, Second printing, **2004**.

2. Leach, A. R. *Molecular Modelling: Principles and Applications*, 1$^{st}$ ed.; Addison Wesley Longman Limited: Singapore: **1996**.

3. Foresman, J. B.; Frisch, A. E. *Exploring Chemistry with Electronic Structure Methods*, 2$^{nd}$ ed.; Gaussian, Inc.: Pittsburgh, **1996**.

4. Leach, A. R. *Molecular Modelling: Principles and Applications*, 3$^{rd}$ ed.; Harlow, England; New York: Prentice Hall, **2001**.

5. Gill, P. M. W. In *Encyclopedia of Computational Chemistry*, 1$^{st}$ ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, **1998**; pp 678.

6. Fock, V. *Z. Physik*, **1930**, *61*, 126.

7. Slater, J. C. *Phys. Rev.*, **1930**, *35*, 120.

8. Roothaan, C. C. J. *Rev. Mod. Phys.*, **1951**, *23*, 69.

9. Hall, G. G. *Proc. Roy. Soc. (London)*, **1951**, *A205*, 541.

10. Pople, J. A. ; Nesbet, R. K. *J. Chem. Phys.*, **1954**, *22*, 571.

11. Goddard, W. A.; Dunning, T. H.; Hunt, W. J.; Hey, P. J. *Acc. Chem. Res.*, **1973**, *6*, 368.

12. Møller, C.; Plesset, M.S. *Phys. Rev.*, **1934**, *46*, 618.

13. Krishnan, R.; Frisch, M. J.; Pople, J. A. *J. Chem. Phys.*, **1980**, *72*, 4244.

14. Bartlett, R. J; Shawitt, I. *Chem. Phys. Lett.*, **1977**, *50*, 90.

15. Raghavachari, K.; Anderson, J. B. *J. Phys. Chem.*, **1996**, *100*, 12960.

16. Chen, J. Y. –J., *PhD Thesis*, Chemistry Department; University of Cape Town, Cape Town, **2004**.

17. Jones, R. O. *Adv. Chem. Phys.*, **1987**, *67*, 413.

18. Lee, C.; Yang, W. Y.; Parr, R.G. *Phys. Rev. B*, **1988**, *37*, 785.

19. Bader, R. F. W. *Acc. Chem. Res.*, **1985**, *18*, 9.

20. Jensen, F. *Introduction to Computational Chemistry*, John Wiley & Sons Ltd., **1999**.

21. Foster, J. P; Weinhold, F. *J. Am. Chem. Soc.*, **1980**, *102*, 7211.

22. Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.*, **1988**, *88*, 899.

23. Weinhold, F.; Landis, C. R. *Chemistry Education: Research and Practice in Europe*, **2001**, *2*, 91.

# Chapter 3

# Computer Simulations: Empirical Methods

## 3.1 Empirical Force Field Models

The need to study large systems such as carbohydrates, proteins and enzymes is of great importance for biomedical applications. The biggest concern facing the computational chemist is how to computationally approach these exceedingly large systems. Using quantum mechanical methods (solving the Schrödinger equation) would require a substantial computational effort due to the vibrant nature of biomolecules when immersed in a condensed phase, such as water, and for this reason such QM methods are not feasible.

To overcome this, empirical methods, more commonly referred to as MM, MD and Monte Carlo (MC) dynamics, are applied. These methods use energetic penalising functions to take into account the interactions and motions experienced by a molecular system.[1-3]

### 3.1.1 Force Fields

With the validity of the Born-Oppenheimer approximation (Section 2.2.1), it is possible to approach atomistic problems by employing classical empirical potentials known as *force fields*, by writing the energy of the electrons, E(electrons), as a parametric function of the nuclear coordinates, thus avoiding the necessity of solving the Schrödinger equation.

Most of the molecular modeling force fields in use today consist of two terms, namely bonded and non-bonded, which can be further broken down into various potential functions, such as a bond stretching function ($E_b$), an angle bending function ($E_\theta$), a dihedral function ($E_\phi$), an improper-dihedral function ($E_\omega$) and several non-bonding interaction functions ($E_{vdw} + E_{el}$).[1-3] These potential functions lead to a simplified

33

mathematical representation of a system's energy, as shown below in equation 3.1, as formulated in the CHARMM force field.[4]

$$E = E_b + E_\theta + E_\phi + E_\omega + E_{VdWaal} + E_{elec} \qquad (3.1)$$

Harmonic potentials (Hooke's Law) can be applied to the bond stretching, angle bending and improper-dihedral functions. This is allowed in molecular mechanics because the atoms and the bonds between the atoms are seen to behave as a ball and spring system and under the provision that the deviations from the equilibrium values are small. The functional forms of these potentials are:[1-3]

- Bond stretching

$$E_b = \sum K_b (r - r_o)^2 \qquad (3.2)$$

The bond stretching potential is the sum of the energies generated by the vibrational motion of the bond between two atoms relative to the natural bond length, $r_o$, calculated over all bonds. The constant, $K_b$, is the force constant, which constrains this vibrational motion of two atoms relative to one another, as a spring would do between two spheres. The deviation in bond length ($r - r_o$) is in place to allow for a change in bond length due to the effects of all other atoms within a particular calculation and using Hooke's law it is seen that the energy sum is only valid for small deviations in bond length.

- Bond angle

$$E_\theta = \sum K_\theta (\theta - \theta_o)^2 \qquad (3.3)$$

The deviation of bond angles from their reference angles ($\theta_o$) can also be shown to follow a harmonic pattern and is described using the above function. The force constant, $K_\theta$, is in this case smaller than the force constant for bond stretching, as the energy required to distort an angle is much less than that required to compress or stretch a bond.

- Improper-dihedral

$$E_{\omega} = \sum K_{\omega} \left( \omega - \omega_o \right)^2 \tag{3.4}$$

The need to maintain local geometry, such as is required in a tetrahedral centre or in a planar aromatic ring, requires an additional term to be incorporated into the force field. This is achieved with the use of an *out-of-plane* bending term. This involves the calculation of the angle between a bond from the central atom and the plane defined by the central atom and the other two atoms, with a value of $0^o$ corresponding to all four atoms being coplanar. If this approach is employed a harmonic potential can be used, where the reference angle is defined by $\omega_o$ and the force constant by $K_{\omega}$.

The dihedral function, $E_{\phi}$, is employed to describe the barriers to rotation about chemical bonds and plays the biggest part in the molecular flexibility of a molecule. It uses energy penalties determined by the torsional angle between the 1, 4 atoms in a system consisting of atoms A-B-C-D, with the lowest energy being derived from a staggered conformation and the highest from an eclipsed conformation of atoms A and D. This can be described using a cosine series expansion such as the one below:

$$E_{\phi} = \sum K_{\phi} \left[ 1 + \cos \left( n\phi - \gamma \right) \right] \tag{3.5}$$

This cosine series expansion allows the desired energy profile to be obtained. The important constants in the above equation are n, which represents the multiplicity, whose value gives the number of minimum points in the function as the bond B-C is rotated through $360^0$, and $\gamma$ (the phase factor), which represents the position the torsion angle adopts at its minimum energy. The force constant, $K_{\phi}$, determines how easily the bond can be rotated and has higher value for a double bond than a single bond.

The final terms, $E_{VdW} + E_{el}$, from equation 2.7 are referred to as non-bonded terms and these energy terms constitute the majority of time required for a computational study. In the CHARMM program, these intermolecular interactions are represented by Coulombic and Lennard-Jones (12-6) pairwise interactions. The importance of these

35

interactions can be seen when modeling a solute interacting with a solvent, which is imperative when comparing experimental data to computational data. The functional forms of these potentials are:

- Van der Waals interactions:

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

(3.6)

The van der Waals terms are calculated using a standard Lennard-Jones curve. This is an example of an "effective pair potential", which can imitate a potential with higher order character by means of an efficient pairwise potential.[1-3] The A and B parameters in the equation above are determined by experimental data and include parameters $\sigma_{ij}$ and $\varepsilon_{ij}$, which represent the collision diameter and well depth for the interaction of atoms i and j respectively. These are described mathematically as:

$$A = 4\sigma^{12}_{ij}\varepsilon_{ij}$$

(3.7)

$$B = 4\sigma^{6}_{ij}\varepsilon_{ij}$$

(3.8)

- Electrostatic interactions:

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

(3.9)

The electrostatic potential is introduced into the overall energy calculation to account for the unequal charge distribution throughout a molecule due to the differing electronegativities of elements constituting the make up of the molecule. In equation 3.9 above, $\varepsilon_0$ represents the effective dielectric constant for the medium and r is the distance between the two point charges.

36

The distinct need for a uniform force field can be attributed to the vast number of varying force fields that are in use today, such as the CHARMM,[4] AMBER[5] and GROMOS[6] force fields. Each force field calculates the energy potential in a different manner, using the same basic potential terms as described above, but altering the way in which parts of the empirical energy function are used. An example of this is the manner in which some force fields incorporate special hydrogen bonding terms, whereas other force fields can model hydrogen bonding adequately by using the Van der Waals interactions and electrostatic potentials as described above.[7]

# 3.2 Force Field Models for Simulations

After establishing the basis of a MM force field as a combination of bonded and non-bonded terms, which is derived from ideal bond angles and geometries, as well as experimental and theoretical constants, it becomes imperative that force field parameterisation can take place in order to accurately reproduce the properties for any system of atoms we wish to study. The only draw back to force field studies is the empirical potential functions.[1] As a result, transferability of individual parameters between force fields is not consistent in terms of results and is dependent on the force fields being used.[8] In this thesis an emphasis was placed on the carbohydrate, water and DMSO models, which are briefly described below.

### 3.2.1 Carbohydrate Models

Numerous force field and parameter sets exist for the modeling of carbohydrates. Early attempts at carbohydrate specific force fields, such as the Hard Sphere Exo-Anomeric approach (HSEA),[9] were accurate in terms of helping to interpret NMR data. Their major drawback was the inability to perform energy minimisations and introduce the flexibility inherent in monomeric saccharides.

The more advanced force fields available were originally developed for protein studies, but have all incorporated carbohydrate parameter sets, such as the force fields

incorporated in CHARMM,[10-13] AMBER[14-18] and GROMOS.[19-21] As this study primarily uses the CHARMM force fields, these will be discussed further.

The initial CHARMM carbohydrate parameter set, developed by Ha et al., was the HGFB force field,[10] appropriately named after the authors. It was developed in order to fit the available structural and experimental data for the *alpha-D-glc-p*. Since then, a CHARMM parameter set has been developed by Reiling et al. by making use of *ab initio* calculations on small representative fragments of various carbohydrates.[11]

Further development lead to the PHLB force field (abbreviated from the authors names) developed by Palma et al.,[12] which replaces the general dihedral angle terms in the HGFB force field with n-fold cosine terms (usually limited to three), used to identify specific dihedrals.

The work in this thesis was performed using the adjusted PHLB force field, which adds improved dynamic behaviour of the hydroxyl methyl groups. This force field, known as the Carbohydrate Solution Force Field (CSFF), was developed by Naidoo et al.[13]

### 3.2.2 Water Models

The interest in water from both a basic and scientific perspective is driven by our need to understand how it sustains life, as well as its fascinating array of properties as both a solvent and a pure substance.

Water's complexity is believed to be as a result of strong, highly directional (linear) hydrogen bonds, which construct a local tetrahedral structure around each water atom and result in an ordered network within the liquid. Bearing this in mind, one is lead to believe that to accurately model water such that all its properties are taken into account, quantum methods should be employed. This is not the case, as simple water models, such as the TIP3P[22] and SPC/E[23] models, which are used for biomolecular simulations, are found to be the best.

The main concern in modeling water is the computational efficiency with regards to energy calculations, as a large number of water molecules are usually under consideration together with a solute. To overcome this time constraint simple models

(TIP3P and SPC) are used. These models make use of effective pairwise potentials with no explicit three body terms or polarisation effects.

Due to the large number of water models in use today, a classification is required and therefore the models are divided into the following three categories:

1. Simple interaction-site models, in which the water molecule is held in a rigid geometry and all interactions are described by pairwise Coulombic and Lennard-Jones expressions,

2. Flexible models which allow for internal changes in the conformation of the water molecule, and

3. Models which explicitly include polarisation and many-body effects.

In 'simple' water models, a rigid geometry and between three and five interaction sites is the order of the day. The classical examples of the TIP3P and SPC models both use three sites for their electrostatic interactions and place partial positive charges on the hydrogen atoms, which are exactly balanced with a negative charge located on the relevant oxygen atom. The four site models, TIP4P[22] and BF,[24] are similar to the three site models, but have the negative charge displaced from the oxygen atom to a point along the HOH angle bisector in the direction of the hydrogens. The five site models, such as the ST2[25] and TIP5P[26] potential model, have their charges on the hydrogens and on two lone-pair sites on the oxygen.

Attempts to develop flexible and polarised models[27, 28] have lead to a slight improvement of reproducing experimental results such as density heat capacity and pair distributions, but the computational time required with these models still has to be reduced to make them viable for study of large systems.

For the scope of this work only 'simple' water models were considered, in particular the TIP3P model, as the CHARMM protein and carbohydrate force fields have been parameterised with the use of this model . Even though the SPC, ST2, TIPS2[22] and TIP4P models are found to give better structural and thermodynamic descriptions of liquid water,[22] the TIP3P model has been shown to adequately predict water structuring around carbohydrates.[29]

39

### 3.2.3 DMSO Models

The importance of DMSO in the study of biological macromolecules requires an adequate force field representation for theoretical studies to be conducted using molecular dynamics simulations. DMSO consists of two hydrophobic methyl groups and a highly polar sulphur-oxygen group; these are arranged in a pyramidal shape, with the oxygen atom, sulphur atom and the two methyl groups positioned at the apices.[30]

Early studies of DMSO used a rigid, united-atom model in which each methyl unit is represented by a single atom center and all three bond lengths for the molecule are held at their equilibrium bond lengths.[30, 31] The respective charge distributions for the various atoms were all derived from *ab initio* QM calculations conducted by Rao and Singh.[32]

Recent work by Strader and Feller has given rise to a flexible all-atom DMSO model, which can be freely used in the CHARMM parameter set.[33] This model was developed using the parameter development procedures on which the CHARMM biomolecular force field is based.

The Strader and Feller model was adopted in this study to compare computational results to the corresponding NMR results for the three diglycosyl disaccharide mimics, which have been calculated in DMSO solvent.

## 3.3 Dynamic Simulation Methods

MD is used to determine system properties such as conformational fluctuations and solvent interactions, by reproducing the 'real' dynamics of a particular system. There are two main dynamics simulation methods, namely, MD and Stochastic Dynamics (SD). Even though these methods use different equations of motion, MD, is used as a general term to refer to either technique.[1-3]

### 3.3.1 Molecular Dynamics

In the MD approach the relative atomic positions of all atoms in the system are derived in sequence via application of Newton's equations of motion.

$$\frac{d^2 r_j(t)}{dt^2} = \frac{F_j}{m_j}$$

(3.10)

$$F_j = \frac{-\partial V(r_j,...r_n)}{\partial r_j}$$

(3.11)

In the above equations, the force exerted on an atom j is given by $F_j$. This is equal to the negative gradient of the potential energy, $V(r)$. This indicates that the force is a differentiable function of the atomic coordinates, $r_j$.

MD methodology is advantageous in research as it is deterministic. This means that the state of a system at any future time can be estimated from its current state. The disadvantage with deterministic methods is the increase in time required for more frequent integration steps (shorter steps ~1-10fs). These are required to obtain more realistic potentials over a time frame relevant to a system being analysed.

### 3.3.2 Stochastic Dynamics

In SD, the trajectory of a system is calculated by the integration of the Langevin equation of motion rather that Newton's equations of motion. The Langevin equation of motion is shown below:

$$\frac{d^2 r_j(t)}{dt^2} = \frac{F_j}{m_j} + \frac{R_j}{M_j} - \gamma_j \frac{dr_j(t)}{dt}$$

(3.12)

41

This equation has two extra terms added to equation (3.10): a stochastic force, $R_j$ that models random collisions with solvent molecules and a frictional drag force (proportional to $\gamma_j$) that dampens the solute's motion through the solvent. The stochastic term introduces energy and the frictional term removes energy from the system.[1]

### 3.3.3 Integration of Equations of Motion

Numerous methods exist for the integration of the equations of motion. The most problematic part is thus choosing a suitable method. The answer to this problem comes from the needs of the overall problem, such as the computational effort required (the number of integrations per time step required) and if energy conservation is required in the MD simulation being run.

The most common method in use is the finite difference method, which is most suited for systems under the influence of a continuous potential (energy conservation). This technique generates MD trajectories which are pairwise additive, the most widely used being the Verlet algorithm. This algorithm uses the positions, $p$, accelerations at time $t$ and the positions from the previous step, $p(t - \partial t)$, to calculate the new positions at $(t + \partial t)$ and $p(t + \partial t)$, as shown below:

$$p(t + \partial t) = 2p(t) - p(t - \partial t) + \partial t^2 a(t) \qquad (3.13)$$

The only problem with the Verlet algorithm is the error that is incurred with the addition of a small acceleration term, $\partial t^2 a(t)$ to a much larger position term. To overcome this, the leap-frog variation of the Verlet algorithm is used.[1] To implement this algorithm the velocities $v\left(t + \frac{1}{2}\partial t\right)$ are calculated from the velocities at time $v\left(t - \frac{1}{2}\partial t\right)$ and the accelerations at time $t$. The positions, $p(t + \partial t)$, are then calculated from the velocities. The leap-frog Verlet explicitly includes the velocity and does not need to calculate the differences of large numbers. The only drawback to this algorithm is that the positions and velocities are not in sync.[1, 34]

$$p(t + \partial t) = p(t) + \partial t v \left( t + \frac{1}{2} \partial t \right) \quad (3.14)$$

$$v \left( t + \frac{1}{2} \partial t \right) = v \left( t - \frac{1}{2} \partial t \right) + \partial t a(t) \quad (3.15)$$

### 3.3.4 Simulation Ensembles

The results of a simulation are required to be comparable with those of a physical experiment.[1-3] For this to be achieved various simulation ensembles are used. Traditionally, MD is run using a microcanonical ensemble (NVE), in which the conditions kept constant are the number of particles, N, the volume, V, and the energy, E.

When compared with experiment, it is found that other ensembles give more correct results. Common alternatives are the isothermal-isobaric ensemble (NPT) where the number of particles, N, the temperature, T and the pressure, P are kept constant or the canonical ensemble (NVT) in which the number of particles, N the temperature, T and volume, V are kept constant.

The easier of the two ensembles to implement is surely the NVT, or canonical ensemble, where the temperature is held constant by rescaling the velocities of the atoms at regular intervals or by implementing a heat bath coupled to the system.

For this thesis, the isothermal-isobaric ensemble, corresponding to the Gibbs free energy, was used. The temperature was held constant by coupling a heat bath to the system, by making use of the Nose-Hoover thermostat method,[35] and the pressure was maintained using a Langevin Piston.[36]

### 3.3.5 Periodic Boundary Conditions

Periodic boundary conditions allow for the study of the macroscopic properties of a molecule using a small number of particles. In other words, the bulk properties of a molecule or solvent can be simulated without the problem of energy fluctuations caused by surface molecule interactions.[1-3]

The atoms of a system are enclosed in a box that is space-filling and the box is usually in the shape of a cube or truncated octahedron. The box is replicated using a rigid translation in all three Cartesian directions so that when a particle moves in the central box its images in the other boxes will move in an identical manner. This is also valid when an atom leaves the central box and simultaneously an atom from an adjacent box enters the central box with an identical velocity, as shown in Figure 3.1.

Each particle $n$ can then not only interact with other particles in the box, but also with a replicated image of itself in another box adjacent to the box it is contained in, thus eliminating surface effects.

## 3.3.6 Truncating the Potential and the Minimum Image Convention

The most time consuming part of any MD simulation is the calculation of the non-bonded energies. This is due to the number of non-bonded terms for a pairwise model, as used in the CHARMM empirical function, which is of the order $N^2$, where N is the number of atoms in the system. In order to reduce the time taken, simple approximations have been applied, which limit the calculation of non-bonded interactions to a particular region of space. These approximations or methods are known as the minimum image convention and the non-bonded cutoff limits.

The minimum image convention comes about as a result of the periodic boundary conditions. In the minimum image convention, each atom 'sees', at most, one representation of every other atom in the system and is infinitely repeated via the periodic boundaries. Thus, the energy for a particular atom is calculated with the closest atom or its image, as represented in the Figure 3.1 by the circle. The application of a cutoff is to implement the effect that all atom pair interactions further apart than the cutoff value are set to zero. For this to work effectively the cutoff value is set to half the length of the simulation box. This is done in order to avoid interactions between identical particles in neighbouring boxes.

The disadvantage arising from the implementation of a cutoff distance is that the cutoff creates a discontinuity in the potential function at that particular cutoff distance, thus creating an artificial increase in the kinetic energy of the system and leading to

fluctuations in the system's temperature. This problem is overcome by the implementation of smoothing functions, where the *shifting* and *switching* functions are commonly used.[1-3]
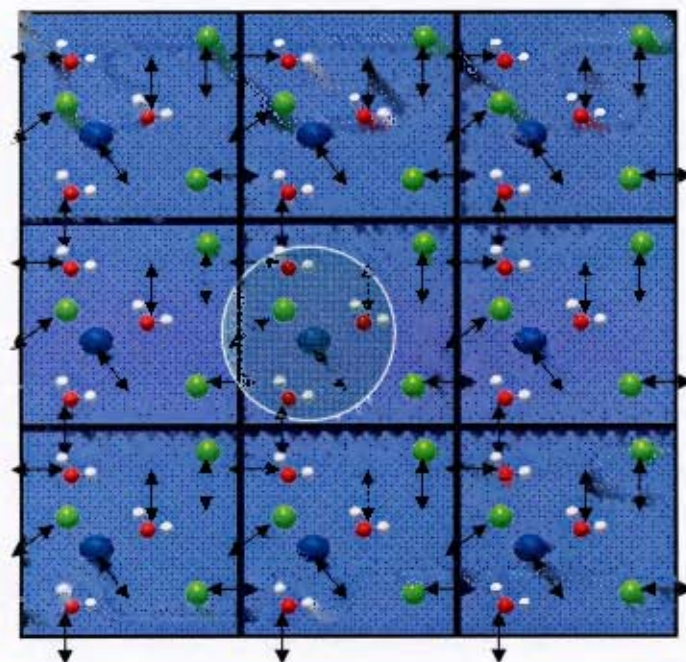


Figure 3.1 A two-dimensional representation of periodic boundary conditions and the minimum image convention with a spherical cutoff

# 3.4 Simulation Analysis Techniques

### 3.4.1 Conformational Analysis

### 3.4.1.1 Pucker Analysis

As with cyclohexane, the closed pyranosyl form of glucose is a six-membered ring and as such can adopt the same conformations, being the boat, the chair and the twist. The most convenient means in which these conformational changes can be analysed is by using the generalised puckering parameters developed by Cremer and Pople.[37]

Experimental evidence from NMR and crystallography has shown that the only significant glucopyranosyl conformer in solution is the $^4C_1$ chair. The analysis of simulation results using the protocols set up by Cremer and Pople allows for the simulated conformations to be checked.

### 3.4.1.2 Cluster Analysis

After completion of a molecular dynamics simulation, it is necessary to classify the various conformers generated according to their structured similarities. This classification of conformers allows for an easier study of the results and focuses the analysis on the important conformers.

Today there are a large number of clustering algorithms available. For this study the ART2[38] adaptive analogue pattern recognition scheme, which is provided in CHARMM, was used. This algorithm clusters time series data as defined by the user and is restricted to the motional and rotational degrees of freedom of the system being studied.

The main constraint is the cluster radius and the assignment of conformations to a cluster is therefore dependent on this radius, such that no member of a cluster is further than the specified distance from the cluster's center. This optimisation is carried out as an iterative refinement procedure that minimises the Euclidean distance between the central cluster and its members.

## 3.4.2 Structural Analysis

### 3.4.2.1 Pair Distribution Functions

The most useful way of describing the structure of a system, particularly liquids, is the pair distribution function (PDF), $g(r_{ij})$. The function $g(r_{ij})$ is defined as the probability of finding two atoms (or molecules) a distance $r_{ij}$ apart, when compared to a completely random distribution at the same distance, r, and is given by the equation:

$$g(r_{ij}) = \frac{1}{4\pi r_{ij}^2} \frac{dN(r_{ij})}{dr_{ij}}$$
(3.16)

The PDF for a liquid is commonly found as an intermediate between its gas and crystalline form and from the ensuing plot it is possible to obtain the nearest neighbours' positions relative to the atom in question, as well as the number of atoms making up the solvation shell.

Apart from the structural information obtained from a PDF, thermodynamic properties can be calculated, since it gives information about the pair distributions within the system studied.[1]

### 3.4.2.2 Spatial Distribution Function

Even though we have information about the number of neighbours for each site and the respective distance from the site, from the pair distribution function, we would like to have a more accurate method of determining the location of a solvent molecule around the solute, which would give a better indication of their role in solvation. For this we require a three-dimensional solvent probability density matrix. The procedure to obtain the three-dimensional probability density is as follows:[29, 39]

1) The translational and rotational diffusion of the solute molecules throughout the simulation are removed in order to match specific conformations, resulting in conformational averaging effects being removed.

2) It is important to run appropriately long MD simulations, so as to generate sufficient frames to obtain a satisfactory solvent density map around a particular conformation. Similarly, equal numbers of frames must be used when comparing the solvent densities around different conformations, thereby reducing the chances of the densities becoming blurred from a large number of frames introducing added motion.

3) For each solvent molecule the site of interest is used to calculate the densities. This site is 'binned' into a three-dimensional grid. A Gaussian function is then centered on each solvent's site of interest and used to determine the distribution of electrons for the site. The form of this function is:

$$G(r_i) = elec \times \left(\frac{a}{\pi}\right)^{\frac{3}{2}} \times e^{-ar_i^2} \qquad (3.17)$$

Where $elec$ is the number of electrons for the site of interest, while $a$ is chosen so that the function drops to 10% of its maximum value at the site of interest's Van der Waal radius.

4) The final solvent density matrix is normalised analogous to the one-dimensional case using the following function:

$$dens_{norm}(i,j,k) = dens(i,j,k) \times \frac{n_{xbin} \times n_{ybin} \times n_{zbin}}{n_{electrons} \times n_{atom} \times n_{frames}} \qquad (3.18)$$

Where $n_{xbin}$, $n_{ybin}$ and $n_{zbin}$ represent the number of divisions in the x, y and z directions of the simulation box respectively.

48

### 3.4.2.3 Hydrogen Bonding

Carbohydrates have the ability to partake in hydrogen bonding due to several hydroxyl groups contained in the molecule. Hydrogen bonds formed across the glycosidic linkage are very important as they contribute to the macromolecular structure and binding of polysaccharides and oligosaccharides.[40-42]

Hydrogen bonds formed for each hydroxyl in solution (aqueous) can be broadly classified into the following categories: those bonded to the solute, those bonded to the solvent and those involved in so called 'bridging' solvent molecules found between two solute hydroxyls. Bridges formed between adjacent hydroxyl groups in carbohydrates are found to be most abundant, even though interresidue dihedrals have been observed.[1,27]

Ideally, the energy used to define whether or not there is a hydrogen bonding interaction is the total non-bonded energy between the hydrogen bonding pairs. Another important factor using the non-bonded energy is that there is no distortion of the hydrogen bonds when looking at the various geometric considerations.[43]

A second method has been proposed, which redefines the energy function for the purpose of analysis such that the carbon of the C-OH pair has a neutral charge and the hydrogen bond is redefined as existing when a rolling average of the pair energy between the C-OH and the solvent molecule drops below a critical value, $E_{HB}$.[29]

For this thesis, a distance criterion for the acceptor hydrogen and an angle criterion of 2.4 Å and 120° respectively, were applied for hydrogen bonding analysis with the angle criterion relaxed to 100° for the intermolecular hydrogen bonding. The lifetimes of the hydrogen bonds were calculated as an average between breakages, leading to a shorter lifetime than the energetic definition due to the dynamic simulations writing structure information every 50fs.

### 3.4.3 Time-Dependent Properties

### 3.4.3.1 Correlation Functions

The correlation of two sets of data values, x and y, is often measured by the correlation function:

$$C_{xy} = \frac{1}{M}\sum_{i=1}^{M} x_i y_i \equiv \langle x_i y_i \rangle \qquad (3.19)$$

where M represents the number of values of $x_i$ and $y_i$ in the data sets. When this correlation function is normalised by dividing the root-mean-square values of x and y the result gives an indication of the level of correlation. A value of zero indicates no correlation and an absolute value of 1 indicates a high level of correlation.

Taking this concept further, an MD simulation gives the data values at specific times and with this the value of a particular property at a particular time may be correlated with the value of the same or another property at a later time. The resulting values are known as time correlation coefficients and the correlation function is written as:

$$C_\psi(t) = \langle \psi(t)\psi(0) \rangle \qquad (3.20)$$

If one looks at a particular correlation, such as ψ for the disulphide disaccharides, the decay of the correlation can be assumed to be exponential, with the general form $Ae^{-\frac{t}{\tau}}$, where τ is the correlation time. This correlation time gives an indication of how quickly the variable becomes uncorrelated as well as any regular motion that is obtained from the peaks that show up later in the functions.

To obtain more information from the frequency of the regular motion, it is possible to calculate the power spectrum of the correlation function by fourier transformation:

$$I(\psi) = \int_0^\infty \cos\psi t \langle \psi_i \psi_j \rangle dt \qquad (3.21)$$

from the peaks that are formed in the spectrum it is possible to determine what motions are responsible for the decay as well as their frequencies.

### 3.4.3.2 Internal and Overall Orientational Relaxation

The use of an appropriate correlation function, such as the angular velocity autocorrelation function, leads to approximate correlation times that describe the rotation of a molecule as a whole, as well as its internal fluctuations. The information obtained from these forms of the orientational correlation functions can be compared with experimental NMR and Raman spectra.[29, 43-45]

Various functions could be used to analyse the tumbling rate of a molecule, but the one specifically relevant to NMR relaxation is:

$$C(t) = \left\langle P_2\left(\overrightarrow{\mu_{mol}}(0) \bullet \overrightarrow{\mu_{mol}}(t)\right)\right\rangle \tag{3.22}$$

# References

1. Leach, A. R. *Molecular Modelling: Principles and Applications*, 3$^{rd}$ ed.; Harlow, England; New York: Prentice Hall, **2001**.

2. Lewars, E. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*, Kluwer Academic Publishers, Second printing, **2004**.

3. Grant, G. H.; Richards, G. W. *Computational Chemistry*, Oxford University Press: Oxford, **1995**.

4. Brooks, B. R.; Bruccoleri, R. E; Olafson, B. D.; States, D. J; Swaminathan, S.; Karplus, M. *J. Comp. Chem.*, **1983**, *4(2)*, 187.

5. Weiner, P. K.; Kollman, P. A. *J. Comp. Chem.*, **1981**, *2*, 287.

6. van Gunsteren *GROMOS: Groningen molecular simulation program package.* Technical report, University of Groningen, The Netherlands, **1987**.

7. van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem. Int. Ed. Engl.*, **1990**, *29*, 992.

8. Mackerell, JR. A.D *J. Comp. Chem.*, **2004**, *25(13)*, 1584.

9. Lemieux, R. U.; Koto, S.; Voisin, D. *The Exo-Anomeric Effect. The Origin and Consequences of the Anomeric Effect.*, ACS Symposium Series, **1979**, 17.

10. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carb. Res.*, **1988**, *180*, 207.

11. Reiling, S.; Schlenkrich, M.; Brickmann, J. *J. Comp. Chem.*, **1996**, *17*, 450.

12. Palma, R.; Himmel, M. E.; Liang, G.; Brady, J. W. In *ACS Symposium Series: Glycosyl Hydrolases in Biomass Conversion*, Himmel, M. E., Ed.; American Chemical Society: Washington, DC., **2002**; 112.

13. Kuttel, M.M; Brady, J. W.; Naidoo, K. J. *J. Comp. Chem.*, **2002**, *23*, 1236.

14. Woods, R. J.; Dwek, R. A.; Edge, C. J. *J. Phys. Chem.*, **1995**, *99*, 3832.

15. Senderowitz, H.; Parish, C.; Still, W. C. *J. Am. Chem. Soc.*, **1996**, *118*, 2078.

16. Senderowitz, H.; Still, W. C. *J. Org. Chem.*, **1997**, *62*, 1427.

17. Gregurick, S. K.; Liu, J. H. –Y.; Brant, D. A.; Gerber, R. B. *J. Phys. Chem. B*, **1999**, *103*, 3476.

18. Momany, F. A; Willet, J. L. *Carbohydr. Res.*, **2000**, *326*, 194.

19. Ott. K.; Meyer, B. *J. Comp. Chem.*, **1996**, *17*, 1084.

20. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comp. Chem.*, **1997**, *18*, 1955.

21. Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. *J. Comp. Chem.*, **2002**, *23*, 1416.

22. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D; Impey, R. W.; Klein M. L. *J. Chem. Phys.*, **1983**, *79*, 926.

23. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.*, **1987**, *91*, 6269.

24. Bernal, J. D.; Fowler, R. H. *J. Chem. Phys.*, **1933**, *1*, 515.

25. Stillinger, F. H.; Rahman, A. *J. Chem. Phys.*, **1974**, *60*, 1545.

26. Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.*, **2000**, *112*, 2000.

27. Niesar, U.; Clementi, G.; Kneller, G. R.; Bhattacharya, D. K. *J. Phys. Chem.*, **1990**, *94*, 7949.

28. Ren, P.; Ponder, J. W. *J. Phys. Chem. B*, **2003**, *107*, 5933.

29. Best, R. *MSc Thesis*, Chemistry Department; University of Cape Town, Cape Town, **2000**.

30. Vishnyakov, A.; Lyubartsov, A. P.; Laaksonen, A. *J. Phys. Chem. A*, **2001**, *105*, 1702.

31. Luzar, A.; Soper, A. K.; Chandler, D. *J. Chem. Phys.*, **1993**, *99*, 6836.

32. Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.*, **1990**, *112*, 3803.

33. Strader, M. L.; Feller, S. E. *J. Phys. Chem. A*, **2002**, *106*, 1074.

34. Thompson, D. L. In *Encyclopedia of Computational Chemistry*; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, **1998**, 2609.

35. Evans, D. J.; Holian, B. L. *J. Chem. Phys.*, **1985**, *83*, 4069.

36. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.*, **1995**, *103*, 4613.

37. Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.*, **1975**, *97*, 1354.

38. Carpenter, G. A.; Grossberg, S. *Appl. Opt.*, **1987**, *26*, 4919.

39. Kuttel, M. M. *MSc Thesis*, Chemistry Department; University of Cape Town, Cape Town, **2000**.

40. Rees, D. A. *Polysaccharide Shapes*, Chapman and Hall Ltd: NY, **1977**.

41. Sharon, N.; Lis, H. *Scientific American* **1993**, 82.

42. Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G.-J. *J. Am. Chem. Soc.* **2001**, *123*, 12238.

43. Chen, J. Y. –J., *PhD Thesis*, Chemistry Department; University of Cape Town, Cape Town, **2004**.

44. Best, R. B.; Jackson, G. E.; Naidoo, K. J. *J. Phys. Chem. B*, **2002**, *106*, 5091.

45. You, L.; Liu, Q.; Shi, Y.; Wang, C. –X.; Lahaye, M.; Tran, V. *Chem. Phys.*, **1997**, *224*, 81.

# Chapter 4

# NMR Spectral Parameters and Relaxation

## 4.1 Introduction

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 molecular geometry are the NOEs, *rotating nuclear Overhauser effects* (ROEs) and scalar spin-spin coupling constants reflecting, respectively, internuclear distances and torsion angles.[1, 2] In conjunction with this data, NMR allows for a closer look into the local dynamics around a given heteronucleus. This is obtained from the longitudinal and transverse relaxation times, $T_1$ and $T_2$, which are a representation of the motional characteristic of the vector X-H between a heteronucleus, X and its proton, as described by the model-free formalism of Lipari and Szabo.[3, 4] This formalism is used to obtain motional information from relaxation parameters by assigning an order parameter $S^2$ to each nucleus. This allows for the related data to be obtained from an MD simulation, as will be described later in the chapter.

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 $B_1$ is then observed as a signal, where the intensity is proportional to the population difference $N_\beta - N_\alpha$, and thus to the sample concentration.

## 4.2 Spectral Parameters

The spectral parameters consist of numerous observables obtained from an NMR experiment.[1] Only those applied in this thesis are discussed below.

### 4.2.1 Spin-Spin Coupling

The nomenclature for the common coupling constants i.e. proton-proton and proton-carbon, are J(H, H) and J(C, H) and the number of bonds through which the nuclei are coupled is given by n preceding the J, such that we have $^nJ$. For the purpose of this study, only $^2J$-geminal coupling and $^3J$-vicinal coupling are considered.

The importance of spin-spin coupling comes into context when one considers the factors that influence the coupling constants. They are:[1,2]

1. The hybridisation of the atoms involved
2. Bond angles and torsional angles
3. Bond lengths
4. The effects of neighbouring bonds and lone-pairs
5. Substituent effects

More specifically, if we obtain a $^2J$ (H, H) coupling from non-equivalent proton pairs on a $CH_2$ group, it is possible to determine the bond angle between the two protons and use this to help in structure determination. This line of thinking is also applicable to the $^2J$ (C, H) coupling. Similarly, if we obtain the $^3J$ (H, H) and $^3J$ (C, H) vicinal couplings, we can determine the bond angles, as well as the torsional or dihedral angles. To better understand vicinal couplings as a function of the torsional angles in saturated systems it is best to use a Karplus relationship curve, which relates the value of the $^3J$-coupling to the dihedral angle $\varphi$.
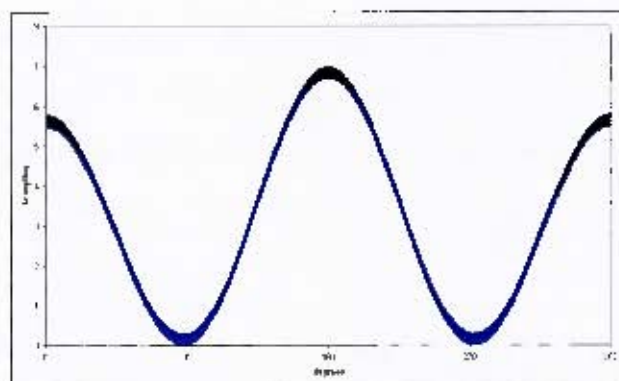
Figure 4.1 Graphical representation of a Karplus relationship

The lower curve in Figure 4.1 corresponds to the theoretical Karplus curve and the hatched area indicates the range in which $^3J$ (H, H) couplings vary in practice.

The Karplus equation, which was introduced by Martin Karplus in 1959,[5] 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.[6-9] The Karplus equation is usually in the form:

$$J = A\cos^2 \varphi - B\cos\varphi + C \tag{4.1}$$

containing the constants A, B and C, which are dependent on the system across which the torsional angle extends. The J-couplings in this thesis were calculated using Karplus equations with Fraser and Tvaroska parametrisation obtained from our Hungarian collaborators and are as follows:

$$J_{HH} = 12.1\cos^2 \varphi - 2.05\cos\varphi - 0.3 \tag{4.2}$$

$$J_{HC} = 6.1\cos^2 \varphi \quad 1.2\cos\varphi + 0.3 \tag{4.3}$$

### 4.2.2 Exchangeable Protons

The study of intra- and intermolecular interactions, and in particular hydrogen bonding, is hampered by the fast chemical exchange of the OH and NH groups with the aqueous phase. Thus, in order to obtain NMR data on fast chemical exchange protons, the solvent can be changed, as was done in this study, by substituting deuterated water ($D_2O$) with a deuterated DMSO ($d^6$-DMSO) solvent. In addition, the temperature dependence of the chemical shifts of the OH and the NH protons is indicative of the degree of solvent exposure and thus leads to a better understanding of the hydrogen bonding interaction.

### 4.2.3 Relaxation

By irradiating at a particular resonance frequency, the thermal equilibrium of the spin system can be disturbed. This process causes alterations in the population ratios and introduces $M_x$ and $M_y$ magnetic field components.

When the system returns to its original state (equilibrium), we say that the system has relaxed. There are two forms of relaxation that must be considered, namely the relaxation in the applied field direction (the longitudinal or spin-lattice relaxation time, $T_1$) and the relaxation perpendicular to the field direction (transverse or spin-spin relaxation time, $T_2$).

### 4.2.3.1 Spin-Lattice Relaxation ($T_1$)

Spin-lattice relaxation is associated with a change in energy of a spin system. This change is caused by rotating the macroscopic magnetisation vector $M_0$ through either a 90° pulse (in the y direction) or a 180° pulse (in the – z direction). This approach changes the $M_z$ value to zero when applying the 90° pulse or to $M_0$ when applying the 180° pulse.

These changes in the $M_z$ value cause an alteration in the population ratios and after the pulse has subsided the equilibrium state is restored. The rate at which this decay occurs is known as the spin-lattice relaxation time $T_1$, described mathematically as follows:

$$\frac{dM_z}{dt} = \frac{(M_0 - M_z)}{T_1}$$ (4.4)

The energy absorbed upon rotating the macroscopic magnetisation vector $M_0$ is transferred to the surrounding molecules (lattice) during the decay process.

### 4.2.3.2 Spin-Spin Relaxation ($T_2$)

A 90° pulse resulting in $M_z$ equaling zero, which is only possible if the populations of $N_\alpha$ and $N_\beta$ are equal, leads to the existence of only a transverse magnetisation component $M_y$. Over time the $M_y$ component decays as the bunched nuclear dipoles fan out. This decay is described as follows:

$$\frac{dM_y}{dt} = \frac{-M_y}{T_2}$$ (4.5)

where the time constant $T_2$ represents the spin-spin relaxation time.

# 4.3 Relaxation Parameters from Molecular Dynamics Simulations

Before obtaining relaxation parameters from a MD simulation it is important to note the two common mechanisms used to describe the exchange of energy between a nucleus and its environment. These are:

1. Dipole-dipole interactions
2. Chemical shift anisotropy (CSA) effects

Since the CSA mechanism plays no role in the relaxation of the $^{13}C$ nuclei in carbohydrates, due to the low dependency of the chemical shift on the molecular

orientation, the rate of relaxation depends solely on the neighbouring nuclei that are directly attached to the nucleus studied. It can thus be shown that the dipolar contribution to both the $T_1$ and $T_2$ relaxation times are related to the motion of the X-H vector as defined in the model-free formalism.[3,4]

The relaxation parameters are a function of the Fourier transform or spectral density $J(\varphi)$ of the correlation function C(t), which is formulated as in equation 4.6, in which the X-H vector is denoted by $\mu_{LF}$. In this formulation X has no effect on the dipole interaction apart from it being connected to H. The spectral density is given in equation 4.7

$$C(t) = \frac{1}{5}\langle P_2(\overrightarrow{\mu_{LF}}(0) \bullet \mu_{LF}(t))\rangle \qquad (4.6)$$

$$J(\varphi) = 2 \int_0^\infty C(t)\cos t\varphi t dt \qquad (4.7)$$

Having defined the spectral density $J(\varphi)$, the $T_1$ and $T_2$ relaxation parameters can be expressed as:

$$T_1^{-1} = \frac{1}{4}D^2[J(\varphi_H - \varphi_X) + 3J(\varphi_X) + 6J(\varphi_H + \varphi_X)] \qquad (4.8)$$

$$T_2^{-1} = \frac{1}{8}D^2[4J(0) + J(\varphi_H - \varphi_X) + 3J(\varphi_X) + 6J(\varphi_H) + 6J(\varphi_H + \varphi_X)] \qquad (4.9)$$

where D is the dipolar coupling constant.

The correlation function C(t) can be examined via MD simulations. This is achieved by making the following approximation: the factors affecting the correlation function come only from the internal motion and molecular tumbling of the molecule. If these properties are taken to be independent of each other, they then yield an expression of the form:

$$C(t) = C_0(t)C_I(t) \qquad (4.10)$$

where $C_0(t)$ is the overall correlation, represented by a decay constant $1/\tau_m$, where $\tau_m$ is the rotational correlation time of the molecule. The internal correlation $C_I(t)$ is calculated with each trajectory frame least square fitted to a reference coordinate set. The rotational correlation time, $\tau_m$ is related to the correlation time via the following function:

$$C(t) = \begin{cases} C_I(t)e^{-t/\tau m} & \text{for } t \leq t_{max} \\ C_I(t_{max})\, e^{-t/\tau m} & \text{for } t > t_{max} \end{cases} \qquad (4.11)$$

where $t_{max}$ is the maximum time up to which the correlation time is calculated.

# 4.4 Calculation of Generalised Order Parameters from Simulation

The steady state limit $S^2$, better known as the generalised order parameter, which indicates the extent of spatial restriction of the motion of a molecule, can be calculated from the MD simulation.[3]

This is achieved by using the addition theorem for spherical harmonics and the simultaneous use of the correlation function in equation 4.6. The resulting formulation is:

$$S^2 = \frac{5}{4\pi} \sum_{-2}^{2} \left| \left\langle y_{2m}(\theta_{mol}, \varphi_{mol}) \right\rangle \right|^2 \qquad (4.12)$$

and gives the order parameter, where $y_{lm}(\theta,\varphi)$ are the spherical harmonics and the C-H vector spherical polar angles are represented by the terms $\theta_{mol}$ and $\varphi_{mol}$ in the molecular coordinate frame model.

In reality, a "molecular" reference frame can be obtained by calculating the least squares rotation for the angles $\theta_{mol}$ and $\varphi_{mol}$ from each trajectory, when placed on a set of reference coordinates, and measuring the angles in the rotated set.

# References

1. Friebolin, H. *Basic One- and Two-Dimensional NMR Spectroscopy*, VCH Verlagsgesellschaft mbH, D-6940 Weinheim (Federal Republic of Germany); **1991**.

2. Abraham, R. J.; Fischer, J.; Loftus, P. *Introduction to NMR Spectroscopy*, John Wiley & Sons: Chichester, **1988**.

3. Lipari, G.; Szabo, A. J. *J. Am. Chem. Soc.*, **1982**, *104*, 4546.

4. Lipari, G.; Szabo, A. J. *J. Am. Chem. Soc.*, **1982**, *104*, 4549.

5. Karplus, M. *J. Chem. Phys.*, **1959**, *30*, 11.

6. Tvaroska, I.; Hricovini, M.; Petrakova, E. *Carbohydr. Res.*, **1989**, *189*, 359

7. Xu, Q.; Bush, A. *Carbohydr. Res.*, **1998**, *306*, 335.

8. Kao, L.-F.; Barfield, M. *J. Am. Chem. Soc.*, **1985**, *107*, 2323.

9. Fraser, R.R.; Kaufman, M.; Morand, P.; Govil, G. *Can. J. Chem.*, **1969**, *47*, 403.

# Chapter 5

# Force-Field Development, Simulation and Analysis

## 5.1 Introduction

To fully explore the influence of structural effects on the proposed properties of three-bond interglycosidic linkages, novel diglycosyl disulfides (chapter 1.4) have been used. As a result of the flexibility of carbohydrates, it is not possible to ascertain their conformational behaviour using only experimental techniques such as NMR. The added information gained from molecular dynamics (MD) computer simulations provides a clear understanding of solution saccharide structures.

To introduce the problem presented here is a simplification of ITZ55, ITZ56 and ITZ57 previously discussed in chapter 1. Each of the disaccharide disulphide mimetics contains a D-glucose ring which is for the purpose of further discussion denoted as residue 1 (ring 1), with residue 2 (ring 2) reserved for the 2-NAc containing D-glucose ring in ITZ55, the D-mannose ring in ITZ56 and the D-galactose ring in ITZ57 respectively.

The naming conventions introduced for the interglycosidic linkage torsion angles (chapter 1.4) have been used throughout this chapter, along with the following definition of the hydroxymethyl groups in all three mimetics:

$$\theta_1 = O5 - C5 - C6 - O6$$
$$\theta_2 = O5' - C5' - C6' - O6'$$

The subscripts 1 and 2 denote ring 1 and 2, respectively. In addition a torsion angle is introduced for ITZ55 in order to define the torsion angle for the 2-NAc group:

$$\sigma = C3 - C2 - N2 - H22$$
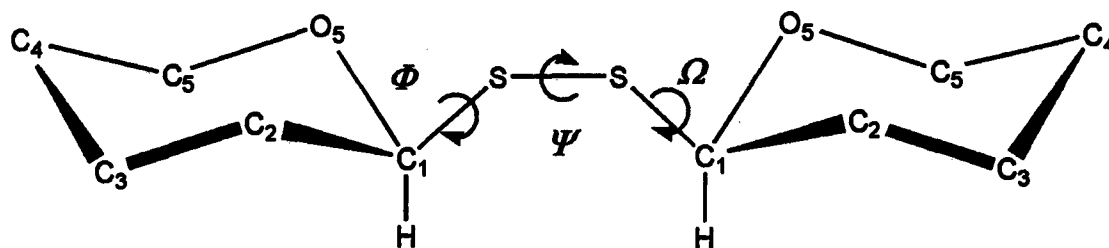
Figure 5.1 Definition of torsion angles characterizing the interglycosidic linkage.

# 5.2 Force Field Parametrisation

Molecular mechanics (MM) is described by the force field description of a molecule as introduced in chapter 3, in which the bonded and non-bonded terms for the atoms making up a molecule are described by potential functions.

The non-bonded terms, consisting of the van der Waals and Coulombic terms, describe atomic physical properties such as the size and charge of the atoms and are used to adequately describe the respective atoms' electronic states. The parameters for the non-bonded terms used for this work were taken from the CHARMM22[1] and Carbohydrate Solution Force Field (CSFF)[2] force-fields respectively and were not reparameterised.

The bonded terms, generally described by standard harmonic potential functions or Fourier terms, are used to approximate the energy associated with bond stretching, bond angular stretching and torsional rotation. As with the non-bonded parameters, the parameters for the description of the SS-linkage were adopted from the CHARMM22 force field, reparameterised to check their validity and incorporated into the CSFF force field. Tables 5.2-5.4 below give details of the reparametrised bond parameters. All other parameters not directly related to the SS-linkage were taken from the CSFF force field.

### 5.2.1 Reparameterisation of Sulphur-Sulphur Linkage

The preceding trend in force field-development for saccharides focused on obtaining MM parameters from *ab initio*/DFT calculations conducted on target molecules. Examples of these are the development of the GLYCAM_93 force field by Woods et al.[3], which was developed for MD simulations of glycoproteins and oligosaccharides, as well as the fitting of the OPLS force field to *ab initio* calculations on monosaccharides by Jorgensen et al.[4]

SS-linkage parameters have previously been incorporated into the CHARMM program via the development of the CHARMM22 force field, to accommodate the SS-linkage joining two cysteine amino acids in proteins. These parameters have been recalculated using DFT calculations on the minimum model compound containing a SS-linkage, dimethyl disulphide (DMDS), shown below in Figure 5.2. This molecule was chosen as it mimics the essential features of the SS-interglycosidic linkage.



Figure 5.2 DMDS fragment showing ($R_1 = R_2 = CH_3$) used for DFT and MD potential energy calculations

Presented below in Table 5.1 is the atom classification used for the sulphur, carbon and hydrogen atoms involved in the parameterisation process. The two different sulphur atoms, SUL and SLF were initially introduced to distinguish them from each other as the constituents may not always be equal as in the case of DMDS. The carbon (CTS) and hydrogen (HAS) atoms were taken from the CSFF force field and were used in the reparameterisation of the SS-linkage parameters.

| Name | Description |
|------|-------------|
| SUL | Sulphur contained in disulphide linkage |
| SLF | Alternate sulphur atom |
| CTS | Sugar aliphatic carbon |
| HAS | Sugar aliphatic hydrogen |

Table 5.1 CHARMM atom classification for the sulfur, carbon and hydrogen atoms required in the disulphide glycosidic linkage.

### 5.2.1.1 Bond and Angle Parameters

Initial bond and angle values for the SS-linkage were obtained from the CHARMM22 force field. Upon recalculating these parameters minor adjustments were made to the reference bond lengths and bond angles based on the calculations.

From the calculations the bond and angle force constants, $K_b$ and $K_\theta$, were found to yield same values as those reported by Mackerell et al in the CHARMM22 force field. The new reference bond and angle values are reported in Tables 5.2-5.3. The old reference values are presented in brackets.

| Bond | $K_b$ | $r_0$ |
|------|-------|-------|
| SUL-SLF | 173.0 | 2.029 |
| SLF-CTS | 214.0 | 1.814 |
| SUL-CTS | 214.0 | 1.814 |

Table 5.2 Bond Parameters for the disulphide glycosidic linkage, where $K_b$ is the force constant in kcal.mol$^{-1}$.Å$^{-1}$ and $r_0$ is the equilibrium bond length in angstroms (Å).

66

| Bond | $K_\theta$ ( kcal.mol$^{-1}$.rad$^{-2}$ ) | $\theta_0$ (°) |
|---|---|---|
| CTS-CTS-SLF | 58.0 | 109.0 (112.5) |
| HAS-CTS-SLF | 38.0 | 111.0 (111.0) |
| CTS-SLF-SUL | 72.5 | 103.3 (103.3) |
| CTS-CTS-SUL | 58.0 | 109.0 (112.5) |
| HAS-CTS-SUL | 38.0 | 111.0 (111.0) |
| CTS-SUL-SLF | 72.5 | 103.3 (103.3) |

Table 5.3  Angle Parameters for the disulphide glycosidic linkage, where $K_\theta$ is the force constant in kcal.mol$^{-1}$.rad$^{-2}$ and $\theta_0$ is the equilibrium bond angle in degrees (°).

## 5.2.1.2 Dihedral Parameters

In addition to the bond and angle parameters, dihedral parameters are required for the SS-linkage. Initial energy profiles of the SS-linkage dihedral angles were obtained using the dihedral parameters contained in the CHARMM22 force field using the CHARMM program[5] for MM and dynamics calculations of macromolecules. This was achieved by varying the torsional angles through 360°, constraining at 10° intervals, followed by steepest descent and conjugate gradients minimisation. Every structure modelled at each interval was studied and the lowest-energy structures were recorded.

Single point DFT calculations were conducted at every point for each of the resulting MM minimised structures. The geometries of the model compound was optimised with the GAUSSIAN98 program[6] using the restricted B3LYP level of theory, and the split-valence 6-31g (d) basis set. Rotational potential energy profiles were generated for each dihedral 10° increments by rotating from 0° to 360° with clockwise rotation around the respective bonds. The MM rotational plot was then fitted to the DFT plot and it was found that the barrier to rotation around the S-S' bond from the –g to +g conformation via an anti transition state was higher with the original force constants.

The force constant, $K_\omega$ for the CTS-SLF-SUL-CTS dihedral, was altered via an iterative method. The MM calculations for the CTS-SLF-SUL-CTS dihedral were repeated at each new $K_\omega$ and the rotational energy profiles were calculated until a best fit to the DFT plots was obtained. The CTS-SLF-SUL-CTS dihedral plot for DMDS is represented in Figure 5.3 below and shows energy minima with the S-S bond in the +g and −g conformations. The torsional parameters for the SS-linkage used throughout this thesis are represented in Table 5.4 below. All other parameters used are already present in the CSFF force field

| Bond | $K_\omega$ ( kcal.mol$^{-1}$.rad$^{-2}$ ) | $N$ | Phase |
|------|------|------|------|
| HAS-CTS-SLF-SUL | 0.158 | 3 | 0.000 |
| CTS-SLF-SUL-CTS | 1.000 | 1 | 0.000 |
| CTS-SLF-SUL-CTS | 3.600 | 2 | 0.000 |
| CTS-SLF-SUL-CTS | 0.900 | 3 | 0.000 |

Table 5.4 Dihedral Parameters for the disulphide glycosidic linkage, where $K_\omega$ is the force constant in kcal.mol$^{-1}$.rad$^{-2}$ and N is the multiplicity.
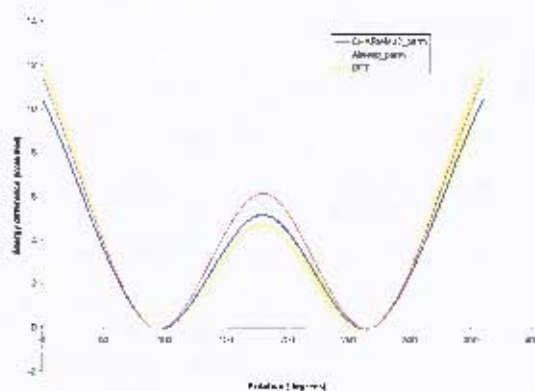


Figure 5.3 Potential energy differences vs. rotation of the S-S torsional angle.

# 5.3 Adiabatic Maps

Simulated annealing is the process whereby the rotational energy barriers of exocyclic functional groups can be overcome by performing molecular dynamics at high temperatures. This is done by heating a system and then slowly cooling it down, allowing a molecule to reach its lowest potential energy conformation. This method has been shown to have a particular application in the study of disaccharides, as demonstrated by Naidoo et al. in the study of GlcNac-β-(1-4)-GlcNac.[7] The method shown below was followed for all three disaccharide mimics and follows the protocol described by Naidoo et al.

1. **Initial Grid Search.** The conformational space of the disaccharide mimics is explored in terms of the dihedral angles rotation. The dihedral angles ($\Phi$, $\Psi$, $\Omega$) energy space is divided into grids by sampling $20°$ increments for each glycosidic dihedral, effectively generating a stack of $\Phi$ - $\Omega$ maps for each value of $\Psi$. This was achieved by implementing harmonic constraints on the glycosidic dihedrals angles. Harmonic constraints were also applied to the endocyclic dihedral angles, such that the residue segments remain in their natural chair conformation and do not flip to their higher energy boat configurations. During the simulated annealing process the disaccharide mimetics were heated from 200 K to 800 K in increments of 100 K.ps$^{-1}$, over 7 ps. The 800 K state was maintained for a further 3 ps of equilibration before the cooling process was implemented. The rate of cooling was in increments of 30 K.ps$^{-1}$ and the mimetics were cooled from 800 K to 300 K, over a 20 ps period. The cooling rate was deliberately slow in order to allow exploration of the mimetics lower energy regions.

2. **Map Refinement and Smoothing.** Although the initial energy surface generated by the first grid search would stand a good chance of finding low energy minima, it does not necessarily give the conformation with the lowest energy. In order for the low energy regions to be more comprehensively explored, a refinement process was conducted at five selected minima for each slice in $\Psi$. Each

minimum, for each mimetic, was then subjected to 35 cycles of simulating annealing, as described in step 1, and the lowest resulting conformations were chosen for at each 20° increment in $\Psi$. Upon completion of the above mentioned process, the five conformations with the lowest energies were selected for a map 'smoothing' procedure. 'Smoothing' was carried out using the same combined-methods minimisation procedure, but on a 10° interval grid search. The resulting maps were studied and the lowest energy conformations were used for the dynamic simulations.

### 5.3.1 Map Features

With the substantial increase in the volume of conformational space available to the disaccharide mimetics (as a result of added bond in the $\beta S(1\rightarrow1)S'$ linkage), it is interesting to find that for each mimetic the majority of the local minima are primarily restricted to two low energy regions, where $\Psi$ is in the +g and −g conformation.

The minimum energy regions were obtained by analysis of each point in the three dimensional energy plots obtained for the respective disaccharide mimetics. Several local minima were observed for each and the five lowest lying wells for ITZ55, ITZ56 and ITZ57 are shown in Figures 5.4 to 5.6 as contour plots with energies plotted in 2 kcal/mol increments. The contour plots were obtained by slicing through the three dimensional energy plots. Energies up to 12 kcal/mol above the respective global minima for each of the disaccharide mimetics have been included.

The local minima for ITZ55, ITZ56 and ITZ57 have been labelled A to E respectively, and are summarised below in Tables 5.5 to 5.7.

70

| Label | $\Phi$ | $\Psi$ | $\Omega$ | $E_{relative}$ (kcal/mol) |
|-------|------|------|------|---------------------------|
| A | 30 | 90 | -40 | 1.666 |
| B | 30 | 80 | -30 | 1.063 |
| C | 0 | 80 | 50 | 0.000[†] |
| D | 40 | -90 | 180 | 1.118 |
| E | 40 | -80 | 30 | 1.488 |

Table 5.5 Potential energy minima from the ITZ55 adiabatic maps, with energies relative to the lowest energy structure, whose energy has been set to 0.000 kcal/mol.
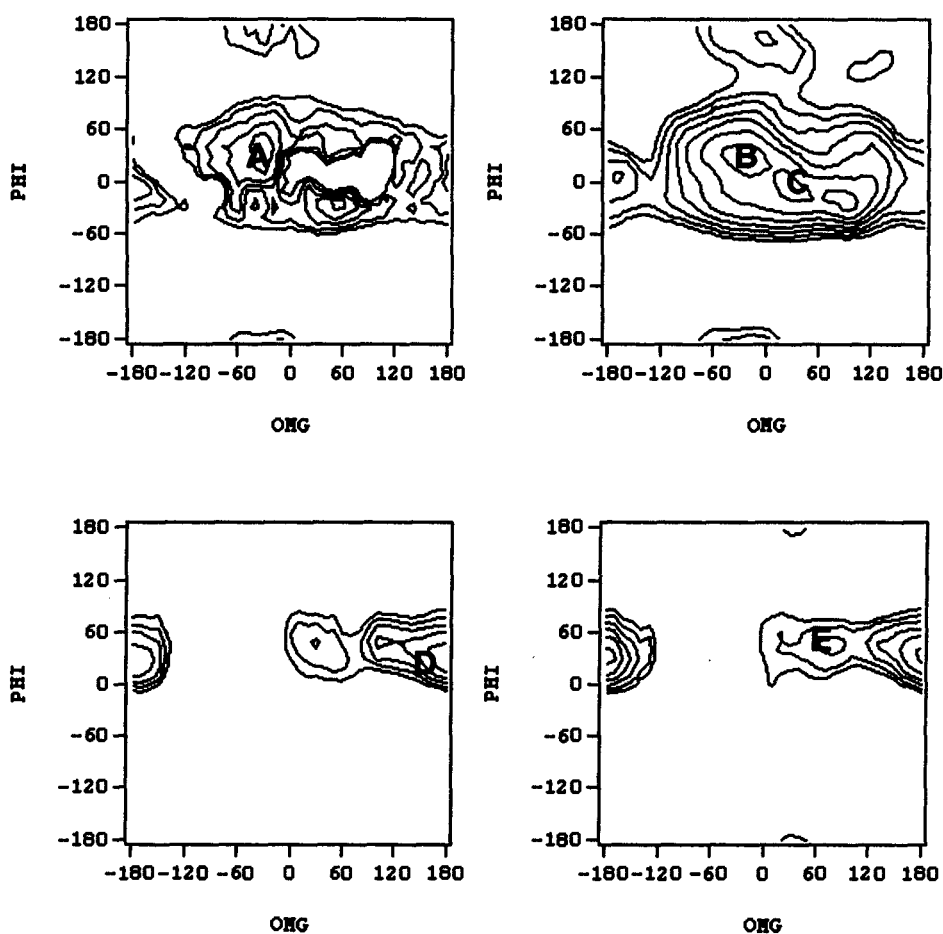


Figure 5.4 Slices through the ITZ55 three-dimensional adiabatic map for $\Psi = 90°$ (top left), $\Psi = 80°$ (top right), $\Psi = -90°$ (bottom left) and $\Psi = -80°$ (bottom right), with energies plotted in 2 kcal/mol increments.

| Label | $\Phi$ | $\Psi$ | $\Omega$ | $E_{relative}$ (kcal/mol) |
|-------|--------|--------|----------|---------------------------|
| A | 40 | 80 | -50 | 0.01 |
| B | 40 | 80 | 30 | 3.229 |
| C | 40 | 90 | 60 | 2.675 |
| D | 30 | -90 | 160 | 2.942 |
| E | 50 | -90 | 50 | $0^\dagger$ |

Table 5.6 Potential energy minima from the ITZ56 adiabatic maps, with energies relative to the lowest energy structure, whose energy has been set to 0.000 kcal/mol.
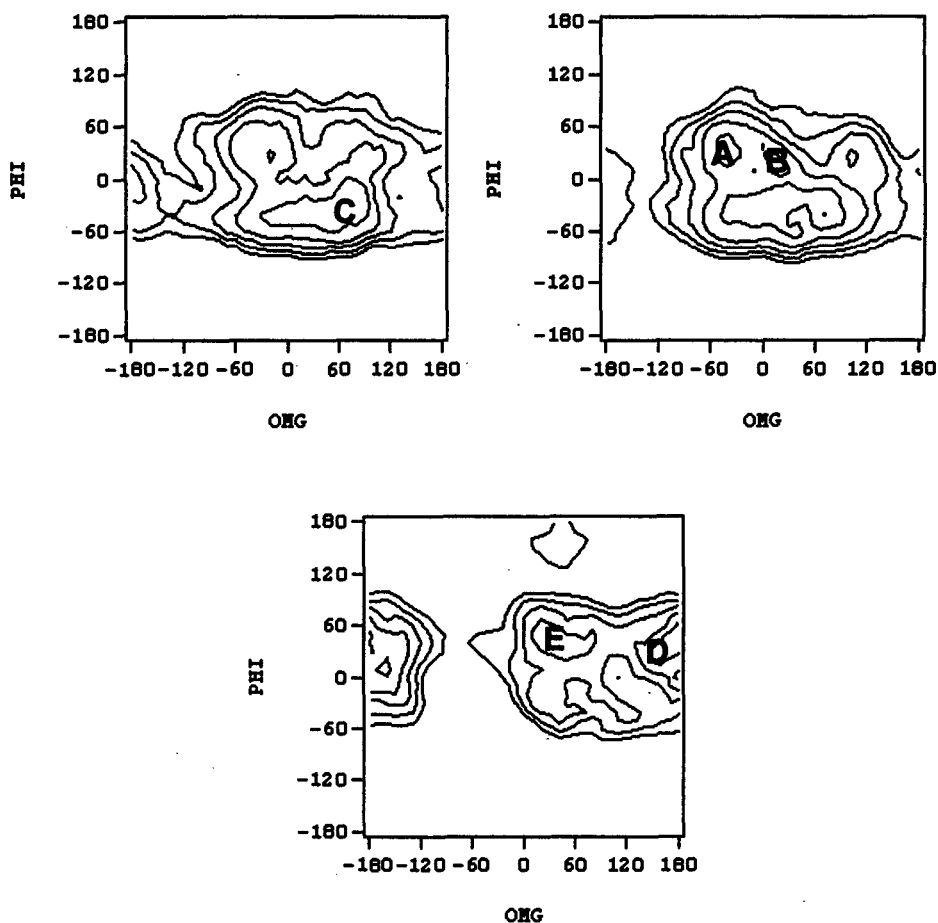


Figure 5.5 Slices through the ITZ56 three-dimensional adiabatic map for $\Psi = 90^\circ$ (top left), $\Psi = 80^\circ$ (top right), $\Psi = -90^\circ$ (bottom), with energies plotted in 2 kcal/mol increments.

72

| Label | $\Phi$ | $\Psi$ | $\Omega$ | $E_{relative}$ (kcal/mol) |
|---|---|---|---|---|
| A | 40 | 90 | -30 | 0.697 |
| B | 10 | 90 | -180 | 0.000[†] |
| C | -20 | 90 | 60 | 0.548 |
| D | 30 | -90 | -180 | 0.462 |
| E | 40 | -90 | 40 | 0.999 |

Table 5.7  Potential energy minima from the ITZ55 adiabatic maps, with energies relative to the lowest energy structure, whose energy has been set to 0.000 kcal/mol.
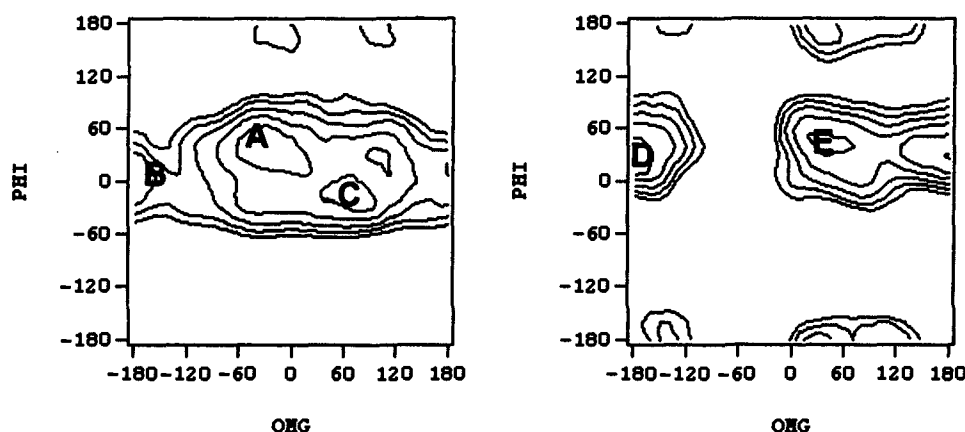


Figure 5.6    Slices through the ITZ57 three-dimensional adiabatic map for $\Psi = 90°$ (left) and $\Psi = -90°$ (right), with energies plotted in 2 kcal/mol increments.

Initial analysis of the energy minima for the disaccharide mimetics indicates that, $\Psi$ is found primarily in the +g and −g conformations, with a variation of approximately $10°$ in ITZ55 and ITZ56. These values are in accordance with previous experimental studies by Szilágyi et al.[8] and Davis et al.[9]

For each disaccharide mimetic $\Phi$ and $\Omega$ show considerably more rotation than $\Psi$, especially $\Omega$. Closer analysis reveals that $\Phi$ fluctuates between -30° and 40° for ITZ55, between -40° and 40° for ITZ56 and between 10° and 60° for ITZ57.

Furthermore $\Omega$ appears to have a greater range of torsional rotation with values ranging from approximately -30° to 180° for each of the respective disaccharide mimetics. Graphically, and in summary, the minima for each disaccharide mimetic

appear as a band with $\Phi$ approximately between -20° and 50°, $\Psi$ in either the +g or –g orientation and $\Omega$ having rotation through the full 360°, as seen in Figures 5.4 to 5.6.

These results can be explained by closer examination of the possible 1, 4 steric clashes expected in a three-bond interglycosidic linkage, using representative Newman-projections as shown in Figure 5.7.



$$(S1\text{-}C1) \equiv \Phi \qquad (S1\text{-}S'1) \equiv \Psi \qquad (S'1\text{-}C'1) \equiv \Omega$$
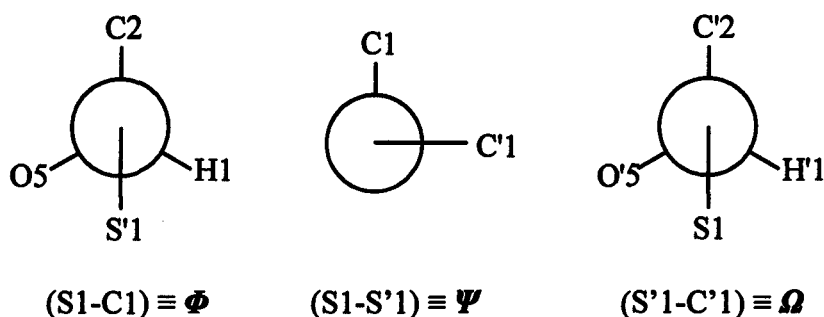
Figure 5.7 Newmann-Projections for the SS-interglycosidic linkage

In general the 1, 4 steric clashes indicate that for respective energy minima, $\Phi$ is orientated such that the S1 atom of residue 2 (S'1) is trans to the C2 atom on residue 1 and gauche relative to the corresponding O5 atom. The exceptions being minima C in ITZ55 and ITZ57, respectively. Minima C for ITZ55 reveal that S'1 is gauche to both the C2 and O5 atoms of residue 1. Whereas, in minima C for ITZ57, S'1 is orientated gauche to C2 and trans to O5.

Closer scrutiny of $\Omega$, reveals that each possible rotamer does exist. This is surprising, because, when the S1 – H'1 torsional angle is found at 180°, S1 is seen to be 'trapped' between O'5 and C'2, which should result in a restriction the S-C bond rotation, due to the arising steric clashes (interactions), and as such, restrict the number of rotamers.

The possibility that these conformations occur may be explained by the overriding effect of the disulfide torsional angle. Upon initial observation of the disaccharide mimetics, one would expect the lowest energy conformations to be obtained where the C1 and C'1 atoms of ring 1 and ring 2 respectively, are in the *trans* orientation. This is not found to occur. The respective C1 atoms are observed to be in the +g and –g

orientations. The reasons behind the preference of the +g and −g orientations of the disulfide torsional angle have been studied further using electronic structure methods and is discussed in section 5.4.

An analysis of the energy minima of disaccharides would not be complete without an analysis of the secondary alcohols found on each glycosyl ring. For each of the diglycosyl disulphide mimetics a clockwise crown is formed in ring 1, and in ring 2 an anti-clockwise crown is formed. The other feature repeated in the diglycosyl disulphide mimetics is the preference for the *gg* rotamer for the hydroxymethyl groups. In conclusion, the general conformational differences between the disaccharide mimetics exist only in the interglycosidic torsional angles and not with respect to their hydroxymethyl groups.

## 5.4 Electronic Analysis of the Disulphide Linkage

Electronic analysis was conducted on a 1, 1'- disulphanediyl diethanol (DSDE) model compound. This model reproduces the chemical environment found around the $\beta S(1 \rightarrow 1)S'$ linkage, without computing the entire molecule.

Single point calculations were conducted with the GAUSSIAN98 program[6] using the restricted B3LYP level of theory and the split-valence 6-31g (d) basis set. Optimisation was obtained upon rotating the C-S bond through 360° at 60° interval so as to determine the effect of the sulphur – oxygen, sulphur – carbon and sulphur – hydrogen interactions. Similarly rotation around the S-S bond was conducted, but at 90° intervals so as to examine the differences between the +g and −g configurations of the disulphide linkage as well as the possible anti and syn transitions.

Analysis of the resulting data was conducted using the Atoms in Molecules (AIM) theory developed by Bader,[10] followed by a Natural Bond Orbitals (NBO) analysis as developed by Weinhold et al.[11]

### 5.4.1 Atoms in Molecules (AIM) Analysis

The AIM2000 program[12] was used to perform an AIM analysis of the DSDE model compound. Wavefunction files consisting of the molecular orbital coefficients were generated using B3LYP/6-31G (d). Bond critical points (3,-1) were determined by searching between the various nuclear attractors (3,-3). Bond paths were constructed by following the eigenvectors corresponding to the negative eigenvalues at specific bond critical points. The collection of these eigenvectors form a line of maximum increase in electron density starting from the (3,-1) critical points leading to (3,-3) nuclear attractors.



Figure 5.8  DSDE representation produced by AIM2000. The bond critical points are indicted by the small red spheres between the atomic nuclei

Contour plots of the laplacian of the electron density, $\nabla^2 \rho(r)$, were constructed. These plots indicate areas of local charge concentration ($\nabla^2 \rho(r) < 0$) and local charge depletion ($\nabla^2 \rho(r) > 0$), where the charge depletion and charge concentration are indicated by dashed and solid lines respectively in the $\nabla^2 \rho(r)$ plots as shown in Tables 5.9 and 5.10.
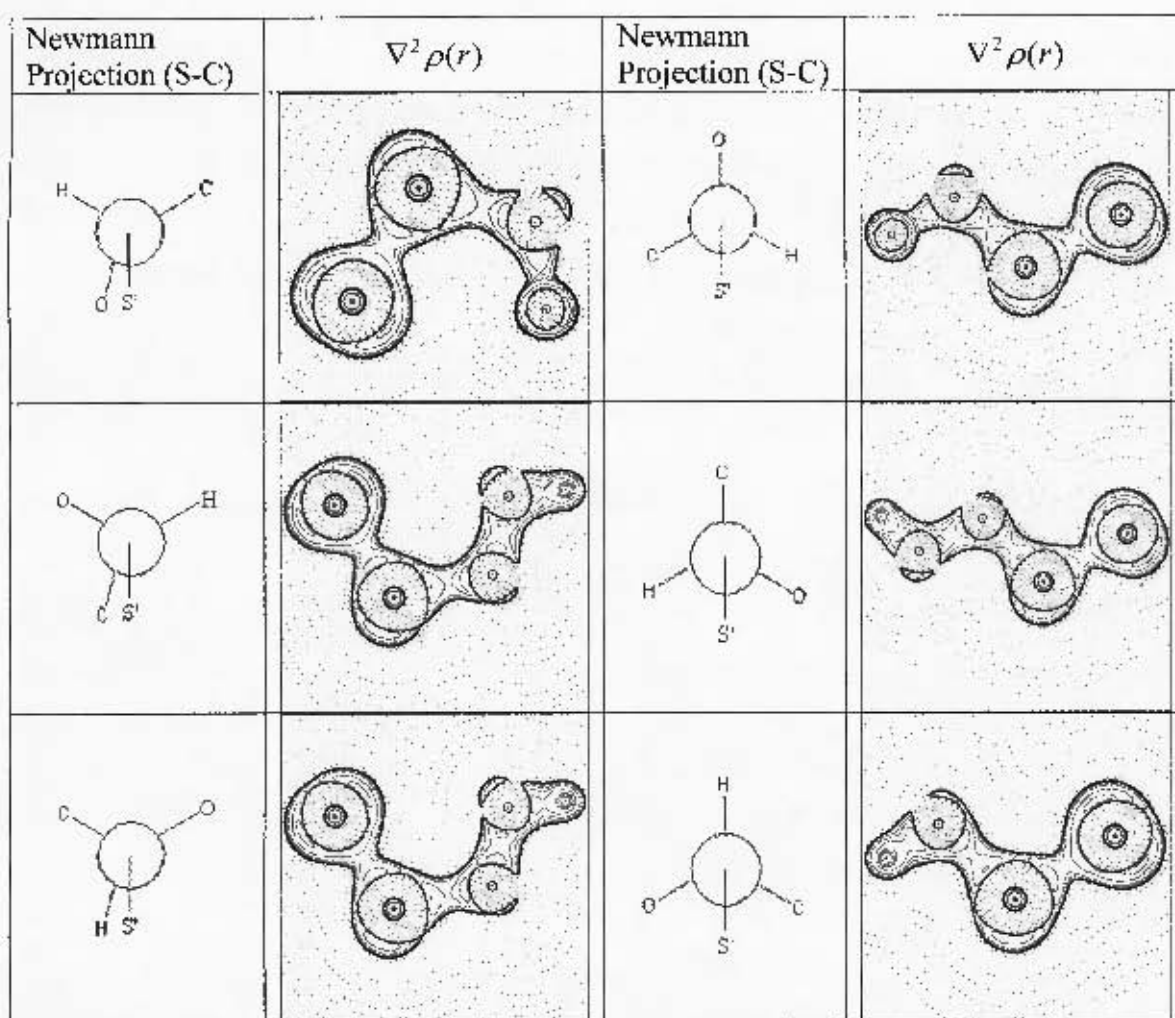
Figure 5.9 Newmann projections of the C-S bond in DSDE representing the possible 1, 4 steric interactions, along with the corresponding contour line diagrams of the Laplacian distribution $\nabla^2\rho(r)$ of the C-S bond. These contours were made with the −g conformation present for the disulphide linkage with the dashed lines indicating charge depletion and solid lines indicate charge concentration.

Contour line diagrams of $\nabla^2 \rho(r)$ indicate that there is an equal charge distribution between the sulphur atoms as would be expected with sulphurs in an equivalent chemical environment. No indication is given of any unexpected distortions in the electron densities for the oxygen or carbon atoms with respect to the sulphur atoms in each of the conformations exhibited in Figures 5.9 and 5.10. Therefore no conclusion can be made regarding the disulfide linkages preference for the +g and −g conformations based on the initial electronic structural analyses.



| Newmann Projection (S-C) | $\nabla^2 \rho$ | Newmann Projection (S-C) | $\nabla^2 \rho$ |
|---|---|---|---|
| | | | |
| | | | |
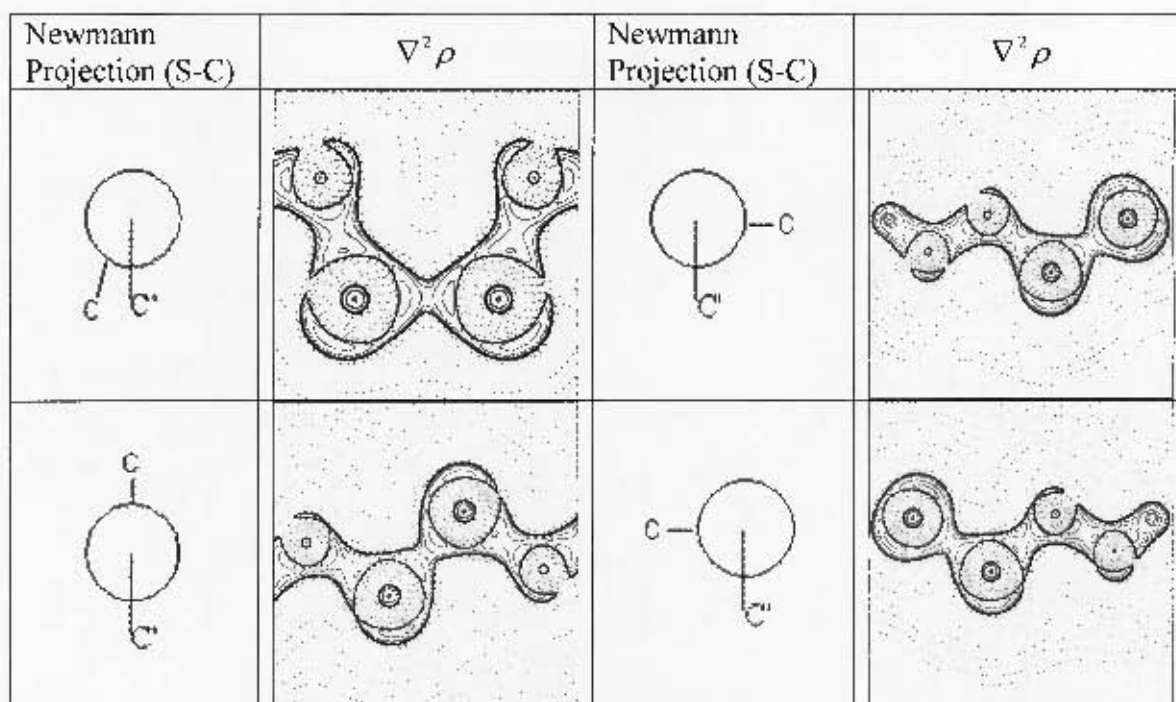
Figure 5.10  Newmann projections of the S-S bond in DSDE representing the possible 1, 4 steric interactions between the carbon atoms, along with the corresponding contour line diagrams of the Laplacian distribution $\nabla^2 \rho(r)$ of the S-S bond. The dashed lines indicate charge depletion and solid lines indicate charge concentration.

### 5.4.2 Natural Bond Orbital (NBO) Analysis

NBO's of the DSDE model were generated by a static NBO calculation using the 2nd order pertubative analysis of the Fock matrix in the NBO basis. B3LYP/6-31G (d) and B3LYP/6-311G (d, p) were both used to check for basis set consistency, with both giving consistent results.

No polarisation was found to exist between the sulphur atoms in the S-S bond, which corresponded to the AIM analysis. The sulphur atoms in DSDE were also found to exhibit more s character upon hybridisation, with a hybridisation of $\sim sp^{2.3}$, whereas the oxygen and carbon atoms were found to have the expected $sp^3$ hybridisation.

An initial indication into the preference of the –g and +g disulfide conformations was obtained. This phenomena is attributed to a two-electron stabilising $n_O \rightarrow \sigma^*_{C-S}$ delocalisation, from the lone pair of the oxygen atom to the carbon-sulphur anti-bonding orbitals which is not present in the anti and syn conformations. The $n_O \rightarrow \sigma^*_{C-S}$ electron delocalisation is found to be stronger in the –g conformation relative to the +g conformation, which may indicate a preference for the –g conformation in the pseudo $\beta S(1 \rightarrow 1)S'$ linkage in DSDE, even though the overall energy is lower for the +g conformation.

The $n_O \rightarrow \sigma^*_{C-S}$ electron delocalisation found for the $\beta S(1 \rightarrow 1)S'$ linkage in DSDE appears to contradict the findings of Chakka et al.[13] on the basis of the bonding - anti-bonding interactions. Chakka et al.[13] used arguments based on pertubational molecular orbital theory, to propose that two interactions stabilise the gauche conformational preference about the S-S bond rather than the one described above. These interactions are the $n_S \rightarrow \sigma^*_{S-C}$ hyperconjugative and the $n_S \rightarrow \sigma^*_{C-O}$ exo-anomeric interactions, respectively.

The difference in the electron delocalisation observed by us ($n_O \rightarrow \sigma^*_{C-S}$) and those published by Chakka et al.[13] ($n_S \rightarrow \sigma^*_{C-O}$) may be attributed to the chemical environment in which the interactions occur and the lack of an alkyl group on the oxygen atoms in the DSDE model when compared to Chakka et al.'s study on the full disaccharide.[13] This second factor also explains why no pseudo exo-anomeric effect is

observed for the DSDE model. Further study on the full disaccharide mimetics is thus required.
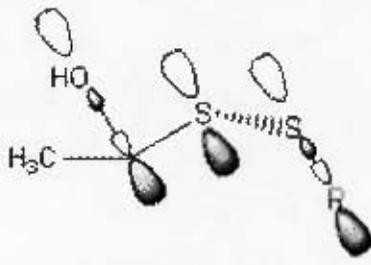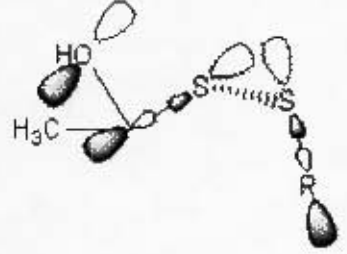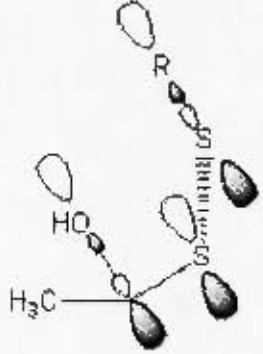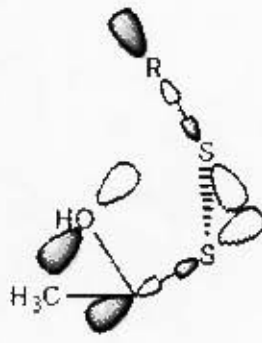
| Conformation about S-S linkage | Pertubational molecular orbital theory representations, Chakka et al.[13] | Natural Bond Orbitals obtained in this work |
|---|---|---|
| -g | | |
| +g | | |



Figure 5.11 Comparative representations of the electron delocalisation affecting the conformations about the S-S bond, as proposed by Chakka et al and those found in this thesis.[13]

# 5.5 Simulation Methodologies

MD simulations give insight into the dynamical behaviour of carbohydrates and for this reason it is used here to help understand the three diglycosyl disulphide mimetics' conformational properties above and beyond the simplistic picture given by a simple adiabatic map. MD also allows for a comparison with experimental data such as nuclear magnetic resonance (NMR) NOE's, J-couplings and relaxation times. This is achieved as MD simulations give a complete time evolution of a molecule.

Starting positions for MD simulations are thus of great importance and are obtained by selecting the energy minima (well's) based on the adiabatic maps. Vacuum simulations were started from each of the minima for ITZ55-ITZ57 found in section 5.3. Water and DMSO simulations were started from two selected minima from each disaccharide mimetic, such that one +g and one –g conformation was studied.

## 5.5.1 Vacuum Simulation

All of the chosen minima were heated from 200 K to 300 K over a period of 1 ns to avoid sudden energy increases. Constraints were placed on the glycopyranosyl dihedrals to maintain the $^4C_1$ boat conformations. The constraints were released and the simulations were allowed to proceed for periods of 10 ns respectively, using the leapfrog Verlet integrator, with data for all configurations stored at intervals of 0.5 ps.

## 5.5.2 Water Simulation

Water dynamic simulations were implemented using the TIP3P[14] water model to represent the solvent explicitly. A cubic box of side 24.8626Å containing 512 TIP3P water molecules was utilised. Each solute was placed at the centre of their respective box and all water molecules with heavy atom distances within 3.0Å from the solute were removed (The number of water and DMSO molecules used for each simulation are given below in Table 5.11). A 1ns equilibration was implemented and then MD simulations were allowed to proceed for 10 ns using the leapfrog Verlet integrator and the

implementation of the isothermal-isobaric ensemble (NPT) where the pressure and temperature are kept constant (P = 1 bar, T = 300 K) using the Langevin piston method[15] as found in CHARMM. Data for all respective configurations was stored at intervals of 0.5 ps.

### 5.5.3 DMSO Simulation

DMSO dynamic simulations were implemented using the DMSO model of Strader et al.[16] Dynamics were run in a cubic box with sides of 41.7262Å, starting with 612 DMSO molecules for each solute. The overlapping DMSO molecules within a heavy atom distance of 3.5Å from the solute were then removed. After an initial 1 ns equilibration the MD simulations were allowed to proceed for a further 6ns using the same procedure as per the water MD simulations.

| Simulation | Water | DMSO |
|------------|-------|------|
| +g (55) | 481 | 568 |
| -g (55) | 486 | 571 |
| +g (56) | 485 | 575 |
| -g (56) | 485 | 574 |
| +g (57) | 483 | 574 |
| -g (57) | 483 | 576 |

Table 5.8 Number of water and DMSO molecules in each solvent MD simulation

# 5.6 Simulation Analysis

### 5.6.1 Pucker Analysis

The initial standard in ascertaining if experimental data is reproduced via simulation is to analyse the glucopyranosyl rings for possible variations in their conformation such as the chair, boat and twist-boat during the simulation run, and see if the chair conformation is maintained. This is allowed as the $^4C_1$ chair conformation is found to be the sole conformation in solution. This ring analysis technique is referred to as puckering.

Puckering of the rings was analysed using the Cremer and Pople parameters for six membered rings,[17] and it was determined that for both the water and DMSO simulations the chair conformers exist with slight deviations due to the asymmetry of the rings.

### 5.6.2 Cluster Analysis

Cluster analyses were performed for the vacuum and solution MD simulations in order to ascertain the conformational preferences of ITZ55, ITZ56 and ITZ57. Clustering is applied to the principle degrees of freedom for the disaccharide mimetics. These degrees of freedom are the interglycosidic torsion angles: $\Phi$, $\Psi$ and $\Omega$ and the hydroxymethyl torsional angles, $\theta_1$ and $\theta_2$ for ITZ56 and ITZ57 allowing for complete analysis. For ITZ55, the second hydroxymethyl group was replaced with the 2-NAc torsional angle, $\sigma$. A cluster radius (section 3.4.1.2) was chosen and the most populous clusters for the various solution dynamics simulations are summarised below in Tables 5.5-5.10. The vacuum clusters for the disaccharide mimetics are given in appendix A.

| sim. | Cluster | # members | Std deviation | $\phi$ | $\psi$ | $\Omega$ | $\theta$ | $\sigma$ |
|---|---|---|---|---|---|---|---|---|
| +g | 1 | 8187 | 1.91E+01 | -20.14 | 93.88 | 161.43 | -66.32 | 74.43 |
| | 2 | 62948 | 1.41E+01 | 47.28 | 85.17 | -27.00 | -62.62 | 60.78 |
| | 3 | 13616 | 1.35E+01 | -12.85 | 84.47 | -24.49 | -62.08 | 57.57 |
| | 4 | 8972 | 1.75E+01 | -24.75 | 82.27 | 39.35 | 52.38 | 61.79 |
| -g | 1 | 20411 | 1.42E+01 | 48.82 | -89.44 | 170.86 | -63.16 | 63.64 |
| | 2 | 43017 | 1.35E+01 | 49.05 | -84.81 | -167.58 | -62.95 | 68.89 |
| | 3 | 14010 | 1.57E+01 | 45.62 | -84.95 | -169.25 | 62.06 | 66.40 |

Table 5.9    Largest clusters arising from ITZ55 trajectories in water, sampled from 100000 members.

| sim. | cluster | # members | Std deviation | $\phi$ | $\psi$ | $\Omega$ | $\theta$ | $\sigma$ |
|---|---|---|---|---|---|---|---|---|
| +g | 1 | 20249 | 1.23E+01 | -16.26 | 84.78 | 43.88 | -64.53 | 61.67 |
| | 2 | 14526 | 1.31E+01 | 36.32 | 86.88 | -17.20 | 59.83 | 52.46 |
| | 3 | 29177 | 1.18E+01 | 54.97 | 87.79 | -13.72 | -62.77 | 53.16 |
| | 4 | 12297 | 1.36E+01 | 24.42 | 91.71 | 25.99 | -66.19 | 58.33 |
| -g | 1 | 26868 | 1.23E+01 | 56.48 | -86.77 | 171.59 | -65.15 | 65.66 |
| | 2 | 20634 | 1.22E+01 | 58.23 | -84.62 | -165.52 | -66.02 | 95.96 |

Table 5.10 Largest clusters arising from ITZ55 trajectories in DMSO, sampled from 80000 members.

| sim. | Cluster | # members | Std deviation | $\Phi$ | $\Psi$ | $\Omega$ | $\theta_1$ | $\theta_2$ |
|---|---|---|---|---|---|---|---|---|
| +g | 1 | 21467 | 1.52E+01 | 44.14 | 82.75 | -56 85 | -63.51 | 52.20 |
| | 2 | 14770 | 1.71E+01 | 66.21 | 77.99 | -42 30 | -64.23 | -61.60 |
| | 3 | 10485 | 1.41E+01 | -30.52 | 86 15 | -55.82 | -61.88 | 51.19 |
| | 4 | 28837 | 1.30E+01 | 36.90 | 84.73 | -61.16 | -63.08 | -62.50 |
| -g | 1 | 7502 | 1.66E+01 | -38.65 | -96.13 | -167.44 | -62.95 | 59.14 |
| | 2 | 4792 | 2.27E+01 | -6.31 | -95.00 | 171.63 | -66.17 | 61.53 |
| | 3 | 7284 | 1.87E+01 | -13.71 | -90.21 | -166.21 | -61.94 | -52.19 |
| | 4 | 8925 | 1.33E+01 | -55.70 | -96.47 | 40.18 | -63.69 | 59.78 |

Table 5.11 Largest clusters arising from ITZ56 trajectories in water, sampled from 100000 members.

| sim. | Cluster | # members | Std deviation | $\Phi$ | $\Psi$ | $\Omega$ | $\theta_1$ | $\theta_2$ |
|---|---|---|---|---|---|---|---|---|
| +g | 1 | 12522 | 1.05E+01 | 49 52 | 85.48 | -56.18 | -60.81 | 53.03 |
| | 2 | 11504 | 1.08E+01 | 46.64 | 85.53 | -54.56 | -62.15 | 76.40 |
| | 3 | 9291 | 1.01E+01 | 30.52 | 79.46 | -62.31 | -69.81 | -72.47 |
| -g | 1 | 16077 | 1.26E+01 | 26.12 | -87.12 | 170.56 | -62.53 | 74.60 |
| | 2 | 10054 | 1.07E+01 | 33.62 | -85.12 | -170.79 | -61.15 | 66.39 |
| | 3 | 10043 | 1.14E+01 | -6.30 | -82.42 | -166.69 | -56.01 | 68.55 |

Table 5.12 Largest clusters arising from ITZ56 trajectories in DMSO, sampled from 800000 members.

| Sim. | cluster | # members | Std deviation | $\Phi$ | $\Psi$ | $\Omega$ | $\theta_1$ | $\theta_2$ |
|------|---------|-----------|---------------|--------|--------|----------|------------|------------|
| +g   | 1       | 8368      | 2.50E+01      | 4.79   | 93.28  | 167.28   | -63.97     | 77.13      |
|      | 2       | 4604      | 2.69E+01      | -9.55  | 95.87  | -171.31  | -63.3      | 79.45      |
|      | 3       | 3109      | 2.74E+01      | -12.7  | 93.81  | 170.18   | -64.25     | -127.95    |
|      | 4       | 3919      | 2.62E+01      | -27.62 | 94.29  | -170.97  | -63.12     | -98.09     |
| -g   | 1       | 20201     | 1.58E+01      | 36.12  | -84.34 | -167.93  | -62.14     | 64.76      |
|      | 2       | 12708     | 1.53E+01      | 37.38  | -85.74 | 171.02   | -63.16     | 65.69      |
|      | 3       | 9429      | 2.09E+01      | 51.77  | -80.24 | -167.03  | 53.77      | 57.18      |

Table 5.13 Largest clusters arising from ITZ57 trajectories in water, sampled from 100000 members.

| Sim. | cluster | # members | Std deviation | $\Phi$ | $\Psi$ | $\Omega$ | $\theta_1$ | $\theta_2$ |
|------|---------|-----------|---------------|--------|--------|----------|------------|------------|
| +g   | 1       | 18775     | 1.31E+01      | -20.49 | 97.59  | -167.72  | -65.64     | 72.21      |
|      | 2       | 8347      | 1.40E+01      | 51.78  | 107.19 | 169.27   | -62.01     | 73.31      |
|      | 3       | 16362     | 1.32E+01      | 53.55  | 109.22 | -167.92  | -60.44     | 71.75      |
| -g   | 1       | 14068     | 1.25E+01      | 48.01  | -80.19 | -168.76  | -62.24     | 163.97     |
|      | 2       | 16688     | 1.38E+01      | 39.05  | -90.14 | -168.75  | -64.14     | 71.35      |
|      | 3       | 10858     | 1.11E+01      | 39.52  | -90.97 | 165.99   | -63.44     | 84.84      |
|      | 4       | 12911     | 1.06E+01      | 32.82  | -87.5  | 161.99   | -61.3      | 62.47      |

Table 5.14 Largest clusters arising from ITZ57 trajectories in DMSO, sampled from 800000 members.

### 5.6.2.1 Clustering in Vacuum

Analysis of the clusters from the vacuum simulations performed for ITZ55 show that very few conformations are sampled. Simulations run from minima A, B and C indicate clustering in the same vicinity of conformational space. The torsion angle $\Phi$ has values ranging from -40° to 40°, while $\Psi$ essentially remains in the +g conformation with a 10° variance. The dihedral values obtained for $\Omega$ indicate that simulations started from each well sampled all three wells during the simulations. The hydroxymethyl group, $\theta_1$ is found to exist in the *gt* and *tg* rotameric orientations for simulations started from A, whereas simulations run from B and C, sample the *gg* rotamer. In each of the simulations, $\sigma$ is found in the *tg* rotamer. The conformations sampled for simulations started from minima D and E (-g conformation) show a reduced torsional variation from simulations A, B and C. $\Phi$ remains at either 30° or 40° respectively. $\Psi$ is consistently in the $-$g conformation, with little or no variation. In simulation D, $\Omega$ shows minimal variation around 180°, whereas in simulation E, $\Omega$ remains constant at 16°. Torsional values $\theta_1$ and $\sigma$ show the same trend as those for simulations started from wells A-C.

Simulations for ITZ56 and ITZ57 appear to have an increased sampling of the allowed conformational space. This may be due to the non-existence of the bulky NAc group present on C2 in ITZ55, thus allowing more rotation around $\Omega$.

Clustering of simulations B and C for ITZ56 reveals sampling of each others minima well's, and also includes sampling of minima A, but clustering of simulation A reveal that only well A is sampled. Similarly, the run from D exhibits sampling of minima E, but not vice versa. For ITZ56 simulations starting from A, B and C, $\Psi$ remains in the +g conformation (with values up to 105° observed for minima C). $\Phi$ ranges from -50° to 50°. $\Omega$ ranges from -60° to 60°.

For simulations D and E, $\Phi$ has values ranging from -170° to 35°, $\Psi$ is primarily in the $-$g conformation (-80° to -105°) and $\Omega$ ranges from -15° and -170°. Indications for the hydroxymethyl groups are that for all minima A through E, $\theta_1$ and $\theta_2$ primarily exist in the *tg* and *gt* rotamers.

ITZ57 clustering reveals that the simulation started from B appears to remain in well B, whereas simulations A and C sample both wells A and C. For simulation B, $\Psi$

87

remains essentially at $100^\circ$. $\Phi$ ranges from $10^\circ$ to $60^\circ$ and $\Omega$ ranges from $-160^\circ$ to $170^\circ$, thus the simulation is 'trapped' in well B. For the simulations from A and C, $\Psi$ is found in the +g conformation, $\Phi$ ranges from $-60^\circ$ to $50''$ and $\Omega$ is found approximately $-40^\circ$ to $-50^\circ$.

Simulations started from D and E revealed that the major clusters were found in well D. $\Psi$ remains in the +g conformation, $\Phi$ remains at $55^\circ$ and $\Omega$ remains at $-155^\circ$. Indications for the hydroxymethyl groups are that for all minima A through E, $\theta_1$ is in the $gt$ rotameric orientations and $\theta_2$ is found as either the $tg$ or $gg$ rotamer. These rotameric orientations exist as a result of the internal hydrogen bonding between hydroxyl groups on C4 (ring2) being in the equatorial position. This trend is repeated in water simulations.

### 5.6.2.2 Clustering in Water

Clusters obtained from water simulations reveal that a larger range of torsion angle space has been sampled over and above that for the vacuum simulations. Simulations started from minima A for ITZ55 reveal that $\Psi$ remains constant at approximately $85^\circ$ (+g), with $\Phi$ ranging from $-30^\circ$ to $50^\circ$ rather than $-40^\circ$ to $40^\circ$ found in the vacuum analysis. $\Omega$ fluctuates from $-40^\circ$ to $160^\circ$. These values give an indication that minimas B and C have been sampled for periods of time when starting from A. The hydroxymethyl group, $\theta_1$, is seen to exist in the $tg$ and $gt$ rotamers, whereas $\sigma$ is restricted to the $tg$ rotamer as there is hydrogen bonding to the water residues which takes preference over the internal hydrogen bonds. However analysis for the simulation started from D shows that only orientations for the $\Phi$, $\theta_1$ and $\sigma$ torsion angles vary from the clustering obtained from the vacuum simulation. The rotation of $\Phi$ has increased to a range of $30^\circ$ to $50^\circ$. $\theta_1$ samples both the $tg$ and $gt$ rotamers and has lost the trans orientation, due to interresidue hydrogen bonding. $\sigma$ follows the same pattern as is obtained from well A. To conclude, wells A, B, D and E were sampled in the water simulations with transitions observed between wells A, B and C.

Clustering for ITZ56 in water reveals that $\Phi$ and $\Omega$ for simulations started from minima C and D are restricted in their rotation, thus resulting in larger populations of respective clusters. The simulation started from C indicates that $\Psi$ remains constant

around 80° (+g), $\Phi$ ranges from -30° to 70° and $\Omega$ fluctuates between -40° to -60°. These values represent that minima A and B have been sampled for periods of time when starting from C. The hydroxymethyl group, $\theta_1$ is seen to exist in the *gt* orientation only, whereas $\theta_2$ exists in both the *tg* and *gt* rotamers. Similarly, the simulation started from D shows transitions to E, giving values showing $\Psi$ remaining constant at -95° (-g) with $\Phi$ ranging from -6° to -55° and $\Omega$ fluctuating from -170° to 40°. Both hydroxymethyl groups have the same distribution as the simulation from minima C.

ITZ57 clustering reveals a similarity to that of ITZ56. $\Phi$ and $\Omega$ for simulations started at B and D are restricted in their rotation, as was found in ITZ56. This is however the only similarity as the simulation started from A shows $\Psi$ constantly around 94°, $\Phi$ ranges from -30° to 4° and $\Omega$ fluctuating between -170° to 170°. These values indicate that minima A and C may not have been sampled when starting from minima B. This is due to steric 1, 4 clashes, where S1 is 'trapped' between O5' and C2' (section 5.3.1). Similarly, simulations started from D reveal that no other minima were sampled. $\Psi$ remains around -85° (-g), $\Phi$ ranges from 36° to 50° and $\Omega$ fluctuates from -170° to 170°. The hydroxymethyl groups, $\theta_1$ and $\theta_2$ are found to exist in the same orientations as was found in the vacuum simulations.

### 5.6.2.3 Clustering in DMSO

The trend set with the clusters obtained from the water simulations of sampling a greater torsion angle space is not repeated in the DMSO simulations. For ITZ55 the range of dihedral values obtained for the simulation from minima B is as follows. $\Phi$ occupies the range from 50° to 70° respectively, with this value increasing from the 50° observed in water. The torsions $\Psi$, $\Omega$ and $\sigma$ appear to remain as they were for the water clusters. $\theta_1$ occupies the *tg* rotamer only. Clusters from well A show no discernable change from those obtained from the water simulations, but for a slight increase in the allowed values for $\Phi$ and $\Omega$ respectively. In summary wells A, B, D, and E were sampled during the simulations

Simulations C and D for ITZ56 reveal larger populations with respect to $\Phi$ and $\Omega$ values but show a restriction in the rotation of the respective torsions. The simulation

started at C appears to have migrated to minima A and populates this well for the majority of the time with $\Psi$ remaining at $85^\circ$ (+g) and $\Phi$ restricted to values between $30^\circ$ to $50^\circ$. $\Omega$ constantly remains at $-60^\circ$. The simulation originating from minima D appears to sample well E for a minimal period of time. The main clusters observed are where $\Psi$ remains in the $-g$ orientation, at approximately $-85^\circ$ with $\Phi$ restricted to values ranging from $-6^\circ$ to $30^\circ$ and $\Omega$ restricted to $-170^\circ$. Not surprisingly, the hydroxymethyl dihedrals $\theta_1$ and $\theta_2$ give the same rotamers for both simulations started from C and D, as was observed for the water simulations.

ITZ57 clustering revealed similar results for $\Omega$ as those obtained in water, for both simulations started from B and D. There was, however, a slight change in $\Psi$, with the dihedral sitting approximately at $107^\circ$ (+g) and $-90^\circ$ (-g) for minima B and D respectively. $\Phi$ appears to fluctuate between $-20^\circ$ to $50^\circ$ for minima B and between $30^\circ$ to $50^\circ$ for minima D. The hydroxymethyl groups for run B and D are found to exist in the same orientations as those found for ITZ56.

### 5.6.3 Molecular Motion Relative to Time

Cluster analysis gives an indication of the preferred conformations sampled during an MD simulation, but is by no means a complete analytical tool, as it can not give an indication of rotational fluctuations or transitions between minima that may have taken place during the simulation.

Fluctuations and transitions between various minima can be determined by analysis of time series plots for the torsional angles used in the cluster analysis (namely, the principal degrees of freedom for the disaccharides). $\Phi$ time series plots for ITZ55 starting from minima B are presented below in Figure 5.12. A full set of time series plots for ITZ55, ITZ56 and ITZ57 starting at selected minima in vacuum, water and DMSO are included in appendix A.

Figure 5.12  Time series plots of $\Phi$ for TZ55 starting from minima C in vacuum (top), water (middle) and DMSO (bottom) for the first 5ns of the MD simulations.

The differences between the simulations in vacuum, water and DMSO are the varying degrees of dihedral rotation. This is observed in the comparison of their respective time series plots. For simulations started from well C in ITZ55, the $\Phi$ time series in vacuum indicates a range of rotation that contains numerous fast transitions between local minima. These transitions are observed to be in the region of 1ps per transition. The $\Phi$ values range from $-70°$ to $50°$ and occasionally values appear in the sterically unfavoured conformation with $\Phi$ around $180°$.

$\Phi$ in the ITZ55 water simulation, has a larger torsional variation, which ranges from $-80°$ to $80°$ and makes an unfavoured transition to $180°$ for a limited period. Transitions between local minima are observed. These appear smoother than those observed in vacuum and are observed to be in the order of 25-30ps. This larger variation and time frame is attributed to the destabilised effect of water on intramolecular hydrogen bonds and the formation of stronger intermolecular hydrogen bonds between the disaccharides and the water molecules.

Simulations run with the aprotic DMSO solvent, give a variation in the $\Phi$ values similar to those observed in water, but with different limits. This variation is observed to be between $-50°$ and $120°$, with transitions found in the order of 10-20ps. This is as a result of the intermolecular hydrogen bonds between the DMSO and the disaccharides in the simulation, which does not appear to model experimental evidence suggesting DMSO should form weaker hydrogen bonds to the disaccharides than water.[16]

A trend of restricted motion in the vacuum over DMSO over water is observed throughout the majority of the remaining one-dimensional time series plots. This however, does not give a holistic indication of the 'true' conformational sampling of the molecules during the simulations and for this reason two- and three-dimensional plots are required. Figure 5.13 shows the two- and three-dimensional plots for the simulations of ITZ55 started at minima B, in vacuum, water and DMSO respectively.
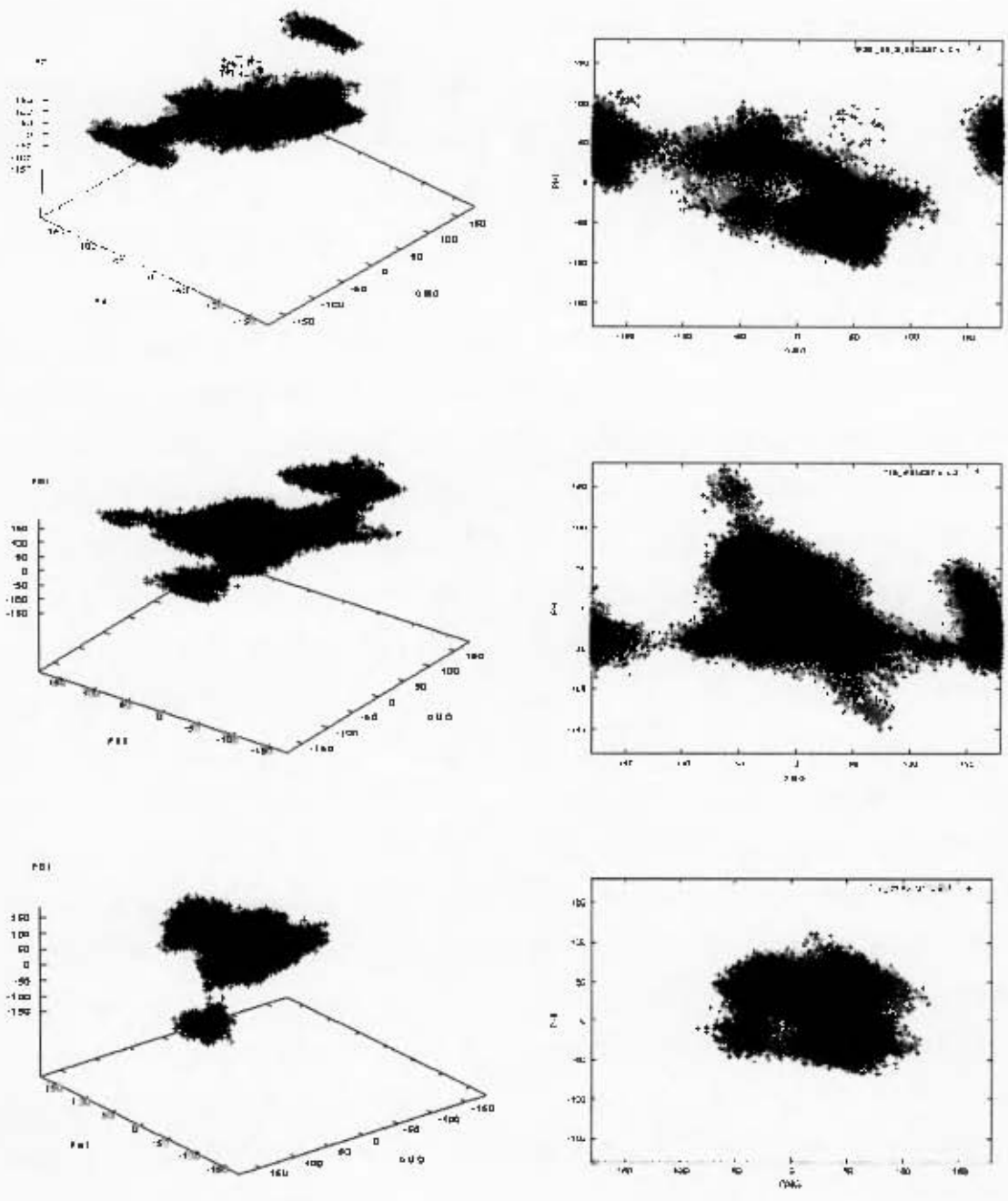
Figure 5.13 Two- and Three-Dimensional time series plots for ITZ55 starting from minima B in vacuum (top), water (middle) and DMSO (bottom).

Figure 5.13 indicates a commonality between the different simulation conditions. $\Psi$ remains primarily in the +g conformation ($90^{\circ} \pm 40^{\circ}$) in all three simulations, with no transitions to the –g conformer observed. The same observation can be made in all simulations conducted on the three disaccharide mimetics and starting with $\Psi$ in the +g and –g conformation.

The 2-dimensional (2D) and 3-dimensional (3D) plots for simulations conducted on ITZ55 starting from minima C (Figure 5.10) indicate that for all simulation conditions used, minima A, B and C found in the simulated annealing were sampled. Closer analysis of the 2D $\Phi$ and $\Omega$ plots reveals that the water simulation samples a far greater region than the vacuum, which in turn samples a greater $\Phi$, $\Omega$ region than the DMSO.

Simulations started from minima D reveal that no transitions to minima E are observed in either the vacuum or solution simulations, but repeats the trend in the ratio of the areas sampled in the simulations started from minima A, such that the water simulation samples a greater region than the vacuum simulation, which is greater then that sampled in DMSO.

The vacuum simulation from minima E in ITZ55 again shows no transition to minima D. For ITZ56, simulations started from minima C, in vacuum, water and DMSO are observed to sample minima A and B. The trend in the conformational area sampling is found to be the same as that observed for ITZ55 simulations. As with ITZ55 all simulations started with $\Psi$ in the –g conformation are observed to be restricted to the minima in which the simulation was started but with a similar trend in the areas sampled in vacuum and solution. Analysis of ITZ57 reveals that solution simulations started from minima B and D do not sample minima A, C and E respectively. However the water simulations spread over a larger region than that observed for DMSO.

A secondary factor is the conformational role played by the respective hydroxymethyl and NAc groups on the disaccharide mimetics. Figure 5.11 below shows the one-dimensional time series plots for the hydroxyl methyl group $\theta_1$ for ITZ55, from the vacuum, water and DMSO simulations started from minima A. A full set of the primary alcohol time series plots are available in appendix B
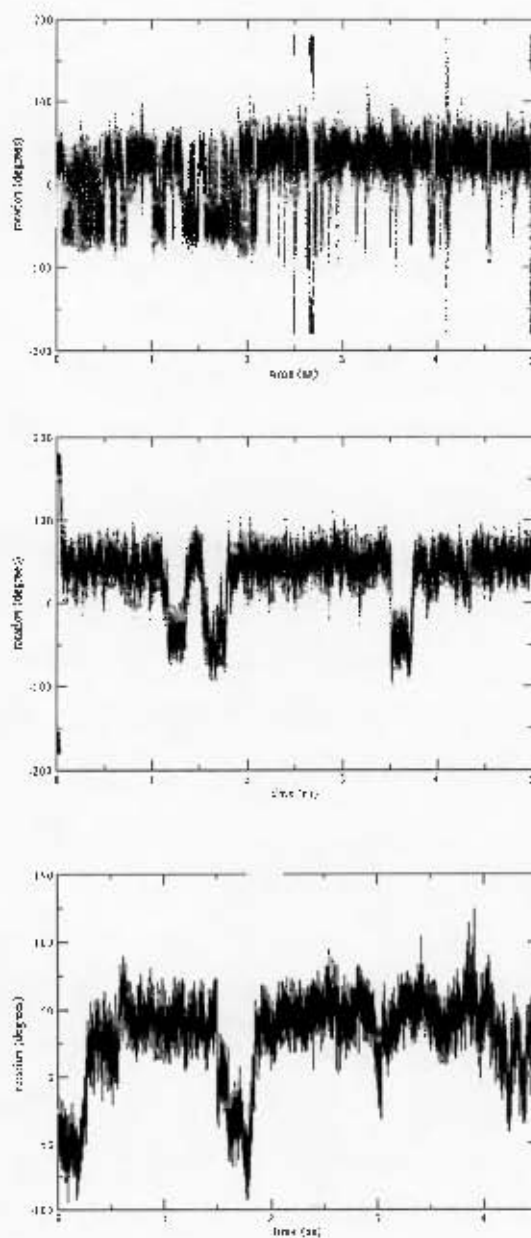
94
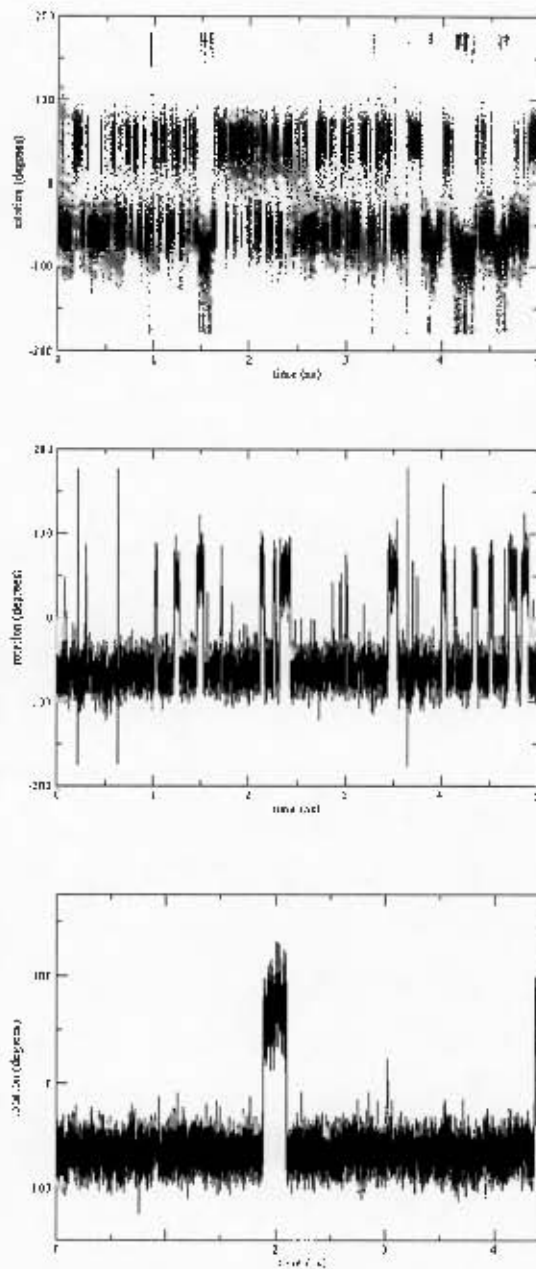
Figure 5.14 Time series plots of $\theta_1$ for ITZ55 starting from minima B in vacuum (top), water (middle) and DMSO (bottom) for the first 5ns of the MD simulations.

The hydroxymethyl time series plots for all three disaccharide mimetics show similar trends. In vacuum there are numerous transitions between the $tg$, $gt$ and $gg$ rotameric orientations with very short life-time. In water and DMSO the $tg$ and $gt$ orientations are observed with the $tg$ orientation being preferred, thus indicating hydrogen bonding to the solution molecules. In DMSO the transitions are of the magnitude of 30 ps, whereas in water they are between 10 ps and 20 ps.

### 5.6.4 Hydrogen Bonding

The ability of a water molecule to form hydrogen bonds with other water molecules and solutes such as disaccharides, leads to a spatial distribution of water around a solute, such as a disaccharide.[18] Similarly hydrogen bonds have been shown to occur between DMSO solvent molecules and disaccharides, but the interactions (bonds) are found to be much weaker, thereby allowing intramolecular hydrogen bonding to dominate over intermolecular hydrogen bonds.[19] For this thesis the criteria used for the hydrogen bond analysis are:

1. An angle of $120^\circ$ to $180^\circ$ for the acceptor-donor-acceptor angle, and
2. an acceptor donor distance of less than 2.4Å.

Criterion 1 was relaxed to $100^\circ$ for the determination of intramolecular hydrogen bonds due to the strained geometry of the participating groups regions. It must be remembered that caution must be taken in calculating the life-time of hydrogen bonds due to the low frequency at which data is saved during the simulation runs. This results in distance and angle changes which may exceed the above criterion and result in erroneous times.

Presented below in Tables 5.15 and 5.16 is the hydrogen bonding data for the ITZ55 water and DMSO simulations starting from minima A obtained using the HBOND function in CHARMM and performed over the entire MD simulations. (Hydrogen bonding results for ITZ55-57 not included in the text are available in appendix A) Both donor and acceptor hydrogen bonds have been included in the analysis where applicable.

| Hydroxyl | Internal | To Water |
|---|---|---|
| 1 OH2 | 0.06 | 0.56 |
| 1 OH3 | 0.13 | 0.49 |
| 1 OH4 | 0.10 | 0.64 |
| 1 O5 | 0.01 | 0.42 |
| 1 OH6 | 0.02 | 0.80 |
| 2 OH3 | 0.10 | 0.70 |
| 2 OH4 | 0.07 | 0.70 |
| 2 O5 | 0.01 | 0.73 |
| 2 OH6 | 0.01 | 0.26 |

Table 5.15 Hydrogen bonding data obtained for ITZ55 from minima A water simulations

The hydrogen bonding data for all the water simulations conducted on ITZ55 through to ITZ57 show similar values for intramolecular and intermolecular hydrogen bonds. Statistically there is a lowering of the number of internal hydrogen bonds from approximately 4.0 in the vacuum simulations to 1.2 internal hydrogen bonds in water. This indicates that there are additional hydrogen bonds to the water and that these have displaced a number of internal hydrogen bonds.

The majority of the hydroxyls form between 0.50 and 0.80 hydrogen bonds with water and between 0.06 to 0.13 intramolecular hydrogen bonds. In ring 1 and ring 2 the hydroxyl OH3 forms the highest number of intramolecular hydrogen bonds as it is between two ring hydroxyls. Interestingly the highest number of hydrogen bonds to water is observed for OH6 in ring 1. Surprisingly OH6 in ring2 has the lowest, and does not follow the pattern observed in the vacuum simulations. This is not repeated throughout the water simulations in which OH6 values increase. The total number of hydrogen bonds throughout all the water simulations is around 4.5, including the intramolecular and intermolecular hydrogen bonds.

Hydrogen bonding in DMSO as seen for ITZ55 in Table 5.16, indicates that there are stronger interactions, with long life-times. This appears to be a false reading. One would expect weaker hydrogen bonds of short life-times between the disaccharides and DMSO due to the size of the DMSO molecules and therefore their lack of ability to approach and get close enough to form hydrogen bonds with the disaccharides. This

97

should result in an increase in the internal hydrogen bonding, but this too does not appear to happen. From the simulations the total number of hydrogen bonds is observed to be approximately 7.0 for each of ITZ55 to ITZ57. The pattern observed for ITZ55 shows that O5 on both rings has the lowest number of hydrogen bonds and that all of OH2, OH3, OH4 and OH6 have relatively the same number of hydrogen bonds.

| Hydroxyl | Internal | To DMSO |
|----------|----------|---------|
| 1 OH2 | 0.06 | 0.95 |
| 1 OH3 | 0.04 | 0.97 |
| 1 OH4 | 0.08 | 0.92 |
| 1 O5 | 0.00 | 0.42 |
| 1 OH6 | 0.01 | 0.99 |
| 2 OH3 | 0.05 | 0.98 |
| 2 OH4 | 0.06 | 0.93 |
| 2 O5 | 0.01 | 0.36 |
| 2 OH6 | 0.04 | 0.96 |

Table 5.16 Hydrogen Bonding data obtained for ITZ55 from minima A in DMSO

### 5.6.5 Solvent Structuring

The difference in conformational space traversed between the solution and vacuum simulations of the three diglycosyl disulphides is evident from the three-dimensional time series plots. It is observed that water allows for a much larger range of conformational space to be explored than that in the DMSO simulations, and the vacuum is even further restricted. To explain this effect, the solution structures of the diglycosyl disulphides in both water and DMSO have been studied using one-dimensional pair distribution functions (PDF, chapter 3.4.2.1) and three-dimensional spatial distribution functions (SDF, chapter 3.4.2.1).

### 5.6.5.1 Pair Distribution Function (PDF)

A one-dimensional PDF gives information regarding the composition (number of atoms and nearest neighbours' positions) of the solvation spheres that are formed when the

disaccharides are simulated in water and DMSO. This information is however limited, as the PDF's are radially averaged.

Analysis of the one-dimensional PDFs in water for ITZ55, ITZ56 and ITZ57 minima shows a great number of similarities, with some minor variations. The same applies for the DMSO PDF analysis. A discussion of the water and DMSO PDFs is therefore provided below.

### 5.6.5.1.1 Water PDFs

From the analysis of the PDFs for the three diglycosyl disulphides in water, it is observed that all of the free hydroxyl groups have a first maximum peak at approximately 2.80 Å, which gives an indication of the of the average distance of the first solvation sphere from each respective free hydroxyl group. The ring oxygens' have peaks at >5 Å, which are less defined and have a far less effect on the structuring of the water, as was found in the previous section on hydrogen bonding. The details above are summarised in Figure 5.15 and Table 5.17 below.
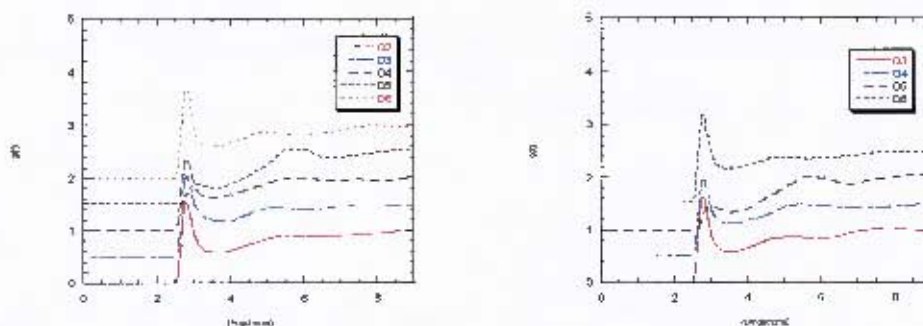


Figure 5.15 Hydroxyl-water PDFs for minima A from ITZ55 with ring 1 (left) and ring2 (right), offset by 0.5 per plot.

| | Ring 1 | | | Ring 2 | | |
|---|---|---|---|---|---|---|
| Hydroxyl | $r_{max}$ | $G(r_{max})$ | Integral | $r_{max}$ | $g(r_{max})$ | Integral |
| O2 | 2.801 | 1.535 | 2.81 | | | |
| O3 | 2.801 | 1.522 | 4.13 | 2.801 | 1.607 | 3.02 |
| O4 | 2.801 | 1.392 | 2.82 | 2.801 | 1.432 | 3.19 |
| O5 | 5.652 | 1.046 | 1.04 | 5.572 | 1.053 | 1.26 |
| O6 | 2.763 | 1.670 | 2.82 | 2.801 | 1.643 | 2.89 |

Table 5.17 Radial distances of the first peak, in the saccharide-O-$H_2O$ PDF for ITZ55 simulations from minima A. Oxygen density integrals are from 0-3.8 Å

One of the most useful bits of information that can be obtained from a PDF is the estimate of the "first shell" coordination number for each oxygen atom in the disaccharide. This is obtained by integrating the PDF over a sphere which has the oxygen being integrated as the centre. The radius of the sphere is equal to the distance at which the first minima in the strongly structured PDFs are found to occur. This was found to be at 3.72 Å.

The total number of waters in the first shell determined in this manner is greater than that determined via hydrogen bonding analysis. The total coordination numbers determined by integration are observed to range between 2.82 to 4 for the hydroxyl groups and just over 1 for all the other oxygens.

### 5.6.5.1.2 DMSO PDFs

The PDFs of the three diglycosyl disulphides in DMSO follow similar trends to those observed in water. All the free hydroxyl groups have a first maximum peak at approximately 2.677 Å, except for O4 on ring 2 in ITZ56 starting from minima A, which has a value of 5.388 Å and all the free hydroxyls are observed to have varying heights.

In DMSO, as with the water, the ring oxygens' have peaks at >5 Å, thus having less effect on the structuring of the aprotic DMSO. These details are summarised in Figure 5.16 and Table 5.8 below.
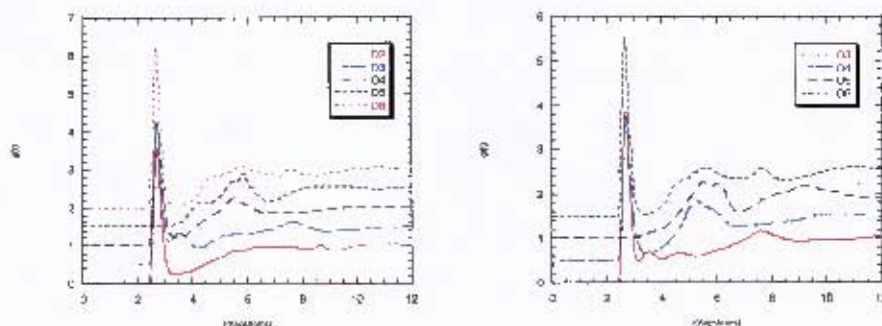
Figure 5.16 Hydroxyl-DMSO PDFs from minima B with ring 1 (left) and ring 2 (right) from ITZ55, offset by 0.5 per plot.

| Hydroxyl | Ring 1 | | | Ring 2 | | |
|---|---|---|---|---|---|---|
| | $r_{max}$ | $G(r_{max})$ | Integral | $r_{max}$ | $g(r_{max})$ | Integral |
| O2 | 2.677 | 3.559 | 1.00 | | | |
| O3 | 2.677 | 3.801 | 1.04 | 2.677 | 3.823 | 0.98 |
| O4 | 2.677 | 3.226 | 0.97 | 2.677 | 3.259 | 0.95 |
| O5 | 5.806 | 1.405 | 7.75 | 5.458 | 1.271 | 7.85 |
| O6 | 2.677 | 4.193 | 0.95 | 2.677 | 4.033 | 0.92 |

Table 5.18 Radial distances of the first peak, in the saccharide-O-DMSO PDF for ITZ55 (simulations from minima B). Oxygen density integrals are from 0-3.4 Å

The "first shell" coordination number for the DMSO simulations was conducted in the same way as for water and was calculated using a sphere radius of 3.31 Å. As with the PDFs for water the total number of DMSO in the first shell determined in this manner is greater than that determined via hydrogen bonding analysis. The total coordination numbers determined via integration are observed to range between 0.9 to 1.1 for the hydroxyl groups and approximately 7.8 for O5. This indicates that there is a lower structuring effect and that there is limited motion of the disaccharides in the DMSO, thus explaining the difference in the hydrogen bonding in the water and DMSO simulations.

### 5.6.5.2 Spatial Distribution Function (SDF)

The spatial orientation of the solvent in relation to the solute can be determined by obtaining the three-dimensional spatial distribution function. This gives three-dimensional information regarding the time-averaged water density and can be used to interpret the results from hydrogen bonding and the one-dimensional PDFs.

SDFs for each cluster from ITZ55, ITZ56 and ITZ57 in both water and DMSO were calculated using a best RMS fit with the cluster centre structure and a cutoff RMS distance was chosen to give an approximate number of members when compared to the cluster sizes obtained from the cluster analysis. The RMS fit was restricted to the dihedrals constituting the interglycosidic linkages, as the free rotation of the hydroxyl groups allows for numerous positions to be obtained and the SDF only shows one. In addition the SDFs were calculated using only the oxygen atoms of the water and DMSO solvent molecules, thus calculating only the oxygen density and limiting the ability to conduct a complete hydrogen bonding analysis.

Presented below in Figure 5.17 are the three dimensional contour plots for water and DMSO simulations for ITZ55 to ITZ57. The water density distributions show that for each of the mimetics their respective ring 1's have a similar water density pattern, whereas each has a slightly altered pattern of water density around their respective second rings (ring 2). This is due to the 2-Nac functional group in ITZ55, and the changes in orientation of the hydroxyls on C2 and C4 in ITZ56 and ITZ57 respectively. Similarly, the isosurfaces appear to be elongated perpendicular to the plane of the ring with the highest densities being close to the equatorial planes. This elongation can be attributed to the hydrogen bond donor atoms. The water density peaks are generally located between the hydroxyls, suggesting bridging hydrogen bonds.
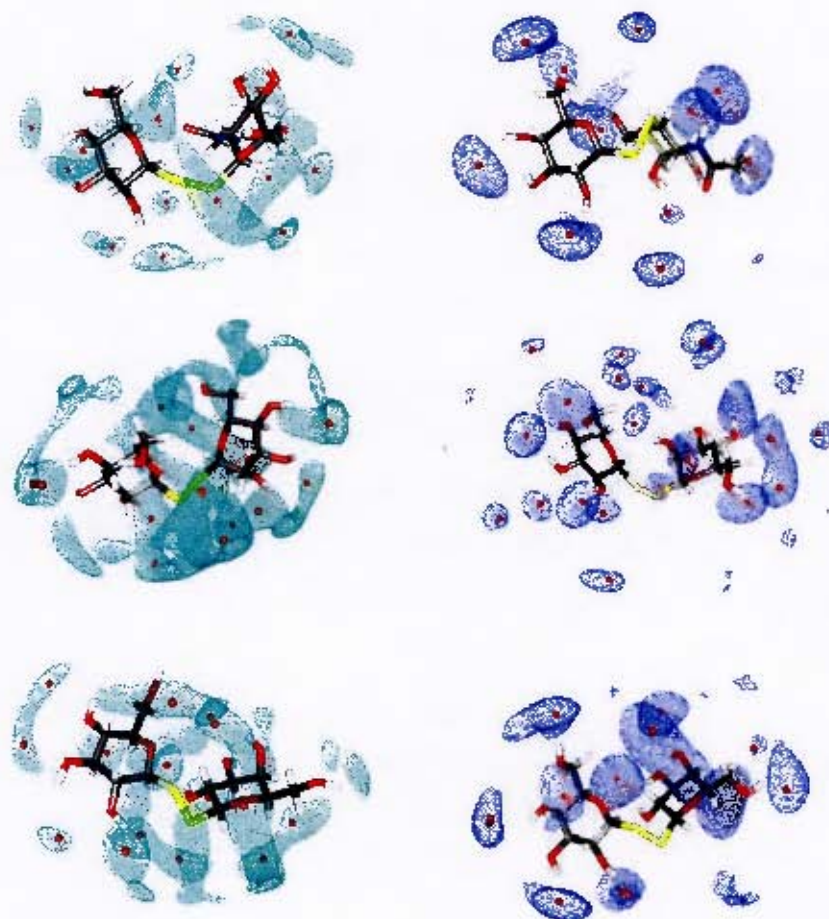
Figure 5.17    Three dimensional contour plots for water (left) and DMSO (right) simulations for ITZ55 (top) from minima A, ITZ56 (middle) from minima C and ITZ57 (right) from minima B. All are plotted at 65% above bulk density.

Closer examination of the disaccharides in water suggests that ITZ56 and ITZ57 have a stronger capacity for structuring water as several of the water peaks are seen to be further than 3.4 Å away from the nearest disaccharide hydroxyl. This may be as a result of secondary hydrogen bonding. In all three mimetics the sugar rings are a fair distance apart due to the increase in the length of the disulphide interglycosidic linkage.

This would usually lead to a lowering of the potential of bridge bonding between the rings, but this is still observed between the OH6 of both ring 1 and ring2 in the disaccharides.

The DMSO SDFs suggest a more rigid conformation is adopted, with less movement of the disaccharides, and shows highly directed hydrogen bonding with all the hydroxyls, as well as the 2-NAc group, thus complimenting the hydrogen bonding and the PDF analysis. Limited elongation perpendicular to the ring plane is observed in ITZ56 and ITZ57, both for ring 2. This data suggests that the DMSO simulation shows limited diffusion of the molecule and is inconsistent with experimental findings for disaccharides in DMSO,[19] in the nature of the directed hydrogen bonds, thus restricting the internal mobility of the torsional angles present in the glycosidic linkage. This is discussed further in chapter 6, along with the details of the experimental NMR spectral parameters.

# References

1. MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, III, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B*, 1998, *102*, 3586.

2. Kuttel, M.M; Brady, J. W.; Naidoo, K. J. *J. Comp. Chem.*, 2002, *23*, 1236.

3. Woods, R. J.; Dwek, R. A.; Edge, C. J. *J. Phys. Chem.*, 1995, *99*, 3832.

4. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comp. Chem.*, 1997, *18*, 1955.

5. Brooks, B. R.; Bruccoleri, R. E; Olafson, B. D.; States, D. J; Swaminathan, S.; Karplus, M. *J. Comp. Chem.*, 1983, *4(2)*, 187.

6. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A.; Gaussian 98, Revision A.7; Gaussian, Inc., Pittsburgh PA, 1998.

7. Naidoo, K. J.; Brady, J. W *J. Chem. Phys.*, 1997, *224*, 263.

8. Szilágyi, L.; Illyés, T.-Z., Herczegh, P. *Tetrahedron Letters* 2001, *42*, 3901.

9. Davis, B. G.; Ward, S. J.; Rendle, P. M. *Chem. Commun.*, 2001,189

10. Bader, R. F. W. *Acc. Chem. Res.*, 1985, *18*, 9.

11. Foster, J. P; Weinhold, F. *J. Am. Chem. Soc.*, 1980, *102*, 7211.

12. AIM2000; *J. Comp. Chem.*, **2001**, *22*, 545.

13. Chakka, N.; Johnston, B. D.; Pinto, B. M. *Can J. Chem.*, **2005**, *83*, 929.

14. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D; Impey, R. W.; Klein M. L. *J. Chem. Phys.*, **1983**, *79*, 926.

15. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.*, **1995**, *103*, 4613.

16. Strader, M. L.; Feller, S. E. *J. Phys. Chem. A*, **2002**, *106*, 1074.

17. Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.*, **1975**, *97*, 1354.

18. Best, R. *MSc Thesis*, Chemistry Department; University of Cape Town, Cape Town, **2000**.

19. Shim, G; Shin, J; Kim, Y. *Bull. Korean Chem. Soc.*, **2004**, *25*, 198.

# Chapter 6

# Conformational Spectral Parameters and Molecular Relaxation

## 6.1 Introduction

The primary NMR spectral parameters related to the molecular geometries of disaccharides in general are the NOE's and the coupling constants (chapter 1.6.1).

Presented below are the NMR techniques applied and the respective NOE and scalar spin-spin couplings related to ITZ55, ITZ56 and ITZ57, calculated experimentally and from the simulations performed in chapter 5. These results were analysed in order to determine if correlations existed between the experimental and computational techniques. This is done in order to determine the applicability and accuracy of the force fields used in the simulations.

In addition molecular relaxation times ($T_1$ and $S^2$, chapter 4.5), and correlation time functions have been determined using the calculated MD simulations. This was performed in order to further determine the applicability of MD simulations and the force fields applied.

## 6.2 NMR Experimental Methods

NMR experiments were run on a Bruker Avance DRX-500 spectrometer and were conducted by the Szilágyi Group at the University of Debrecen, Hungary.

10-15mg of the samples was dissolved in DMSO-$d_6$ and the probe temperature was set to 300K. 2D NOE (NOESY) and rotating frame NOE (ROESY)[1, 2] spectra in phase sensitive mode (TPPI) were obtained using standard Bruker pulse programs. A CW spinlock field of 3.57 kHz strength was used for the ROESY experiments.

HOHAHA effects[2, 3] and J-relayed cross peaks [4, 5] were identified by repeating the experiment using a different offset frequency for the spinlock field.[6] The raw datasets typically consisted of 2KH512 complex data points. The cross peak intensities were determined by volume integration from the baseplane corrected spectra. The individually assigned cross peaks were checked for the absence of the artefacts above before converting them into distances using the Isolated Spin Pair Approximation (ISPA).[7]

The distances between 1, 3 and 1, 5 diaxial protons within the glycopyranosyl rings (with $^4C_1$ chair geometry [8, 9]) were used for calibration of the integrals. Long range proton-carbon coupling constants were determined using a sensitivity enhanced gradient selected $^{13}C$-filtered TOCSY pulse sequence and HSQMBC experiments.[10]

# 6.3 Experimental and Simulation NOE Results

Interannular NOEs (i.e. between protons on the sugar rings) from NMR experiments were obtained using the NOE intensities between the 1, 3-diaxial protons in the glucopyranosyl rings as reference. These NMR distances measured ~ 2.46 Å in distance for each ring, while the corresponding value from the simulations is ~2.40 Å for each of the disaccharides in the simulations run.

Time averaged NOE results for the diglycosyl disulphide mimetics from experiment and simulations in DMSO are presented below in Table 6.1. No interannular NOE were detected for ITZ57 via NMR methods, due to the almost exact overlap of the mannose (M1) and glucose (G1) resonances which precludes detection of the characteristic NOEs between anomeric protons that can clearly be seen in ITZ55 and ITZ56. Simulation NOEs for ITZ57 were calculated and have been included in appendix B, but are not discussed here.

| Compound | NOE assignment | $^1$H/$^1$H distance (Å) | | | |
|---|---|---|---|---|---|
| | | NMR | MD (+g) | MD (-g) | MD (average) |
| 1 (ITZ55) | N1 / G1 | 3.4 | 4.4 ± 0.3 | 4.0 ± 0.5 | 4.2 |
| | N1 / G2 | 3.7 | 6.7 ± 0.5 | 5.2 ± 0.5 | 6.0 |
| | N1 / G-OH2 | 4.2 | 7.1 ± 0.4 | 6.3 ± 0.5 | 6.7 |
| | N-NH / G-OH6 | 3.7 | 9.3 ± 0.7 | 7.9 ± 1.4 | 8.6 |
| | N-Me / G2 | 4.5 | 8.0 ± 0.6 | 8.2 ± 1.4 | 8.1 |
| | N-Me / G-OH6 | 5 | 8.7 ± 1.0 | 8.2 ± 1.5 | 8.5 |
| 2 (ITZ56) | M1 / G1 | 3.7 | 3.4 ± 0.6 | 4.2 ± 0.3 | 3.8 |
| | M1 / G2 | 4.5 | 5.0 ± 0.4 | 4.0 ± 0.44 | 4.5 |
| | M1 / G-OH2 | 3.8 | 6.0 ± 0.6 | 5.8 ± 0.4 | 5.9 |
| | M1 / G-OH6 | 4.2 | 4.7 ± 1.6 | 4.0 ± 1.0 | 4.4 |
| | M2 / G2 | 4.3 | 4.9 ± 0.5 | 3.6 ± 0.6 | 4.3 |
| | M2 / G4 | 4.8 | 5.9 ± 0.7 | 3.5 ± 0.7 | 4.7 |

Table 6.1 Interannular NOE derived distances from NMR from an ensemble average based on $r^6$ averages and directly calculated NOE distances from MD simulations in DMSO

The values obtained from the MD simulations for ITZ55 are not in good agreement with those obtained from NMR experiments. This is believed to be as a result of the orientation of the NAc functional group being 'fixed' throughout the simulation, as a consequence of the original parameterisation of the force field. In addition, the restricted motion observed about the $\Omega$ torsion angle may be as a result of the use of the original C-S torsion parameters from the CHARMM22 force field.[11] To correct this, reparametrisation of the C-S torsion parameters may be required.

However in the case of ITZ56 good agreement is obtained for all the NOEs, excluding the M1/G-OH6 NOE. This is true for both simulations started in the +g and –g conformations. Interestingly, if the average of the two simulations is calculated, the average NOEs are found to be within 0.2 Å of the NMR values obtained.

This indicates that the parameters set up for the sulphur-sulphur linkage are excellent for un-substituted saccharide rings.

# 6.4 Experimental and Simulation J-Coupling Results

In principle it is possible to determine interglycosidic torsion angles by measuring the three-bond carbon-carbon couplings across the S-S-linkage (C1-S-S-C1'). Natural $^{13}C$ abundances presently restrict such measurements. Karplus-type dependence of these couplings has been demonstrated both theoretically and experimentally,[12] application to carbohydrates is still problematic because appropriate parameters are not yet available.[13-14] Therefore to collect experimental data on the conformational features of the molecules under study we have relied on $^{1}H$-$^{1}H$ NOESY and/or ROESY measurements to reveal internuclear distances suitable for closer scrutiny of molecular shapes.

Coupling constants were obtained from simulations, this was achieved by averaging the coupling constants over the MD trajectory and by using Karplus equations with Fraser and Tvaroska parametrisation as shown in chapter 4.2.1. The $^{1}H$-$^{1}H$ and $^{1}H$-$^{13}C$ vicinal coupling values for ITZ55, ITZ56 and ITZ57 are included in Tables 6.2 to 6.6. Unfortunately no $^{1}H$-$^{13}C$ proton couplings were obtained experimentally for ITZ56.

| ITZ55 | NMR | MD (+g) | MD(-g) |
|---|---|---|---|
| GlcNAc-NH/C1 | 0.3 | 1.1 ± 0.6 | 0.8 ± 0.6 |
| GlcNAc-NH/C3 | 0.8 | 1.0 ± 0.6 | 0.8 ± 0.6 |
| GlcNAc-OH3/C4 | 1.0 | 1.12 ± 0.6 | 1.2 ± 0.6 |
| GlcNAc-OH6/C5 | 1.5 | 1.3 ± 0.6 | 1.3 ± 0.6 |
| Glu-OH2/C1 | 2.8 | 2.37 ± 1.4 | 2.68 ± 1.6 |
| Glu-OH2/C3 | 0.0 | 1.11 ± 0.6 | 1.11 ± 0.6 |
| Glu-OH4/C3 | 2.1 | 2.32 ± 1.4 | 2.46 ± 1.2 |
| Glu-OH4/C5 | 3.2 | 3.65 ±2.2 | 3.23 ±2.4 |
| Glu-OH6/C5 | 2.8 | 2.33 ± 1.4 | 2.68 ± 1.4 |

Table 6.2 $^{1}H$-$^{13}C$ vicinal coupling values in Hz obtained from experimental NMR studies and MD simulations run for ITZ55 in the +g and -g conformations. The $^{1}H$-$^{13}C$ coupling obtained from NMR experiments has a standard deviation of 0.3 Hz.

| ITZ57 | NMR | MD (+g) | MD(-g) |
|---|---|---|---|
| Glu-OH2/C1 | 3.0 | 2.8 ± 0.6 | 3.4 ± 1.0 |
| Glu-OH2/C3 | 2.0 | 1.6 ± 1.2 | 2.4 ± 0.8 |
| Glu-OH3/C2 | 1.6 | 1.4 ± 0.8 | 1.4 ± 0.6 |
| Glu-OH3/C4 | 2.0 | 1.8 ± 0.4 | 2.3 ± 0.4 |
| Glu-OH4/C3 | 2.3 | 2.5 ± 0.1 | 2.1 ± 0.4 |
| Glu-OH4/C5 | 3.0 | 2.7 ± 0.5 | 3.6 ± 0.4 |
| Glu-OH6/C5 | 2.6 | 2.4 ± 0.8 | 2.6 ± 1.2 |
| Gal-OH2/C1 | 3.3 | 3.0 ± 0.4 | 3.4 ± 0.8 |
| Gal-OH2/C3 | 2.6 | 2.2 ± 0.2 | 2.1 ± 0.5 |
| Gal-OH3/C2 | 2.0 | 2.2 ± 0.6 | 2.2 ± 0.6 |
| Gal-OH3/C4 | 3.0 | 2.6 ± 1.2 | 2.9 ± 0.8 |
| Gal-OH4/C3 | 1.6 | 1.1 ± 0.6 | 1.8 ± 0.5 |

Table 6.3 $^1$H-$^{13}$C vicinal coupling constants in Hz obtained from experimental NMR studies and MD simulations run for ITZ57 in the +g and –g conformations. The $^1$H-$^{13}$C coupling obtained from NMR experiments has a standard deviation of 0.3 Hz.

| ITZ55 | NMR | MD (+g) | MD(-g) |
|---|---|---|---|
| GlcNAc-NH | 9.5 | 7.9 ± 0.3 | 10.2 ± 1.4 |
| GlcNAc-OH3 | 5.4 | 8.4 ± 0.4 | 6.2 ± 0.6 |
| GlcNAc-OH6 | 5.6 | 8.2 ± 0.6 | 6.0 ± 0.8 |
| Glu-OH2 | 5.9 | 5.4 ± 0.8 | 5.8 ± 0.8 |
| Glu-OH4 | 5.0 | 5.4 ± 0.8 | 5.2 ± 0.6 |
| Glu-OH6 | 5.1 | 5.0 ± 0.8 | 4.8 ± 1.2 |

Table 6.4 $^1$H-$^1$H vicinal coupling constants in Hz obtained from experimental NMR studies and MD simulations run for ITZ55 in the +g and –g conformations. The $^1$H-$^1$H coupling obtained from NMR experiments has a standard deviation of 0.1 Hz.

| ITZ56 | NMR | MD (+g) | MD(-g) |
|---|---|---|---|
| Glu-OH2 | 5.9 | 6.3 ± 1.4 | 5.6± 1.2 |
| Glu-OH3 | 4.8 | 4.6 ± 0.8 | 5.4± 0.7 |
| Glu-OH4 | 5.4 | 5.3 ± 0.5 | 5.2 ±1.1 |
| Glu-OH6 | 5.7 | 6.3± 0.7 | 6.3± 0.9 |
| Man-OH2 | 5.4 | 5.2 ± 1.2 | 5.1 ± 0.8 |
| Man-OH3 | 5.4 | 5.4± 0.9 | 5.7 ± 1.2 |
| Man-OH4 | 5 | 4.8 ± 1.0 | 4.6 ± 0.6 |
| Man-OH6 | 5.7 | 5.8 ± 1.4 | 5.8± 1.3 |

Table 6.5 $^1$H-$^1$H vicinal coupling constants in Hz obtained from experimental NMR studies and MD simulations run for ITZ56 in the +g and –g conformations. The $^1$H-$^1$H coupling obtained from NMR experiments has a standard deviation of 0.1 Hz.

| ITZ57 | NMR | MD (+g) | MD(-g) |
|---|---|---|---|
| Glu-OH2 | 5.7 | 6.0 ± 1.4 | 5.3 ± 0.4 |
| Glu-OH3 | 4.8 | 5.3 ± 1.0 | 4.6 ± 1.0 |
| Glu-OH4 | 5.3 | 6.0 ± 0.7 | 5.5 ± 0.3 |
| Glu-OH6 | 5.6 | 5.2 ± 0.6 | 6.0 ± 1.2 |
| Gal-OH2 | 6.1 | 6.0 ± 1.1 | 5.8 ± 0.8 |
| Gal-OH3 | 5.8 | 6.0 ± 2.1 | 6.0 ± 1.0 |
| Gal-OH4 | 5 | 4.5 ± 1.1 | 5.3 ± 0.6 |
| Gal-OH6 | 5.6 | 5.8 ± 1.4 | 6.0 ± 0.9 |

Table 6.6 $^1$H-$^1$H vicinal coupling constants in Hz obtained from experimental NMR studies and MD simulations run for ITZ57 in the +g and –g conformations. The $^1$H-$^1$H coupling obtained from NMR experiments has a standard deviation of 0.1 Hz.

$^1$H-$^{13}$C and $^1$H-$^1$H vicinal coupling constants obtained from the MD simulations for ITZ55 are in poor agreement with those obtained from NMR experiments. Whereas the coupling constants for ITZ56 and ITZ57 show good agreement with those obtained from NMR. This data supplements the observations made in chapter 6.3. In reaching the conclusion that the parameters set up for the sulphur-sulphur linkage are functional when working with non-substituted saccharide rings and that the original parametrisation of the NAc functional group, is not applicable to the diglycosyl disulphide mimetics.

# 6.5 Molecular Relaxation: Correlation Functions

### 6.5.1 Dihedral Time Correlation Functions

Correlations of the fluctuations in torsional angles, i.e. $\alpha$, are given by $C(t) = \langle \Delta\alpha(0).\Delta\alpha(t) \rangle$ gives an indication of the scope of the motion of the torsion angle correlated with the same torsion at any time $t$. The rate of decay of these correlated fluctuations gives a clear indication of the rate of the torsional fluctuations and can clearly be observed in correlation plots, as shown below in Figures 6.1 and 6.2. These figures represent the rate of decay of the interglycosidic torsion angles (Figure 6.1) and the primary alcohol (N-Acetyl) torsion angles (Figure 6.2) for ITZ55 in the +g conformation from water and DMSO simulations. Correlation plots for ITZ55 in the −g conformation, as well as those of ITZ56 and ITZ57 are included in appendix B.

Figure 6.1 and Figure 6.2 indicates the loss of correlation that takes place in both the water and DMSO simulations for interglycosidic and N-Acetyl torsions of the +g conformer in ITZ55. The time taken for the correlation functions to decay to approximately zero varies for each torsion angle and dissacharide. The loss of correlation in both the water and DMSO for the interglycosidic torsions: $\Psi$, $\Omega$ and $\sigma$ ranges from 250ps − 500ps. The rates observed are substantially slower than those observed in vacuum, which can be attributed to the 'dampening effect' of the solvent molecules, thus restricting the rate of torsional motion.
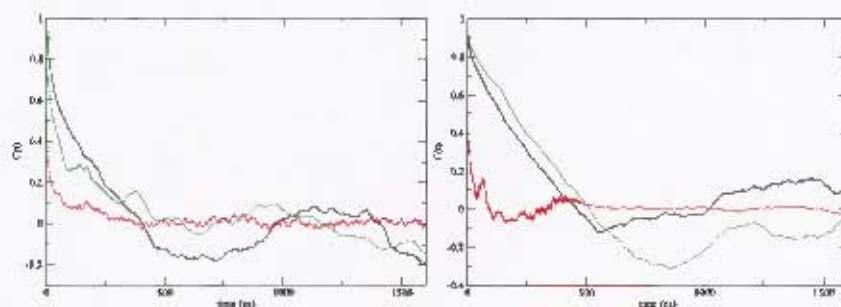
Figure 6.1 Time correlation functions for the interglycosidic torsions in ITZ55 from water (left) and DMSO (right) simulations with the disaccharide starting in the +g conformation. Representing $\Phi$ (black), $\Psi$ (red) and $\Omega$ (green).

.

The rate of decay for the hydroxymethyl torsions occur within 10ps of the simulations in water, but only after 150-200ps in DMSO. Whereas, the loss of correlation of the NAc torsion occurs at around 400ps for both water and DMSO.

In summary, both the water and DMSO simulations exhibit most rapid loss of correlation of the hydroxymethyl torsions due to the lack of connectivity associated with the oxygen bond termini and for the interglycosidic torsions angles, generally $\Psi$ has the most rapid loss of correlation, whereas $\Phi$ and $\Omega$ have similar losses of correlation.
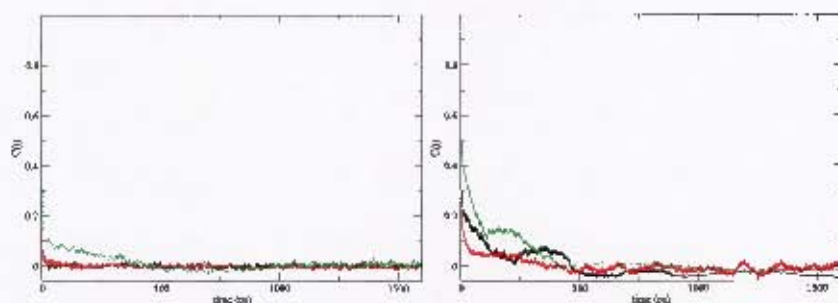


Figure 6.2 Time correlation functions for the hydroxymethyl and N-Acetyl torsions in ITZ55 from water (left) and DMSO (right) simulations. The disaccharide mimetic is in the +g conformation. The colours representation is as follows, $\theta_1$ (black), $\theta_2$ (red) and $\sigma$ (green).

114

### 6.5.2 Overall Molecular Correlations

The use of a $P_2$ autocorrelation function for a vector rigidly connected to molecule and an RMS fit autocorrelation gives an indication of the overall contributions made by molecular tumbling and internal fluctuations of a molecule.
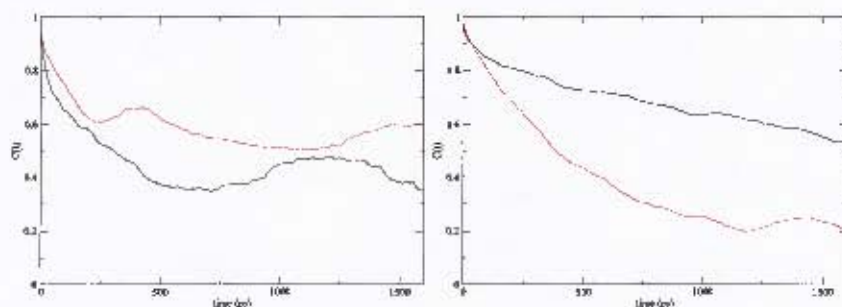


Figure 6.3 $P_2$ autocorrelation functions describing molecular tumbling in ITZ55 from water (left) and DMSO (right) simulations. For both the +g (black) and –g (red) conformations.

A vector was chosen between the C1 atom of ring 1 and the C1 atom of ring 2 to calculate the $P_2$ autocorrelation function used to describe molecular tumbling. Of all the water and DMSO simulations, only the ITZ55 and ITZ56 disaccharide in the –g conformers were found to have decayed to zero, within the confines of the simulation lengths. This decay took approximately 1ns for ITZ55 and 0.7ns for ITZ56 respectively. Using a least squares fit of the exponential decay to the first 100ps of all the functions it is possible to extract tumbling times, $\tau_M$ . These were calculated in the range from 233ps to 552 ps. RMS fit autocorrelations were also obtained for each of the solvent simulations and all internal correlation is found to have been lost within the first 70ps.

# 6.6 Relaxation Parameters

### 6.6.1 $T_1$ Relaxation Times from Solution Molecular Dynamic Simulations

The calculation of the Longitudinal $T_1$ Relaxation times (chapter 4.3) directly from solution simulations may only be carried out if a rotational tumbling time, $\tau_M$ is available. $T_1$ relaxation times were estimated from simulations using the $\tau_M$ values calculated in section 6.5.3. These were calculated at 400 MHz and the results for ITZ55 from both water and DMSO simulations are presented below in Tables 6.7 and 6.8.

| Ring | Carbon | $T_1(+g)$ 400MHz | $T_1(-g)$ 400 MHz |
|---|---|---|---|
| 1 | C1 | 0.95 | 0.73 |
|   | C2 | 0.98 | 0.74 |
|   | C3 | 1.03 | 0.87 |
|   | C4 | 1.02 | 0.83 |
|   | C5 | 0.89 | 0.80 |
|   | C6 | 0.43 | 0.41 |
| 2 | C1 | 0.46 | 0.39 |
|   | C2 | 0.46 | 0.40 |
|   | C3 | 0.34 | 0.42 |
|   | C4 | 0.41 | 0.40 |
|   | C5 | 0.44 | 0.35 |
|   | C6 | 0.20 | 0.17 |

Table 6.7 Longitudinal relaxation times ($T_1$), in seconds for ITZ55 from simulations started in at minima A (+g) and B (-g) in water, calculated at 400 MHz.

In general the $T_1$ values in water are seen to be higher than those calculated in DMSO for each of the three disaccharide mimetics. This may be due to errors incurred in the estimation of the tumbling times or as a result of the disaccharide mimetics appearing to tumble faster in the DMSO solvent, which is not necessarily the case, but may be an artifact due to the limited accuracy introduced, by the small number of rotations observed in the shorter DMSO simulations.

116

In all the water simulations the $T_1$ values for the carbons on each ring appear to be of similar values which reflect the rigidity of the rings and show the motions of each of the carbons to be the same. Interestingly, the values on ring 1 are generally higher than those observed for ring 2 in all the water simulations. This can however not be explained, as no definite answer is forthcoming.

| Ring | Carbon | $T_1$(+g) 400MHz | $T_1$(-g) 400 MHz |
|---|---|---|---|
| 1 | C1 | 0.29 | 0.22 |
| | C2 | 0.30 | 0.23 |
| | C3 | 0.30 | 0.23 |
| | C4 | 0.29 | 0.23 |
| | C5 | 0.29 | 0.23 |
| | C6 | 0.13 | 0.17 |
| 2 | C1 | 0.24 | 0.23 |
| | C2 | 0.25 | 0.23 |
| | C3 | 0.26 | 0.23 |
| | C4 | 0.27 | 0.23 |
| | C5 | 0.25 | 0.23 |
| | C6 | 0.13 | 0.15 |

Table 6.8 Longitudinal relaxation times ($T_1$), in seconds for ITZ55 from simulations started at minima A (+g) and B (-g) in DMSO, calculated at 400 MHz.

In the DMSO simulations the $T_1$ values for the carbons on each ring follow the same pattern observed for the water simulations, but the values observed for ring 1 and ring 2 are also of the same magnitude, thus indicating that the motion for individual rings is of the same magnitude. The general trend observed for both the water and DMSO $T_1$ values is that the C6 atoms on all the rings have the lowest values, which is attributed to fact that these carbons have two hydrogen atoms connected to them.

## 6.6.2 Generalised Order Parameters from Molecular Simulations

The information obtained from order parameters pertains only to the C-H vector and its restriction of motion. This makes it a popular tool in comparing simulated and experimental simulated relaxation data as it does not require the calculation of molecular tumbling or the rate internal motion. The order parameters were extracted from MD simulations by direct calculation as shown in chapter 4.4. Presented in Tables 6.9 and 6.10 are the order parameters calculated for ITZ55 simulations in water and DMSO.

The order parameters obtained for both the water and DMSO simulations indicate that generally there is more restriction of motion observed in the DMSO simulations. This observation can be made as the DMSO values are observed to be closer to 1.

| Ring | Carbon | $S^2(+g)$ | $S^2(-g)$ |
|------|--------|-----------|-----------|
| 1 | C1 | 0.35 | 0.26 |
|   | C2 | 0.39 | 0.29 |
|   | C3 | 0.37 | 0.22 |
|   | C4 | 0.40 | 0.23 |
|   | C5 | 0.55 | 0.21 |
|   | C6 | 0.46 | 0.33 |
| 2 | C1 | 0.30 | 0.57 |
|   | C2 | 0.41 | 0.51 |
|   | C3 | 0.49 | 0.61 |
|   | C4 | 0.47 | 0.40 |
|   | C5 | 0.75 | 0.61 |
|   | C6 | 0.49 | 0.21 |

Table 6.9 General Order parameters for ITZ55 simulations starting at minima A (+g) and B (-g) in water.

| Ring | Carbon | $S^2(+g)$ | $S^2(-g)$ |
|---|---|---|---|
| 1 | C1 | 0.57 | 0.78 |
| | C2 | 0.55 | 0.74 |
| | C3 | 0.56 | 0.72 |
| | C4 | 0.52 | 0.72 |
| | C5 | 0.36 | 0.74 |
| | C6 | 0.41 | 0.70 |
| 2 | C1 | 0.33 | 0.90 |
| | C2 | 0.33 | 0.88 |
| | C3 | 0.34 | 0.86 |
| | C4 | 0.33 | 0.87 |
| | C5 | 0.45 | 0.86 |
| | C6 | 0.18 | 0.40 |

Table 6.10 General Order parameters for ITZ55 simulations starting at minima A (+g) and B (-g) in DMSO

# References

1. Bothner-By, A.A.; Stephens, R.L.; Lee, J.; Warren, C.; Jeanloz, R.W. *J. Am. Chem. Soc.*, **1984**, *106*, 811

2. Bax, A.; Davis, D.G. *J. Magn. Reson.*, **1985**, *63*, 207.

3. Bax, A. *J. Magn. Reson.*, **1988**, *77*, 134.

4. Farmer II, B.T.; Macura, S.; Brown, L.R. *J. Magn. Reson*, **1987**, *72*, 347.

5. Szilágyi, L.; Pusztahelyi, Z., *Magn. Reson. Chem.*, **1992**, *30*, 107.

6. Neuhaus, D.; Keeler, J. *J. Magn. Reson.*, **1986**, *68*, 568.

7. Williamson, M. P.; Neuhaus, D. *J. Magn. Reson.*, **1987**, *72*, 369.

8. Mo, F.; Jensen, L.H. *Acta Crystallogr. Sect.B*, **1978**, *34*, 1562.

9. Szilágyi, L.; Forgó, P. *Carbohydr. Res.*, **1993**, *247*, 129.

10. Williamson, R. T.; Marques, B. L.; Gerwick, W. H.; Kövér, K. *Magn. Reson. Chem.*, **2000**, *38*, 265.

11. MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R.L.; Evanseck, J.D.; Field, M.J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.T.; Prodhom, B.; Reiher, III, W.E.; Roux, B.; Schlenkrich, M.; Smith, J.C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B*, **1998**, *102*, 3586.

12. Krivdin, L. B.; Della, E. W. *Progr. NMR Spectrosc.*, **1991**, *23*, 364.

13. Serianni, A. S. *J. Am. Chem. Soc.*, **2004**, *126*, 15668.

14. Xu, Q.; Bush, A. *Carbohydr. Res.*, **1998**, *306*, 335.

# Chapter 7

# Conclusions and Future Work

The primary objective of this work was to develop a sulphur-sulphur interglycosidic linkage to be used in molecular dynamics simulations for three disaccharide mimetics, in order to obtain comparative NOE and spin-spin coupling data as obtained from NMR experiments. This was achieved with limited success. The parameters of the respective torsion angles in the glycosidic linkage, especially the C-S torsion parameters should be re-examined as they appear to be too rigid and do not allow the sampling of the appropriate conformational space, as found for ITZ55. Following reparameterisation, the parameters can be expanded to include other novel glycosidic linkages, such as the three bonded sulphur-nitrogen linkage.

The secondary objective was to find a reason for the preference of the sulphur-sulphur torsion to be in the +g and –g conformations. This objective was met by showing electron delocalisation between the lone pair of the oxygen and the anti-bonding orbitals contained in the carbon-sulphur bond using natural bond orbital analysis. This study can be expanded to a further study of the effects surrounding this preference by increasing the size of the substituents connected to the respective sulphurs and may also be expanded to include other novel glycosidic linkages.

The results for the DMSO simulations seem to suggest that there is excessive hydrogen bonding to the solvent which contradicts experimental disaccharide studies in DMSO. This in turn impedes the internal motion of the disaccharide mimetics studied and may explain their conformational preferences. Further study of the DMSO solvent model should be conducted with disaccharides to obtain closer correlation with experimental data.

In conclusion, this work has shown that further parametrisation of a carbohydrate force field for use in DMSO solvent simulations is required, as is further work required on the sulphur-sulphur glycosidic linkage.

# Appendix A

# Vacuum Clusters and Solution Dynamics Plots

## A1.  Vacuum Cluster Analysis

| sim. | cluster | # members | Std deviation | $\Phi$ | $\Psi$ | $\Omega$ | $\theta$ | $\sigma$ |
|---|---|---|---|---|---|---|---|---|
| a | 1 | 31599 | 1.29E+01 | 33.79 | -85.92 | 171.86 | -66.19 | 103.01 |
|  | 2 | 19679 | 1.26E+01 | 29.88 | -87.57 | -173.90 | 69.68 | 103.02 |
|  | 3 | 29461 | 1.17E+01 | 33.11 | -86.26 | 170.69 | -63.00 | 105.17 |
|  | 4 | 19162 | 9.82E+00 | 28.94 | -87.48 | -173.85 | -64.09 | 105.12 |
| b | 1 | 43815 | 1.99E+01 | 41.81 | 77.36 | -44.67 | 44.91 | 101.47 |
|  | 2 | 14509 | 2.50E+01 | 106.59 | 103.01 | -3.02 | -84.20 | 102.01 |
|  | 3 | 13993 | 1.76E+01 | 32.65 | 73.60 | -51.71 | 52.32 | 105.06 |
| c | 1 | 46388 | 1.58E+01 | 38.87 | 75.42 | -45.92 | 63.47 | 100.06 |
|  | 2 | 14349 | 2.07E+01 | -37.05 | 80.77 | 29.86 | 91.53 | 109.60 |
|  | 3 | 19123 | 1.63E+01 | 35.04 | 73.79 | -49.83 | 51.64 | 104.99 |
| d | 1 | 51552 | 1.45E+01 | -41.98 | -90.07 | 16.86 | 95.28 | 102.20 |
|  | 2 | 48349 | 1.19E+01 | -41.68 | -90.89 | 14.00 | 79.88 | 104.83 |
| e | 1 | 42880 | 1.53E+01 | 38.23 | 74.80 | -47.50 | 64.24 | 100.21 |
|  | 2 | 22841 | 1.78E+01 | 32.49 | 74.36 | -47.76 | 52.95 | 106.44 |
|  | 3 | 10649 | 2.80E+01 | -39.92 | 81.05 | 32.70 | 92.65 | 103.47 |

Table A.1 Clusters from ITZ55 vacuum Simulations from minima A to E

| sim. | Cluster | # members | Std deviation | $\Phi$ | $\psi$ | $\Omega$ | $\theta_1$ | $\theta_2$ |
|---|---|---|---|---|---|---|---|---|
| a | 1 | 6393 | 2.93E+01 | 23.61 | 81.86 | -43.71 | 58.31 | 50.52 |
|   | 2 | 7433 | 2.85E+01 | -24.31 | 97.90 | 165.99 | -60.64 | -62.55 |
|   | 3 | 9292 | 2.09E+01 | 23.48 | 82.39 | 167.81 | 103.95 | -63.53 |
|   | 4 | 6109 | 2.13E+01 | -48.70 | 104.98 | 163.87 | -66.63 | 53.85 |
| b | 1 | 11708 | 1.84E+01 | 22.85 | 80.52 | -41.44 | 68.15 | 48.78 |
|   | 2 | 21848 | 1.43E+01 | 51.40 | 82.54 | -62.94 | 45.84 | 49.06 |
|   | 3 | 26160 | 1.70E+01 | 43.51 | 81.43 | -59.13 | 40.27 | -58.65 |
| c | 1 | 10419 | 1.28E+01 | 2.20 | -79.46 | -172.39 | 139.40 | -75.64 |
|   | 2 | 11986 | 1.75E+01 | -70.85 | -106.64 | 169.65 | -23.12 | -58.13 |
|   | 3 | 11300 | 1.25E+01 | -58.15 | -104.34 | -168.70 | -35.65 | -58.04 |
|   | 4 | 7013 | 1.15E+01 | -61.90 | -106.60 | -168.61 | -28.46 | 48.28 |
|   | 5 | 7439 | 2.04E+01 | -67.01 | -107.21 | 170.14 | -20.99 | 53.62 |
|   | 6 | 7137 | 1.44E+01 | 35.48 | -81.30 | -170.50 | -5.19 | -57.38 |
| d | 1 | 12117 | 3.19E+01 | -22.52 | -101.59 | 166.21 | -64.27 | -22.08 |
|   | 2 | 12002 | 3.05E+01 | -159.98 | -85.89 | -157.29 | -60.81 | -5.87 |
|   | 3 | 13617 | 2.79E+01 | 21.96 | -84.50 | 165.82 | 103.60 | -36.22 |
|   | 4 | 14225 | 2.54E+01 | -169.00 | -93.53 | -15.08 | -22.05 | -58.34 |
| e | 1 | 7996 | 2.48E+01 | -27.33 | 80.70 | 32.04 | 85.54 | 52.47 |
|   | 2 | 33423 | 2.16E+01 | 40.23 | 80.87 | -57.85 | 53.75 | 57.17 |
|   | 3 | 36836 | 1.92E+01 | 34.55 | 81.12 | -59.55 | 41.78 | -57.41 |

Table A.2 Clusters from ITZ56 vacuum simulations from minima A to E

| sim. | cluster | # members | std deviation | $\Phi$ | $\Psi$ | $\Omega$ | $\theta_1$ | $\theta_2$ |
|---|---|---|---|---|---|---|---|---|
| a | 1 | 11183 | 1.31E+01 | 58.121 | 109.975 | -167.31 | -59.588 | 61.603 |
|   | 2 | 11954 | 1.92E+01 | 11.774 | 100.909 | -170.835 | -61.642 | -57.194 |
|   | 3 | 11644 | 1.27E+01 | 41.923 | 103.967 | 169.767 | -62.037 | -57.129 |
| b | 1 | 41752 | 2.26E+01 | 55.382 | -85.683 | -155.708 | -64.801 | 130.417 |
|   | 2 | 17188 | 1.05E+01 | 54.895 | -85.338 | -156.432 | -64.648 | -166.156 |
|   | 3 | 7555 | 3.52E+01 | 27.467 | -103.476 | -167.107 | -29.313 | 96.777 |
| c | 1 | 33551 | 2.35E+01 | -155.142 | -86.26 | -155.417 | -63.962 | 123.491 |
|   | 2 | 14092 | 1.82E+01 | -154.97 | -85.734 | -155.276 | -54.007 | -165.461 |
| d | 1 | 9358 | 2.16E+01 | -36.434 | 83.705 | 39.45 | -56.826 | 62.78 |
|   | 2 | 14571 | 1.59E+01 | -63.201 | 109.06 | 49.289 | -57.904 | 69.095 |
|   | 3 | 8325 | 2.31E+01 | 47.491 | 87.522 | -13.284 | -59.585 | 63.676 |
| e | 1 | 14106 | 2.86E+01 | -41.899 | 83.201 | 35.329 | -59.349 | 90.436 |
|   | 2 | 10145 | 2.95E+01 | 27.329 | 81.411 | -38.027 | 45.53 | 98.42 |
|   | 3 | 11441 | 2.61E+01 | 43.781 | 88.742 | -4.026 | -61.931 | 67.572 |
|   |   | 10698 | 2.94E+01 | -38.658 | 83.374 | 36.89 | 53.565 | 93.321 |

Table A.3 Clusters from ITZ57 vacuum simulations from minima A to E

## A2.    Time Series Plots



$\Phi$

$\psi$

$\Omega$

$\theta_1$

$\sigma$
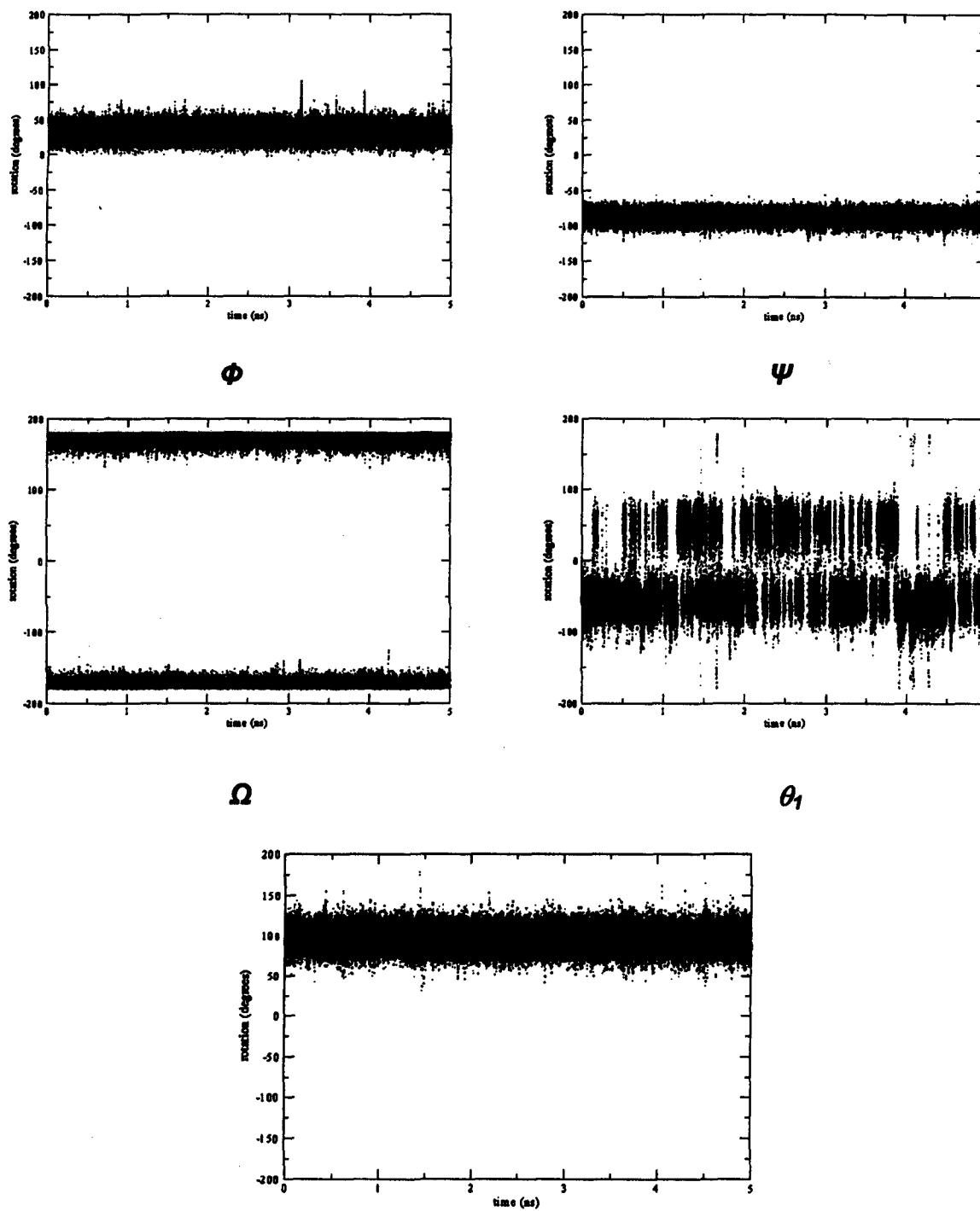
Figure A.1 Vacuum time series plots of dihedral angles from ITZ55 starting from minima
D with the sulphur-sulphur torsion n the –g conformation.

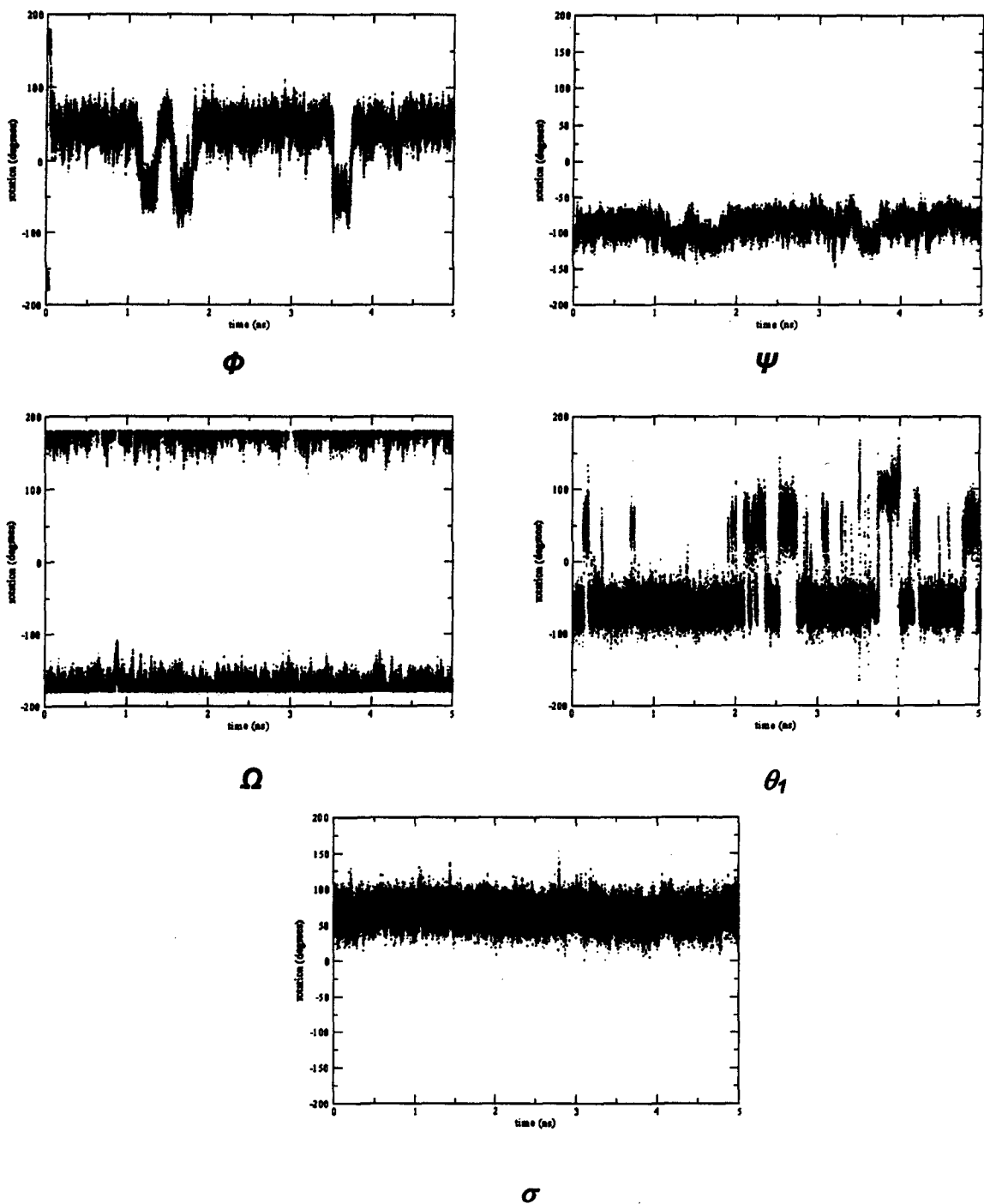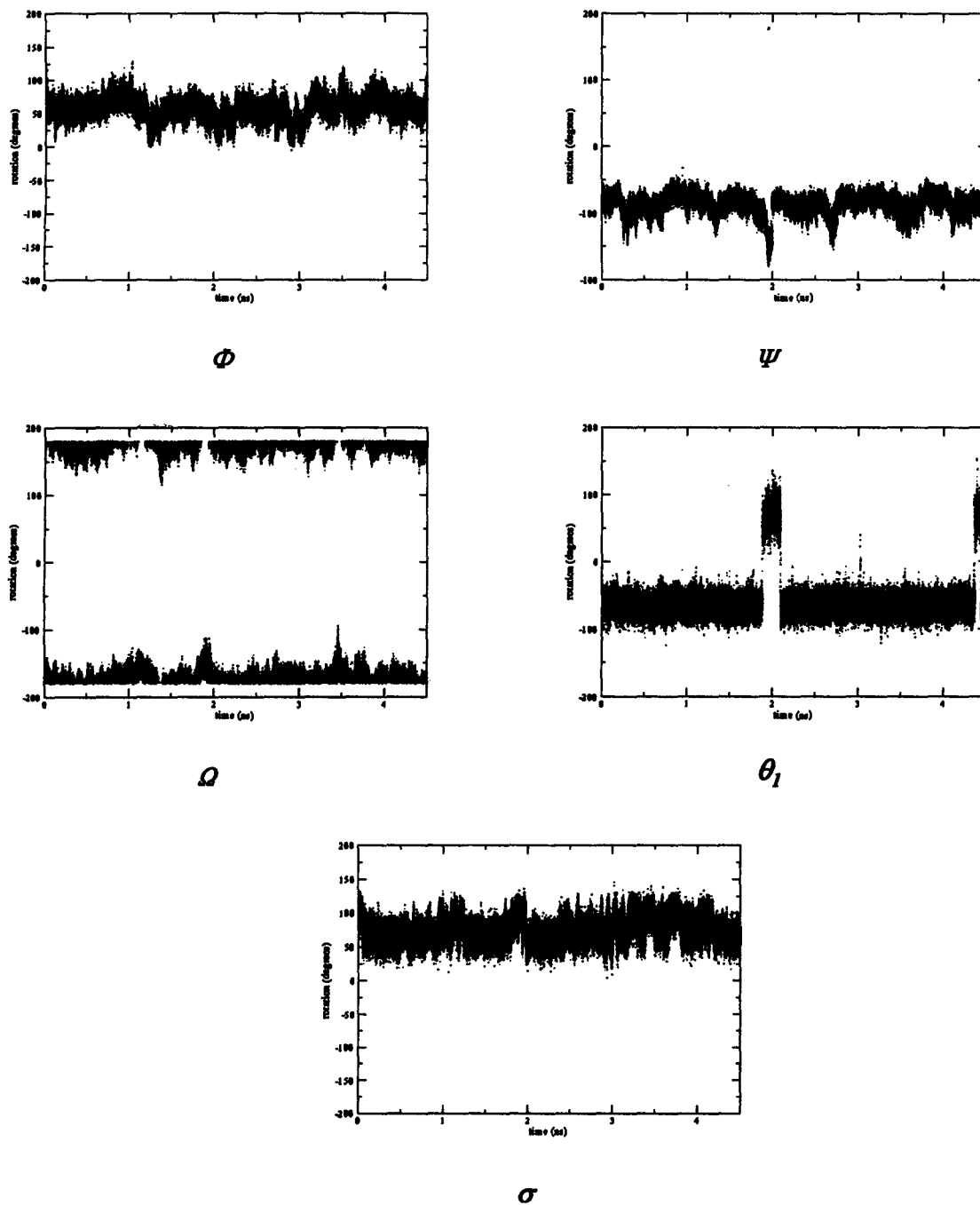Figure A.2 Time series plots of dihedral angles from ITZ55 starting from minima D (-g conformation) in water

*Φ*

*ψ*

*Ω*

*θ₁*

*σ*

Figure A.3 Time series plots of dihedral angles for ITZ55 starting from minima D (-g conformation) in DMSO

127

# Appendix B

# Relaxation Parameters

## B1.  NOE, $^1$H-$^1$H and $^1$H-$^{13}$C coupling values

| ITZ57 | +g | -g |
|---|---|---|
| Gl2-OH / Gl1 | 3.42176 | 3.36241 |
| Gl2-OH / Gl2 | 2.29457 | 2.34693 |
| Gl2-OH / Gl3 | 2.99365 | 2.89482 |
| Ga2-OH / Ga1 | 3.40132 | 3.3783 |
| Ga2-OH / Ga2 | 2.25895 | 2.2499 |
| Ga2-OH / Ga3 | 3.13008 | 3.20246 |
| Gl3-OH / Gl2 | 3.02532 | 3.12705 |
| Gl3-OH / Gl3 | 2.31313 | 2.34858 |
| Gl4-OH / Gl4 | 2.30113 | 2.30591 |
| Ga3-OH / Ga4 | 3.35013 | 3.26591 |
| Ga3-OH / Ga2 | 2.91813 | 3.0109 |
| Ga3-OH / Ga3 | 2.39823 | 2.36799 |
| Ga6-OH / Ga4 | 4.15913 | 4.07227 |
| Ga6-OH / Ga6a,b | 2.38164 | 2.39511 |
| Gl6-OH / Gl6a | 2.32954 | 2.32821 |
| Gl6-OH / Gl6b | 2.43346 | 2.54713 |
| Gl6-OH / Gl5 | 3.59458 | 3.85394 |
| Gl6-OH / Gl4 | 3.91133 | 3.41498 |
| Ga4-OH / Ga4 | 2.99833 | 2.99646 |
| Ga4-OH / Ga2 | 7.33344 | 8.43281 |
| Ga4 / Ga3 (ref.) | 2.38805 | 2.39715 |
| Ga4 / Gl4 | 8.88699 | 7.76366 |
| Gl6a / Gl6b | 1.75817 | 1.75832 |

Table B.1 Calculated Interannular NOE values from ITZ57 simulations. Distances are all in Å

## B2.    Correlation Plots



Figure B.1 Glycosidic torsion autocorrelation functions for ITZ56 in water (top left) and DMSO (top right) for the +g conformer, as well as ITZ56 in water (bottom left) and DMSO (bottom right) for the −g conformer. Representing $\Phi$ (black), $\Psi$ (red) and $\Omega$ (green).
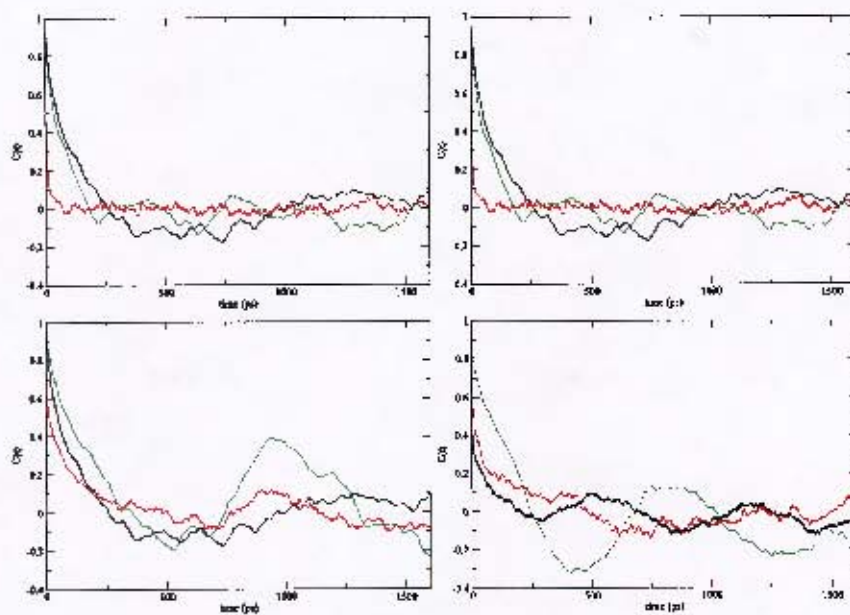
Figure B.2 Glycosidic torsion autocorrelation functions for ITZ57 in water (top left) and DMSO (top right) for the +g conformer, as well as ITZ57 in water (bottom left) and DMSO (bottom right) for the −g conformer. Representing $\Phi$ (black), $\Psi$ (red) and $\Omega$ (green).
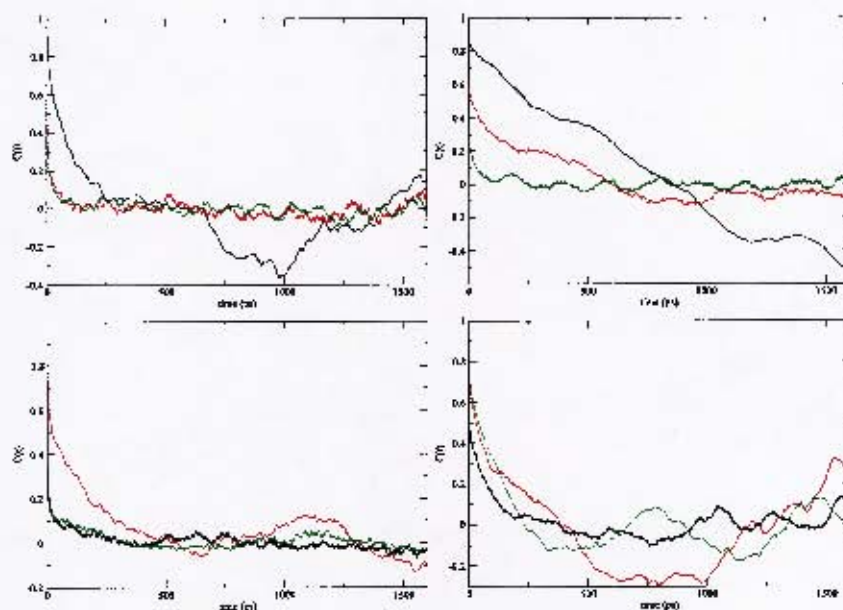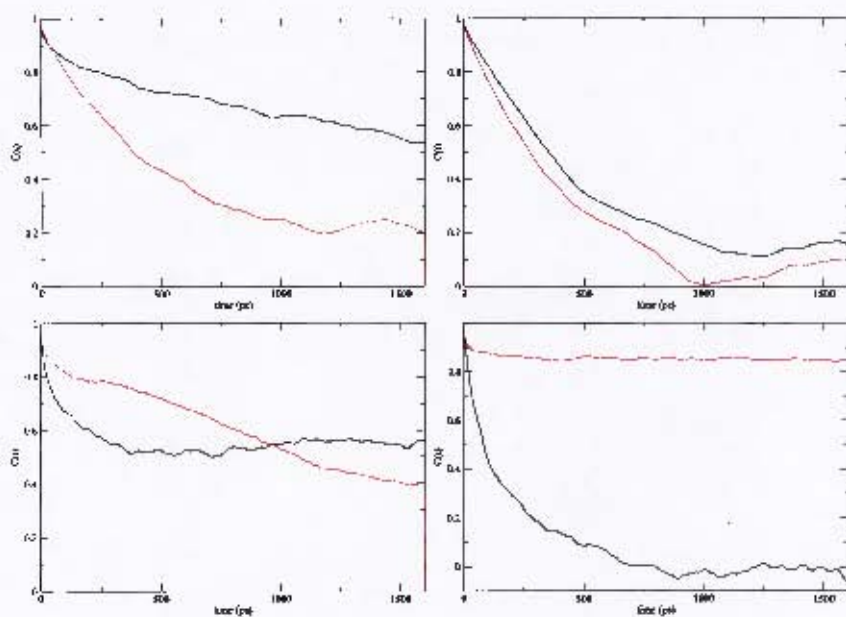
Figure B.3 $P_2$ correlation functions for ITZ56 in water (top left) and DMSO (top right) and ITZ57 in water (bottom left) and DMSO (bottom right). The black line indicates the +g conformer and the red line indicates the −g conformer

## B3. $T_1$ and $S^2$ Data

| Ring | Carbon | $T_1(+g)$ 400MHz | $T_1(-g)$ 400 MHz |
|------|--------|------------------|-------------------|
| 1    | C1     | 0.56             | 0.68              |
|      | C2     | 0.87             | 0.71              |
|      | C3     | 0.79             | 0.76              |
|      | C4     | 0.66             | 0.71              |
|      | C5     | 0.91             | 0.61              |
|      | C6     | 0.42             | 0.45              |
| 2    | C1     | 0.16             | 0.39              |
|      | C2     | 0.80             | 0.44              |
|      | C3     | 0.54             | 0.51              |
|      | C4     | 0.53             | 0.50              |
|      | C5     | 0.14             | 0.47              |
|      | C6     | 0.38             | 0.32              |

Table B.5 Longitudinal relaxation times ($T_1$), in seconds for ITZ56 from simulations started with the sulphur-sulphur torsion in the +g and -g conformations in water, calculated at 400 MHz.

| Ring | Carbon | $T_1(+g)$ 400MHz | $T_1(-g)$ 400 MHz |
|------|--------|------------------|-------------------|
| 1 | C1 | 0.26 | 0.23 |
|   | C2 | 0.26 | 0.23 |
|   | C3 | 0.26 | 0.23 |
|   | C4 | 0.27 | 0.24 |
|   | C5 | 0.26 | 0.23 |
|   | C6 | 0.17 | 0.17 |
| 2 | C1 | 0.24 | 0.21 |
|   | C2 | 0.25 | 0.23 |
|   | C3 | 0.24 | 0.21 |
|   | C4 | 0.24 | 0.24 |
|   | C5 | 0.24 | 0.21 |
|   | C6 | 0.30 | 0.34 |

Table B.6 Longitudinal relaxation times $(T_1)$, in seconds for ITZ56 from simulations started with the sulphur-sulphur torsion in the +g and –g conformations in DMSO, calculated at 400 MHz.

| Ring | Carbon | $T_1(+g)$ 400MHz | $T_1(-g)$ 400 MHz |
|---|---|---|---|
| 1 | C1 | 0.76 | 0.72 |
| | C2 | 0.78 | 0.77 |
| | C3 | 0.82 | 0.76 |
| | C4 | 0.80 | 0.71 |
| | C5 | 0.86 | 0.61 |
| | C6 | 0.38 | 0.31 |
| 2 | C1 | 0.60 | 0.77 |
| | C2 | 0.62 | 0.77 |
| | C3 | 0.71 | 0.46 |
| | C4 | 0.38 | 0.40 |
| | C5 | 0.74 | 0.61 |
| | C6 | 0.47 | 0.47 |

Table B.7 Longitudinal relaxation times $(T_1)$, in seconds for ITZ57 from simulations started with the sulphur-sulphur torsion in the +g and −g conformations in water, calculated at 400 MHz.

134

| Ring | Carbon | $T_1(+g)$ 400MHz | $T_1(-g)$ 400 MHz |
|------|--------|------------------|-------------------|
| 1 | C1 | 0.22 | 0.24 |
|   | C2 | 0.26 | 0.24 |
|   | C3 | 0.26 | 0.25 |
|   | C4 | 0.24 | 0.25 |
|   | C5 | 0.23 | 0.24 |
|   | C6 | 0.26 | 0.15 |
| 2 | C1 | 0.23 | 0.22 |
|   | C2 | 0.22 | 0.23 |
|   | C3 | 0.28 | 0.23 |
|   | C4 | 0.22 | 0.26 |
|   | C5 | 0.27 | 0.22 |
|   | C6 | 0.23 | 0.21 |

Table B.8 Longitudinal relaxation times ($T_1$), in seconds for ITZ57 from simulations started with the sulphur-sulphur torsion in the +g and −g conformations in DMSO, calculated at 400 MHz.