COMBINED NMR AND SIMULATION STUDY OF CARBOHYDRATE LINKAGE DYNAMICS

A thesis submitted to the
UNIVERSITY OF CAPE TOWN
in fulfillment of the requirements for the degree of

MASTER of SCIENCE

by
Robert Best
B.Sc(Hons)(University of Cape Town)

April 2000

Abstract

Biomaterials have recently gained significance as a result of environmental pressures and their increased viability through crop engineering. Starch harvested from maize crops is one example of a cheap and abundant biopolymer, which could substitute conventional polymers such as polyethylene as a material in manufacturing. Its practical application, however, will depend on understanding its physical behaviour, so that intelligent modifications can be made to enhance its properties.

The most effective way of changing a polymer's properties is by modification at the chemical level, since ultimately it is this which determines the macroscopic features, such as viscosity, plasticity and tensile strength. Since the origin of these features is difficult to study experimentally, this thesis will tackle the problem by means of computer simulation of small carbohydrate fragments. In particular, the two principal types of carbohydrate linkage (namely $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$) will be examined in this thesis, since these are the chief elements of conformational flexibility in these molecules. The $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkages are studied both separately in the maltose and isomaltose molecules respectively and together in panose and the resultant water structuring and dynamic behaviour are studied.

Although the computational approach provides insights not available experimentally, it is nonetheless important to compare the results obtained with an experimental reference as a check on their validity. NMR T_1 relaxation measurements, which give indirect information on the atomic scale dynamics, have been measured for each of the model fragments and compared with values calculated directly from simulation.

Acknowledgments

I would like to thank:

- My supervisors:
  - Dr Kevin Naidoo, for his suggestions and assistance in the Computational aspects of the project.
  - Prof. Graham Jackson for his advice with the NMR section of the work.
- The Foundation for Research Development for my MSc bursary.
- The USDA-ARS for their support in providing the computers on which the simulations were performed.
- Peter Sandor of Varian, Inc. for running the 1-D TOCSY experiments.
- My fellow group members Michelle Kuttel, Chas Simpson, Günter Klatt and Jeff Chen for their help and support.

List of Abbreviations

Å	angstrom
CHARMM	Chemistry at Harvard Macromolecular Mechanics
fs	femtosecond
ps	picosecond
MD	Molecular Dynamics
MC	Monte Carlo
NVE	Microcanonical Ensemble
NVT	Canonical Ensemble
T_1	Longitudinal Spin Relaxation Time
T_2	Transverse Spin Relaxation Time
NMR	Nuclear Magnetic Resonance
CD	Circular Dichroism

Contents

1 Introduction **1**

1.1 General Importance of Carbohydrates 1

1.1.1 Biological Role of Carbohydrates 1

1.1.2 Applications of Carbohydrates as Materials 3

1.2 Macromolecular Conformation and Property Relations 4

1.3 Methods for Investigating Conformation 5

1.3.1 X-Ray Diffraction Experiments 5

1.3.2 NMR Experiments 5

1.3.3 Force Fields and Molecular Dynamics Simulations 6

1.4 Objectives 7

2 Computer Simulation Methods **9**

2.1 Molecular Mechanics Formalism 9

2.1.1 Background 9

2.1.2 Force Fields 11

2.2 Molecular Dynamics 14

2.2.1 Integration of Equations of Motion 15

2.2.2 Potential Truncation and Switching 15

2.2.3 Periodic Boundary Conditions 17

2.2.4 Simulation Ensemble 20

2.2.5 Limitations of the Molecular Dynamics Method 20

2.3 Analysis Methods for Condensed Phases 21

2.3.1 Structural Analysis 21

2.3.2 Correlation Functions and Transport Properties 25

2.3.3 Conformational Analysis 27

2.4 Models for Water 29

2.5 Models for Carbohydrates 30

3 NMR Relaxation Theory and Experiments **33**

3.1 Measurement of NMR Relaxation Parameters 33

3.1.1 Measurement of T_1 . . . 35
3.2 Physical Basis for NMR Relaxation . . . 35
3.3 Calculation of Dipolar Relaxation Parameters from Simulation . . . 37
3.4 Model Free Formalism . . . 37
3.5 Calculation of Generalised Order Parameters from Simulation . . . 39

4 Maltose (α(1→4)-linkage) 41
4.1 Simulation Details . . . 42
4.1.1 Adiabatic Map and Simulation Conditions . . . 42
4.2 Basic Trajectory Analysis . . . 43
4.3 Cluster Analysis . . . 45
4.4 Hydrogen Bond Analysis . . . 49
4.5 Water structuring about Maltose . . . 50
4.6 Internal relaxation behaviour . . . 55
4.7 Calculation of NMR Properties and Experimental Comparison . . . 59

5 Isomaltose (α(1→6)-linkage) 63
5.1 Relaxed Adiabatic Map of Isomaltose . . . 65
5.1.1 Methodology . . . 65
5.1.2 Features of Map . . . 67
5.2 Molecular Dynamics Simulations of Isomaltose . . . 74
5.2.1 Simulation Methodology . . . 74
5.2.2 Pucker Analysis . . . 75
5.2.3 Cluster Analysis . . . 75
5.2.4 Time dependence of molecular motion . . . 79
5.2.5 Hydrogen Bonding Analysis . . . 87
5.2.6 Solvent Structuring about Isomaltose . . . 93
5.3 Molecular Relaxation . . . 101
5.3.1 Dihedral Angle Time Correlation Functions . . . 101
5.3.2 Overall Molecular Correlation Functions . . . 103
5.4 Isomaltose T_1 Relaxation Experiments . . . 105
5.5 Comparison of Relaxation Parameters from Experiment and Simulation . 108

6 Modelling an Amylopectin branch point 111
6.1 Trisaccharide Model: Panose . . . 111
6.1.1 Molecular Dynamics Simulations . . . 113
6.1.2 Simulation Events . . . 113
6.1.3 Cluster Analysis . . . 114

6.1.4 Hydrogen Bond Analysis 116
6.1.5 Solvent Structure Analysis 119
6.1.6 NMR Relaxation parameters from Simulation and Experiment . . 123
6.2 Tetrasaccharide Model 130
6.2.1 Molecular Dynamics Simulations 131
6.2.2 Dynamic Events 131
6.2.3 Cluster Analysis 132
6.2.4 Hydrogen Bond Analysis 135
6.2.5 Solvent Structure Analysis 135
6.3 Branch models in the context of larger polysaccharides 140

A Isomaltose dynamics plots **149**
A.1 Time Series plots 149
A.2 Correlation plots 151

B Panose Dynamics Plots **167**

C Detailed hydrogen bond analyses for branch models **171**

D NMR Spectral Plots for Panose **173**

Chapter 1

Introduction

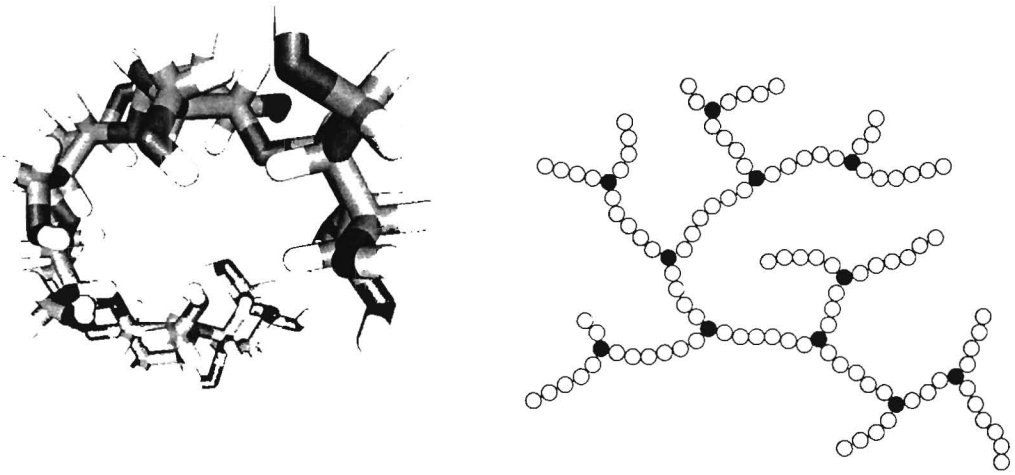
1.1 General Importance of Carbohydrates

1.1.1 Biological Role of Carbohydrates

Structural and Storage Functions

The best known examples of the biological function of polysaccharides are the structural role played by cellulose (a $\beta(1\rightarrow4)$-linked polyglucose, or *glucan*, which forms flat, sheet-like structures) in plant cell walls and the function of starch as a plant storage polysaccharide. Starch is actually a mixture of *amylose*, an $\alpha(1\rightarrow4)$-linked glucan, and amylopectin, a branched polysaccharide consisting of $\alpha(1\rightarrow4)$-linked chains joined together by $\alpha(1\rightarrow6)$-linked branch points (figure 1.1). The average distance between branches in amylopectin is typically 24 to 30 glucose residues. Amylose prefers to form helices, of which three forms have been identified. X-ray diffraction[1,2] and NMR studies[3] have shown that the naturally occurring A and B forms consist of a right-handed double helix, while the C form, formed in the prescence of iodine, adopts a left-handed single helix. About six residues per turn are found for the A and B polymorphs.

Starch occurs in granules of up to $100\mu m$ in diameter, formed from an outward growth process from the central point (or hilium). In the conventional model, the starch granule, shown in figure 1.2, consists of concentric growth layers, alternating between crystalline and amorphous lamellae. It is generally accepted[4–6] that the crystalline regions consist of amylopectin double helices (consisting of the amylose-like $\alpha(1\rightarrow4)$-linked regions) and that the amorphous regions contain essentially the branches, which are $\alpha(1\rightarrow6)$-linked. In the crystalline region, the helices are oriented parallel to one another and perpendicular to the granule surface.



(a) (b)

Figure 1.1: Schematic illustrations of the two components of starch: (a) amylose, (b) amylopectin. Note that the interbranch distances in amylopectin have been shortened for clarity.

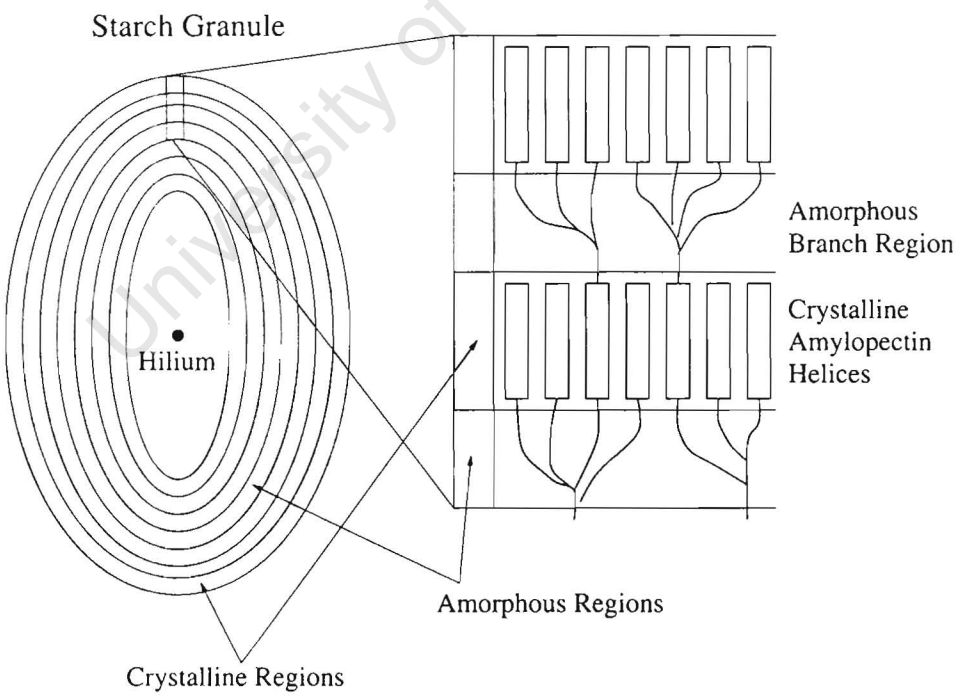


Figure 1.2: Model for the composition of a starch granule.

Other Biological Roles of Carbohydrates

Smaller carbohydrates play many important biological roles, particularly in molecular recognition and signalling processes, which are at least as important as the role of polysaccharides in food storage. The large family of mammalian carbohydrate binding proteins known as lectins are responsible for many types of cellular recognition and adhesion. For example, L-selectin on the surface of lymphocyte cells is able to facilitate their transport between the vascular and lymphatic systems by binding the Gly-CAM protein on the surface of certain endothelial cells.[7]

The targeting of proteins to their intended location in the cell is also sometimes mediated by oligosaccharides, for example in the targeting of lysosomal enzymes to the lysosome. In cell signalling pathways, small oligosaccharides are sometimes involved as messengers; inositol triphosphate produced by phospholipase C triggers the release of the second messenger Ca^{2+} from the endoplasmic reticulum. The attachment of the influenza virus to red blood cells depends on a carbohydrate motif on the surface of the cell. Vaccines containing carbohydrate moieties designed to mimic the carbohydrates on the cell surfaces of viruses have also recently been developed.

Other roles include glycoconjugate hormones, enzyme cofactors and as integral parts of proteins. It can thus be seen that oligosaccharides are involved in a multitude of important cellular functions.

1.1.2 Applications of Carbohydrates as Materials

Starch, being the main food storage molecule in plants, is available in almost unlimited quantities from commercial harvests of maize, wheat and potatoes. Its properties have recently been recognised as being favourable for use as a polymeric material to substitute the plastic products used today. Current environmental pressures have caused considerable interest in such applications, which could range from packing materials to eating utensils.

Experiments done in the 1970's revealed that extrusion of starch with low water content at high temperatures removes its ordered crystallinity and renders it suitable for use in molded materials. Experiments with cross-linking and co-melts with other polymers such as poly(ethylene-co-acrylic acid) and polyethylene[8,9] have yielded a material suitable for use as a packing material, or in injection molded plastics.

The practical application of starch as a material, however, is currently limited by its physical properties, the most undesirable being its increased plasticity in the prescence of water.[10,11] One possible method of altering the properties of starch would be through genetic changes in the maize crop. Another would be through chemical modification of

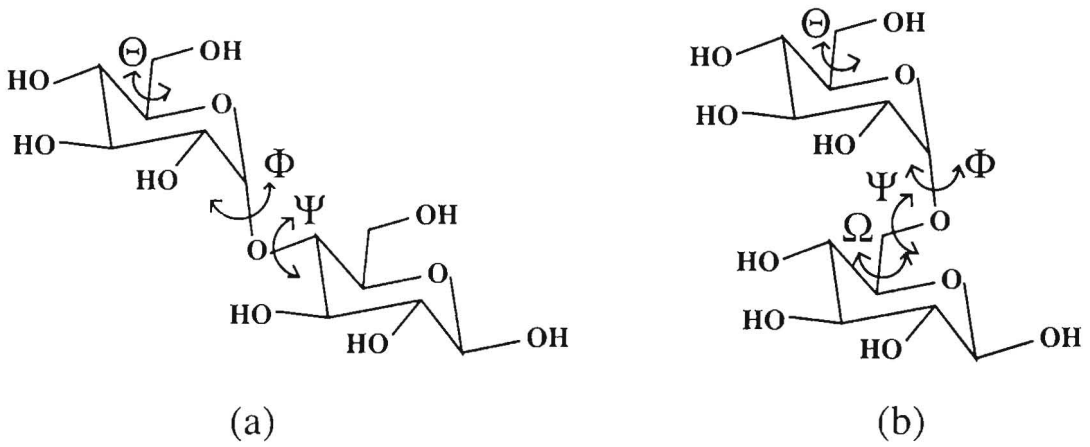


Figure 1.3: Schematic illustrations of disaccharides containing $\alpha(1\rightarrow4)$ (a) and $\alpha(1\rightarrow6)$ (b) linkages, showing the general nomenclature for the torsion angles.

the extracted starch product. However, in order to make an intelligent guess at what modifications might prove fruitful, it is necessary to have a proper understanding of the structure and dynamics of starch at a molecular level, and the influence of water on its properties.

1.2 Macromolecular Conformation and Property Relations

In order to understand the role and mechanism by which the various biologically relevant oligosaccharides operate, it is necessary to have a description of their structure, and possibly dynamics. For example, in considering the binding of a lectin to a glycoprotein, the shape and functionalization of the sugar moiety will determine how tightly it binds to a particular site on the lectin protein. Similarly, in understanding the properties of the carbohydrate storage polymers in starch, a description of the favoured conformations will be required. In this case, the dynamic behaviour of the polymer would also be important.

A first step to understanding the properties of a molecule is the identification of the main degrees of conformational freedom. The saccharides with which we will be chiefy concerned are the simple $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linked components of starch. Schematic diagrams of molecules containing $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkages are given in figure 1.3.

The rings in this type of carbohydrate are fairly rigid, for the most part remaining in a chair conformation in which carbon 4 is above the mean plane of the ring and carbon 1 is below, known as the $^4C_1$ chair. Therefore the flexibility in these molecules originates primarily in rotations about the dihedrals comprising the glycosidic linkage, which are named Φ and Ψ for the $\alpha(1\rightarrow4)$ linkage and Φ, Ψ and Ω for the $\alpha(1\rightarrow6)$ linkage as

shown in figure 1.3. An extra degree of conformational freedom is rotation about the C5-C6 bond of the hydroxymethyl group. The secondary alcohols are not a fundamental degree of freedom due to the low barrier to rotation between gauche rotamers and hence rapid rotation, and the fact that the orientation of these alcohols does not have much effect on the structure of the carbohydrate backbone.

Thus a first step in analyzing the conformation of a polysaccharide is the description of the molecular energy surface as a function of the glycosidic dihedrals, usually for a model disaccharide. Low energy wells present in this surface will give a general indication of the regions of conformational space likely to be sampled. The properties of the polysaccharide may then be reduced to the properties of only a few minima.

1.3 Methods for Investigating Conformation

As discussed above, a good understanding of carbohydrate conformation is important for understanding their behaviour as materials and in various biological roles. Several experimental and theoretical methods are available for studying both conformation and dynamics.

1.3.1 X-Ray Diffraction Experiments

High accuracy structural and conformational information may be extracted from single-crystal X-ray structures. However, due to difficulties with crystallizing larger saccharides, structures for only a few small oligosaccharides[12–14] have so far been available. Crystal structure data also provide only a single conformational data point and do not include dynamic or solvent effects.

Fibre diffraction and X-ray microfocus diffraction[4] give insight into the macroscale structure of polysaccharides. The latter method has been used to study the packing of amylopectin in starch granules. Fibre diffraction can provide general information such as unit cell sizes, which, in conjunction with a model or other crystal structure data, allows the packing of the polymer to be deduced. While this is useful on a large scale system, it does not provide any detailed conformational information.

1.3.2 NMR Experiments

High resolution NMR spectroscopy is currently the only method for the experimental determination of precise structural and dynamic information from molecules in solution. ^3J Coupling between 1,4 related nuclei yields information on the dihedral angle between them via the Karplus relationship (of which several functional forms have been

developed for carbohydrates[14,15]). This has principally been used in the determination of the rotameric distribution of hydroxymethyl groups,[16–18] since it is not applicable to most glycosidic dihedrals (except for Ω in the $\alpha(1\rightarrow6)$ linkage, which has the requisite attached protons).

The nuclear Overhauser effect (nOe) gives information on the distances between protons which are less than 5Å apart. The rate of nOe buildup, as measured by NOESY spectroscopy is proportional to the inverse sixth power of the distance between the protons under appropriate conditions (slow molecular tumbling relative to the rate of internal motion). Inter-ring nOe's can in principle be used to determine the environment around the glycosidic linkages, which is of greatest interest. Unfortunately, there are several practical difficulties: there are usually only two nOe's observed between the rings, which does not constrain the range of possible glycosidic conformations very much when the errors in the distances are considered.[7] Moreover, if the glycosidic dihedrals are undergoing transitions between several rotamers, then the observed nOe's will be averaged over the different conformations, leading to a false conformation if a static interpretation is adopted. Nonetheless, nOe's are very useful, especially in obtaining conformations for larger oligosaccharides, where nOe coupling between non-sequential residues can give information on folding.

Dynamical properties of polysaccharides may be extracted from ^{13}C relaxation data, such as longitudinal and transverse relaxation times, T_1 and T_2 respectively.[19,20] These relaxation parameters contain information on molecular tumbling as well as internal motions, which may be extracted by suitable analysis (for example the model-free formalism of Lipari and Szabo[21]). However, the motional information available from such interpretations is limited to a general description of the dynamics.

One of the main limitations of all the NMR methods is the similarity in chemical shift of the corresponding carbons and protons in the different rings of an oligosaccharide. Often signals are overlapped, which hinders assignments and prevents the use of such signals in any of the above experiments.

1.3.3 Force Fields and Molecular Dynamics Simulations

Considerable interest has recently centred on developing accurate theoretical models for carbohydrates using force fields and computer simulations.[22–25] A strong motivation for this is the difficulty of interpreting the experimental data for saccharides; in this respect, force-field calculations can often augment the results obtainable by experiment.[26] For example, a simple grid search of the glycosidic dihedrals can identify whether or not there are likely to be multiple minima, and thus whether the NMR parameters such as nOe's are averaged or not.

Molecular dynamics simulations using these force fields provide a method for studying the detailed dynamics of carbohydrates. Since a full set of coordinates is available throughout the simulation, almost any quantity of interest can be calculated. Thus a precise analysis of the dynamics may be done, and calculated dynamic parameters may be compared with those from experiment to verify that the simulation is valid.[27]

1.4 Objectives

The principal aim of this work will be to gain an understanding of the conformation and dynamics of the different starch linkages by considering small oligosaccharide model compounds.

To get an initial description of each type of linkage, molecular dynamics simulations of maltose (α-D-glucopyranosyl-(1→4)-D-glucopyranose) and isomaltose (α-D-glucopyranosyl-(1→6)-D-glucopyranose) both in vacuum and with explicit inclusion of water as solvent will be run (chapters 4 and 5). This will enable both the range, time evolution and relaxation behaviour of the principal degrees of conformational freedom discussed in section 1.2 to be calculated from the simulation data. From the analysis of the simulations in vacuum and water, conclusions about the effect of solvation in water and the differences in behaviour between the two types of linkage may be made.

To further mimic the branch point in amylopectin, a three residue model (panose, α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→4)-D-glucopyranose) and a four residue model (α-D-glucopyranosyl-(1→4)- [α-D-glucopyranosyl-(1→6)]-α-D-glucopyranosyl-(1→4)-D-glucopyranose) will be simulated in water under the same conditions as for the simpler models (chapter 6).

As a check on the validity of the force-field and simulations, and as a description of dynamics itself, ^{13}C longitudinal relaxation rates, T_1, will be calculated from the simulations. These will be compared with the corresponding experimental results obtained here and elsewhere.

Chapter 2

Computer Simulation Methods

Due to the conformationally flexible nature of the starch components amylose and amylopectin, it is difficult to get a precise description of their solution behaviour experimentally. Thus far, no single crystal X-ray structure has been possible for large carbohydrates, although crystal structures for small di- and trisaccharide moieties have been solved.[12-14] The available fibre diffraction data are only able to provide limited insights. Traditional spectroscopic techniques such as NMR nuclear Overhauser effect (nOe) or Circular Dichroism (CD) spectroscopy (§1) are not easily applicable when a large number of conformational variants exist in solution, as the measurable properties are averaged over all the structures present. In addition, the similarity between the residues in oligosaccharides limits the amount of specific information that can be extracted spectroscopically. Computer simulation with explicit inclusion of the solvent, however, is able to provide a detailed time evolution of the system, allowing almost any equilibrium physical quantity to be calculated.

2.1 Molecular Mechanics Formalism

2.1.1 Background

The complete description of systems of atomic and molecular scale is provided by quantum mechanical theory. However, for species consisting mainly of heavy atoms, a great many physical properties may be succesfully calculated without recourse to the computationally intensive task of solving Schrödinger's equation for a realistically sized problem. This convenient simplification is provided by the Born-Oppenheimer approximation,[29] which treats the electrons and nuclei separately. The molecular Hamiltonian may be

written

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

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

Because the nuclei move much more slowly than the electrons (the mass of the lightest nucleus is about 1840 times heavier than an electron), it is possible for the purposes of the electronic problem to consider them as being fixed and neglect T_{nucl}, while $V_{nucl-nucl}$ is simply a constant. Since the addition of a constant to the Hamiltonian will only add a constant to the eigenvalues, with no effect on the eigenfunctions, the electronic wavefunctions can then be obtained by solving the stationary state Schrödinger equation with the electronic Hamiltonian, given in equation 2.2.

$$\mathcal{H}_{elec} = T_{elec}(\{\mathbf{r_i}\}) + V_{elec-elec}(\{\mathbf{r_i}\}) + V_{elec-nucl}(\{\mathbf{r_i}, \mathbf{R_i}\}) \quad (2.2)$$

This gives a set of "electronic" wavefunctions, which depend explicity on the electron positions only (since they are functions of $\mathbf{r_i}$ only), while implicitly depending on the set of fixed nuclear coordinates chosen (since these appear as parameters in the Hamiltonian). Thus for each set of nuclear coordinates, a different set of electronic wavefunctions would need to be generated. This is the domain of electronic structure theory, or *ab-initio* methods. The approximation is usually very good, leading to errors in the energy of the order of 10^{-4} for the hydrogen molecule,[29] and better for molecules with heavier nuclei. It is certainly negligible compared to the other approximations involved in actually solving the Schrödinger equation for the electrons.

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

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

Since the nucleus-nucleus repulsion also depends only on the $\mathbf{R_i}$, the nuclei are effectively moving in a potential entirely described by their coordinates (equation 2.4). Thus the nuclear hamiltonian is given by equation 2.5, which can be solved in the nuclear coordinates only. The resultant nuclear wavefunctions $\Phi_{nucl}(\{\mathbf{R_i}\})$ describe all the vibrational, translational and rotational properties of the molecule, and therefore form a

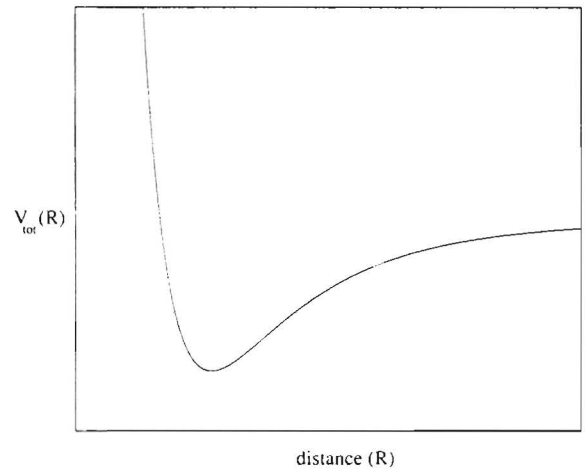


Figure 2.1: Schematic nuclear potential for a diatomic system

basis for the calculation of macroscopic properties from statistical mechanics.

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

$$\mathcal{H}_{nucl}(\{\mathbf{R_i}\}) = T_{nucl}(\{\mathbf{R_i}\}) + V_{tot}(\{\mathbf{R_i}\}) \quad (2.5)$$

The actual form of the potential energy described by the two rightmost terms of equation 2.5 has a form similar to the classic Lennard-Jones potential (figure 2.1) in the special case of diatomic systems. This consists of long range attractive interactions due to the electronic term and short range repulsive interactions due to the nucleus-nucleus repulsion as well as contributions from the electronic potential related to the pauli exclusion principle.

2.1.2 Force Fields

For molecular mechanics, the significance of the Born-Oppenheimer approximation is that the potential function for the nuclear Hamiltonian (equation 2.5) is a function of the nuclear coordinates only. However, the system is not treated quantum mechanically using equation 2.5, but by a set of classical empirical[1] potential functions known as a *force field*. Because they are sufficiently heavy, the nuclei can be treated classically: a common criterion which is applied is that the de Broglie wavelength (h/p) is much less than the distance between the nuclei; alternatively, one can observe that at room temperature the available nuclear energy states form a continuum ($\triangle E \ll k_B T$), which is not the case for electrons.[30] Most force fields consist of pairwise potentials for the non-bonded interactions and pairwise and higher order potential forms for intramolecular

[1]The potential may actually be derived both from *ab initio* calculations and experimental results

bonded, angle and dihedral terms. The CHARMM[31] force field can be written as the sum of bond, angle, dihedral and other terms:

$$E = E_b + E_\theta + E_\phi + E_\omega + E_{VdW} + E_{el} \tag{2.6}$$

These parameters can be broken into non-bonded ($E_{VdW} + E_{el}$) (often referred to as "non-constrained") and "constrained" ($E_b + E_\theta + E_\phi + E_\omega$) parts, which are discussed below. The internal energy terms are:

Internal Energy Terms

$$E_b = k_b(r - r_0)^2 \tag{2.7}$$

$$E_\theta = k_\theta(\theta - \theta_0)^2 \tag{2.8}$$

$$E_\phi = \sum_{n \in N} k_\phi(1 - \cos(n\phi)) \tag{2.9}$$

$$E_\omega = k_\omega(\omega - \omega_0)^2 \tag{2.10}$$

Harmonic terms are used to describe the deformations of each bond (2.7) and angle (2.8) from certain "reference" values. These references need not necessarily be the equilibrium values, since those are obtained by considering the whole force field, which will have many competing terms. Large values of k_b and k_θ are assigned to the bond and angle terms, and consequently these are not responsible for much of the variation between different conformations. Each type of bond or angle will generally have a different force constant, depending on the chemical properties of the participating atoms.

For each pair of dihedrally related atoms (A-X-X-B), there is a fourier potential (equation 2.9). This potential is usually represented by only a few terms of the series and is frequently dominated by a term with $n = 2$ or $n = 3$, corresponding to minima in planar or gauche geometry respectively. For any bond, there is generally such a term for every pair of 1,4 related atoms across that bond (again a function of the terminal atoms A and B), so that by defining general parameters k_ϕ for each type of atom pair, one can get the dihedral contribution to bond rotation simply by summing over all the 1,4 related pairs. Since dihedral force constants are much lower than those for the bonds and angles, they are more significant determinants of conformation, the dihedral angles being the principal degrees of freedom in flexible biomolecules.

Analogous to the dihedrals, although with a different functional form, are the so-called improper dihedrals (equation 2.10). These are defined between four atoms A-B-C-D where B and C are not actually bonded. This type of term is used to enforce local geometry, for example, at a tetrahedral centre or in a planar aromatic ring.

Every atom in the CHARMM force field is assigned a specific type, according to what type of molecule it is in and its position in the molecule. Thus in the carbohydrate

force field used here, there are different atom types for each carbon in a glucospyranosyl ring and there is even a distinction between the α and β forms for some atoms. The parameters used in the force field are specific to the atom types involved, allowing them to be highly "tuned". Other more general force fields have only one type for, say, a generic carbonyl carbon.

The non-bonded interactions between and within molecules, comprising the closed-shell Van der Waals and electrostatic energies are modelled using the equations 2.12 and 2.11 respectively. The electrostatic potential is cast in its standard form, usually with a dielectric of 1. The charges used are generally partial in order to reproduce bond polarity, and their values are very important in the final force field.

Nonbonded Terms

$$E_{el} = \sum_{i>j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.11}$$

$$E_{VdW} = \sum_{i>j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \tag{2.12}$$

The Van der Waals terms are calculated using the standard Lennard-Jones curve (2.12). This is an example of an "effective pair potential", since it tries to emulate a potential which has significant higher-order character (e.g. three-atom terms) by means of an effective pairwise potential. The hetero-atomic parameters σ_{ij} and ϵ_{ij} are calculated from the corresponding homo-atomic parameters using 2.15 and 2.16, unless the interaction is fixed for a particular pair. Some force fields include a term similar to the Lennard-Jones potential to describe hydrogen bonding (the 10-12 relation); however most modern force fields, such as those used with CHARMM, attempt to include hydrogen bonding in the electrostatic part of the potential.

$$A = 4\sigma_{ij}^{12}\epsilon_{ij} \tag{2.13}$$

$$B = 4\sigma_{ij}^{6}\epsilon_{ij} \tag{2.14}$$

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{2.15}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{2.16}$$

The compilation of force field parameters is a comprehensive exercise, attempting to include as many experimental and theoretical results as possible, within the conditions of interest. A large set of model compounds known as a "training set" is chosen, and

from the parameters derived from these, a more general force field can be constructed. Out of the terms 2.7-2.12, the bond and angle contributions are the most straightforward to parameterize, the reference positions being readily available (from x-ray and neutron diffraction, quantum chemical calculations and other experimental methods) as well as the force constants (from IR-Raman spectroscopy and *ab initio* calculation). In addition, the bond and angle terms are reasonably independent of each other and of the other terms, making them easily transferrable between different force fields.

Despite being the most important in determining the distribution of conformations, the dihedral and non-bonded terms are the most difficult to parameterise. NMR data such as ^3J coupling and nOe buildup rates can give information on rotameric distributions for some dihedral angles, as well as the rotation frequency in some cases. However this is usually limited to only a few dihedrals per molecule. Since the energies of rotamers in the force field will be determined by the dihedral, Van der Waals and electrostatic energy terms, it is hard to separate the influence of these and consequently they are the source of greatest variation between force fields. Dihedral energies are often extracted from quantum chemical calculations in modern force fields, while Van der Waals radii are available from solid-state data. Since there is no way to measure "atomic charges" (which are an artificial construct in the first place), they must be chosen so as to reproduce the dipole characteristics of the molecule. A common method is to fit the charges to reproduce an electrostatic potential calculated by an *ab initio* method, as was done in deriving the AMBER force field.[32]

2.2 Molecular Dynamics

Once a satisfactory potential energy function is available, it is possible to generate configurations of the system of interest in a suitable ensemble (§2.2.4). One approach to this is the use of *Monte-Carlo* (MC) methods, for example that of Metropolis,[33,34] in which configurations are randomly generated and selected based on a Boltzmann criterion. The method used in this thesis is a simple integration of the Newtonian equations of motion 2.17 and is known as *Molecular Dynamics* (MD). Since the MD approach provides a full time evolution of the system, it enables time correlation functions and transport properties, such as diffusion to be calculated directly. There are several recent reviews, of the method.[35-37]

$$\frac{\delta^2 x_i}{\delta t^2} = -\frac{\nabla V(x_i)}{m_i} \tag{2.17}$$

Although virtually all molecular dynamics simulations to date have been done using a force field potential, it should be noted that "first-principles" molecular dynamics has

recently been implemented.[38] While still using the Born-Oppenheimer approximation and classical equations of motion, this recalculates the ground-state electronic configuration and potential at each timestep by applying the technique of "simulated annealing" to the electrons. This has been applied to glucose in water,[39] however the small system size (12.42Å in the glucose simulation), short simulation times (less than 10ps) and very computationally intensive calculations limit the utility of such methods at present.

2.2.1 Integration of Equations of Motion

Given that the generation of MD simulation data involves integrating a large number of differential equations, it is important that the algorithm be efficient. The most computationally intensive part of a molecular dynamics simulation is the calculation of forces at each timestep, which far exceed the computational requirements of even the most demanding integration algorithm. Consequently, the most important feature of any integration algorithm is the ability to support a long integration timestep with small numerical error, rather than being fast itself. One of the most stable and the most commonly used method is the Verlet algorithm[40] given in equation 2.18

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t) \tag{2.18}$$

A deficiency in the standard Verlet algorithm is that a very small term ($\delta t^2 a(t)$) is being added to a much larger term ($2r(t) - r(t-\delta t)$), which can lead to loss of precision in computer floating point arithmetic. One improvement on this is the *leap-frog* algorithm (2.19,2.20) which first calculates the velocities halfway between the timesteps using equation 2.19; the coordinates are then obtained from the previous set using the velocities calculated at the half-step 2.20, so that the coordinates and velocities appear to "leapfrog" over each other. This process can be shown to be formally equivalent to the standard Verlet algorithm (2.18), while avoiding the need to include terms in δt^2 and the consequent inaccuracies.

$$r(t+\delta t) = r(t) + \delta t v(t+\tfrac{1}{2}\delta t) \tag{2.19}$$

$$v(t+\tfrac{1}{2}\delta t) = v(t-\tfrac{1}{2}\delta t) + \delta t a(t) \tag{2.20}$$

2.2.2 Potential Truncation and Switching

The time taken for an MD simulation increases rapidly with the size of the system: for a system with N molecules, it is around $\mathcal{O}(N^2)$ due to the double loop used for calculating the forces. In addition to this practical consideration, it is also undesirable to have long range forces, especially in a system with periodic boundaries (§2.2.3), since they lead to long range order, which is not what is wanted in a simulation of bulk solvent. A

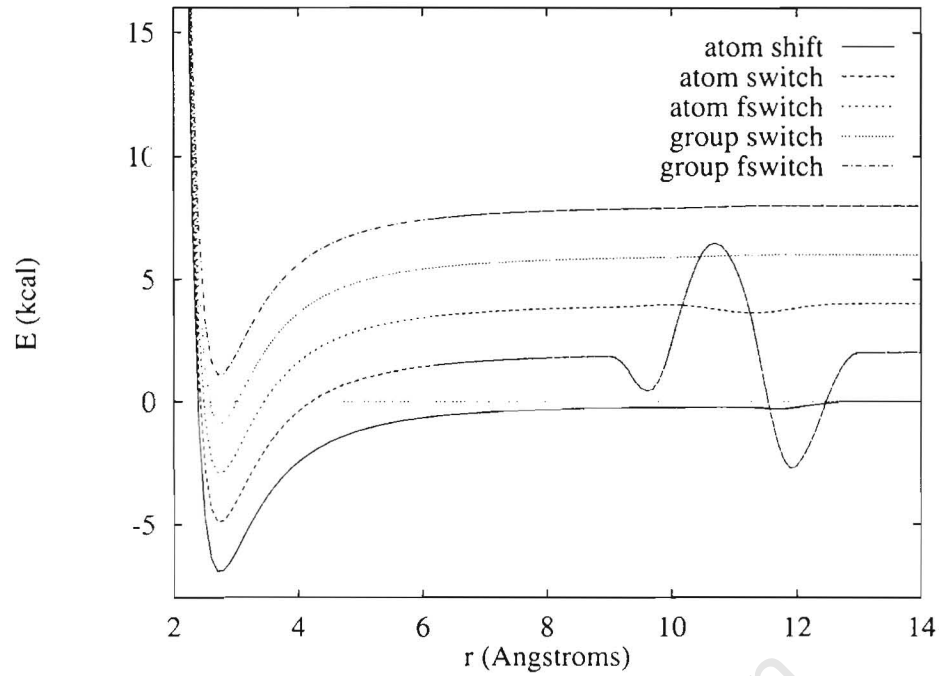


Figure 2.2: Dimer energy (kcal/mol) for two SPC/E water molecules in a linear hydrogen-bonded geometry as a function of O-O distance, using several cutoff methods provided by CHARMM (offset by 2 kcal).

frequent practice is simply to truncate the pair potentials for the nonbonded (Van der Waals, electrostatic) terms at a certain distance from each atom (present cutoff radii generally lie between 9Å and 12Å), since the pair potential is usually negligible at this distance. This introduces serious errors, since the potential is no longer continuous, so the system is non-conservative and the total energy therefore cannot be expected to be conserved. In such a case it is necessary to either scale velocities or couple the system strongly to a heat bath to keep the temperature constant. Both methods are unphysical and potentially detrimental, since they may obscure underlying defects in the simulation setup. An especially damaging feature of the atom-based cutoff is that it causes dramatic fluctuations in the molecular potential near the cutoffs (2.2). Consider two water molecules, represented by point charges on each atom: Although the dipole-dipole electrostatic potential for the two molecules drops as $1/r^3$, this in fact arises from the sum of fairly large monopole-monopole interactions which only decay as $1/r$. Therefore cutting off the atomic (monopole) terms one by one causes large oscillations in the electrostatic potential near the cutoff distance,[41] a phenomenon exhibited by the atom-based switching feature of CHARMM in figure 2.2. A solution to this problem is provided by dividing the system into groups of neutral charge and basing the cutoff on some property of the group (e.g. the centre of mass or, better, when any atom in the group is within the cutoff). Since all the atom-atom contributions to the $1/r$ decay of

the dipolar interaction are switched off simultaneously, it does not exhibit any of the wild fluctuations seen with atom-based cutoffs. However, the resultant potential is still discontinuous and suffers from constant heating of about $4K/ps$.[42] This is due to the fact that a group entering the sphere of influence of another is generally at a higher energy than one leaving, due to the ordering that takes place while in the region of interaction.[43] A net increase in energy results, and indirectly an increase in temperature.

This problem can be overcome by applying a *switching* or *shifting* function on the groups, so that the potential decays smoothly to zero, with no discontinuities in the force, and no loss of energy conservation. A switching function is multiplied by the potential to scale it to zero over a short range near the cutoff, while a shift function does the same sort of thing, but over the whole range. In CHARMM, the switching function is implemented as shown in 2.21 and the standard shifting function as shown in 2.22.

$$S(r, r_{on}, r_{off}) = \begin{cases} 1 & \text{for } r \leqslant r_{on} \\ \frac{(r_{on}-r_{off})^2(r_{off}+2r-3r_{on})}{(r_{off}-r_{on})^3} & \text{for } r_{on} < r \leqslant r_{off} \\ 0 & \text{for } r > r_{off} \end{cases} \tag{2.21}$$

$$S(r, r_{off}) = (1 - \frac{r}{r_{off}})^2 \tag{2.22}$$

It should be noted that in applying the switch and smoothing functions to group-based cutoffs, the same switch/shift factor should be applied to the whole group, so as not to reintroduce the fluctuation effects the groups were designed to avoid!

Group switching has been applied in all the calculations done here, since it solves all the problems mentioned above, and gives excellent energy conservation. That said, there are many other ways of producing good cutoff behaviour (see figure 2.2), for example, applying the switching function directly to the forces, so that the potential is now given by the integral of the switched forces ("force-based" methods[42]).

2.2.3 Periodic Boundary Conditions

Since the simulation time grows so rapidly with the size of the system, doing a simulation with enough explicit solvent molecules to represent the bulk liquid would be unmanageable, and we are forced to devise ways of making the small systems we can study behave like "bulk" solvent. The more primitive method of doing this is to place the solute in a "droplet" of solvent and hope that the solvent cohesive forces hold it together, or alternatively to place a surface potential on the drop to do this. However, this is a poor reproduction of the bulk, since the molecules on the surface are a significant part of the total, and would behave differently from those in "bulk" solution.

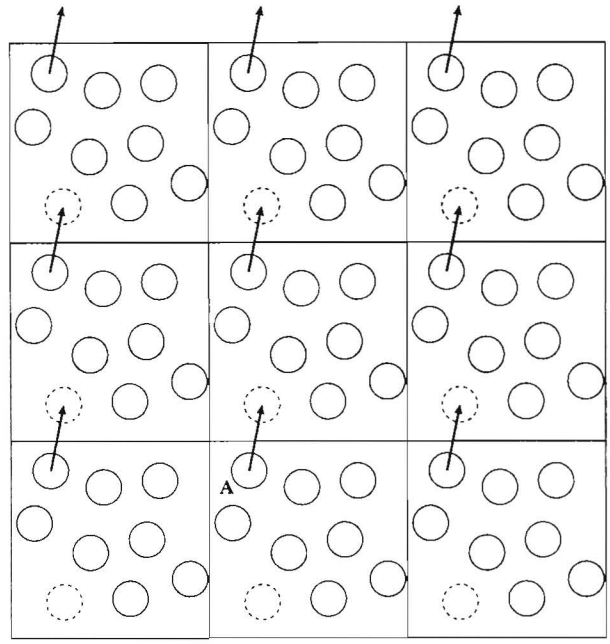


Figure 2.3: Periodic boundary conditions illustrated for a two-dimensional system. The "molecule" **A** is shown leaving the box from one side and re-entering from the other

Since a "bulk" solvent is generally what is desired, there must be some way of eliminating surface effects. This may conveniently be achieved by putting the system in a "cell" or "box" which is repeatedly tiled throughout space, so that a crystal is effectively simulated. If a particle passes out one side of the box, it immediately re-enters from the opposite side. An illustration of such *periodic boundary conditions* is illustrated for a two-dimensional system in figure 2.3. The formulation of the potential is exactly the same, except that atoms are now able to interact with image atoms. In principle, each atom now interacts with all of its infinite set of images as well as those of all the other atoms. This is undesirable for both practical and theoretical reasons: the calculation of such forces in a periodic system is possible, but more computationally intensive, and long range interactions in the potential will give rise to long range order, which is not desirable in a simulation of bulk liquid. For this reason, it is common to restrict the interaction of atom i to the nearest image (including primary atoms) of atom j, a condition known as the *minimum image convention.*

There is an inherent risk in setting up a simulation of solute in a periodic box that the system may behave too much like a crystal. A first guess at a *minimum* acceptable box size may be made by estimating a sphere, sufficiently large to always contain the whole solute, being placed inside the simulation cell (Figure 2.4 (a)). For the molecule not to interact with itself, the minimum side-to-side distance in the box must be at least the diameter of the sphere plus the cutoff distance for non-bonded interactions. However, in the prescence of solvent, it is necessary to also preclude any water molecules

from interacting with the solute twice. This may be enforced by extending the minimum side-to-side distance to the solute diameter plus two cutoff radii.

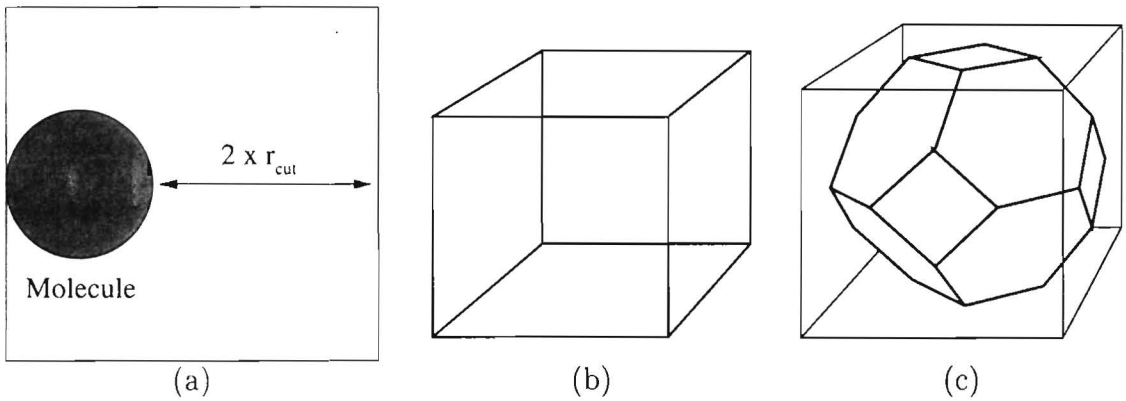


Figure 2.4: (a) Minimum cell size, (b) Cubic cell, (c) Truncated Octahedron cell

Any space-filling polyhedron can in principle be chosen as the simulation box; in fact only five are able to do so (the parallelepiped, hexagonal prism, the truncated octahedron, the rhombic dodecahedron and the 'elongated' dodecahedron). The merits of a box can be judged by how small a volume is required to satisfy the above minimum box size criterion. It would be expected that a more spherically shaped cell would have a smaller volume given a minimum side-to-side difference, and indeed this is the case. A truncated octahedron (Figure 2.4) is such a solid, obtained by cutting the corners from a cube so that it has exactly half the volume. It was first applied to the simulation of ionic systems,[44] in which the minimum side-to-side distance must be as large as possible due to the long-range potentials. The ratio of the volume of the inscribed sphere to the volume of the simulation box is 0.52 for a cube, but 0.68 for a truncated octahedron (TO). Although the complexity of implementing the minimum image convention is marginally increased for a TO box, it is more than offset by the reduced number of atoms, since the time taken for force calculations is dominant.[45] However, periodic boundaries are sufficiently complex for the other space-tiling shapes, to make their use prohibitive. A recent argument[46] has been made that a trucated octahedron can be transformed to a cube and be more effectively implemented in that form, but this entails rewriting the molecular dynamics code. Certainly, the cubic simulation box is much more versatile, allowing facile implementation of constant pressure dynamics and the application of sheer stresses to the system.

Although the use of periodic boundary conditions is probably the best for dealing with homogeneous solutions under equilibrium conditions, it is clearly not suitable for modelling solid/liquid interfaces or a restricted region of a macromolecule such as an enzyme active site. It is also limited to modelling fluctuations which have a wavelength

less than that of the simulation box, longer wavelength fluctuations being suppressed. For this reason, it cannot be used to study certain phase transitions. However, for water under biological conditions, it is adequate.

2.2.4 Simulation Ensemble

Computer simulations are most credible if they can be compared with the results of a physical experiment. Thus, it would be ideal to run the dynamics in an ensemble that corresponds closely to experiment, the most obvious choices being under conditions of constant temperature and volume (NVT) or temperature and pressure (NPT). Various methods have been proposed for obtaining these ensembles using have been proposed.[36,37,47] The NVT, or canonical, ensemble is the easiest of the two to implement in a program, since the periodic box has a fixed volume. The temperature would be kept constant by rescaling the velocities, coupling the system to some sort of heat bath, or using the Langevin equations of motion. Rescaling velocities has the disadvantage of being an unphysical effect, and uniform scaling leads to the phenomenon of "hot solvent, cold solute". Coupling to a heat bath also introduces unphysical phenomena, while the randomness of Langevin dynamics prevents the reliable calculation of long term time correlation functions. It was therefore decided to use the microcanonical (NVE) ensemble for the current simulations (with a few named exceptions), on the grounds that it would be best for the calculation of correlation functions for measuring experimentally available parameters. Once the system has equilibrated, the potential and kinetic energy will be stably partitioned and the system will maintain a near constant temperature. Thus the physical treatment of the problem is strictly correct, even though there are no experimental conditions which correspond exactly to the simulation conditions.

2.2.5 Limitations of the Molecular Dynamics Method

Molecular Dynamics has the attractive feature of providing precise "trajectories" with atomistic detail, allowing the simulation to be probed and data to be extracted, often far more easily than from experiment, in addition to providing properties which are inaccessible via experiment. Furthermore, it allows the user to exactly control and manipulate conditions to test for the effect of modifications, so as to investigate the molecular basis for a particular conformation, for example. However, several issues must be borne in mind when drawing conclusions. Firstly, one has to consider how effectively conformational space has been sampled; commonly this is done by starting the simulations from different points and by running them for long periods, typically more than one nanosecond. By comparing the results with adiabatic vacuum maps,

one is able to assess how well the vacuum minima have been sampled. Secondly, as the results are obtained from an approximate force field, they may not reflect properly the appropriate experimental distributions; thus as many experimentally measurable static and dynamic properties should be calculated and compared to experimental benchmarks. Finally, since the simulation is intended to reproduce solvation in a bulk solvent, the effect of the system size on the results ought to be taken into account, possibly by comparing the results for different system sizes.

2.3 Analysis Methods for Condensed Phases

Molecular dynamics simulations provide a wealth of information on an atomistic scale, which, beyond simply reproducing experimental results, can be probed in order to gain further insight into the behaviour of solutes in the condensed phase. It is possible to obtain precise structural information such as solvent distributions about the solute and details of the hydrogen bond patterns, as well as dynamical properties, such as diffusion, rotation and vibrational modes. By applying such analyses to carbohydrates, one should be able to extract useful information, which will aid in the design of carbohydrate-based drugs, polymers and transport agents.

2.3.1 Structural Analysis

The Pair Distribution Function

In analyzing the local structure of liquids, the most common method is the calculation of the pair distribution function (PDF), $g(r)$. In the case of simple monatomic liquids, the function $g(r)$ is defined as the ratio of the probability of finding two atoms a distance r apart, relative to the probability of finding them at that distance in a completely random distribution, and may be cast in the form given in equation 2.23.

$$g(r) = \frac{1}{4\pi\rho r^2}\frac{dN(r)}{dr} \tag{2.23}$$

Figure 2.5 shows this function for liquid water, as measured experimentally using neutron diffraction,[48] which is typical of the pair distribution for a liquid, being intermediate between a gas (no structure beyond first peak) and a crystal (perfectly repeating). From this plot, it is possible to obtain the positions of nearest neighbours to the atom in question, as well as the total number of atoms which form part of the "solvation shell". This latter quantity can be calculated from the running integral of the density. The pair distribution can easily be extended to molecular liquids such as water by calculating site-site distributions, for example O-O or O-H PDF's. Furthermore, the distribution of

solvent around solute sites can be analyzed using a site-site $g(r)$ with a solute atom or residue as one of the sites.

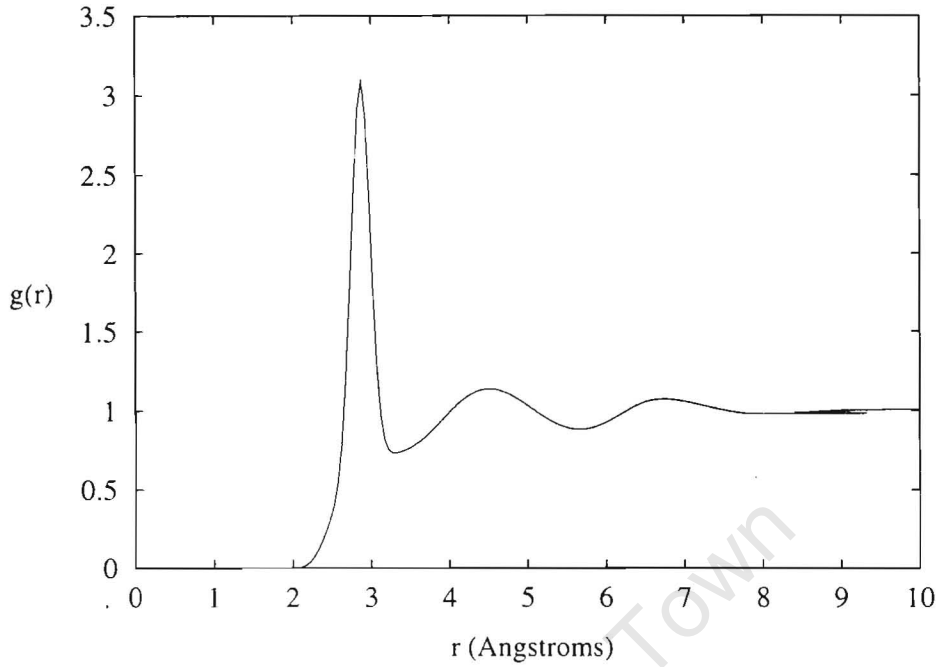


Figure 2.5: Experimental pdf $g_{OO}(r)$ for water[49]

Not only does the PDF give important structural information, it can also be used to calculate any thermodynamic pair function, since it provides complete information about the distribution of pairs within the configuration. In addition, the structure factor of water, which can be measured by X-ray and neutron diffraction, is also simply related to the pair distribution. Thus the calculation of the PDF for a solvent model, for example, is vital for comparison with experimentally measured data. However, the pair distribution is even more significant in the case of solute-solvent interactions; it cannot be obtained directly from experiment, due to the radial averaging of the experimental observables. Computer simulations are able to supply the information in these cases, however.

Three-dimensional Solvent Structure around Molecules

Since the information available from simple one-dimensional pair distribution functions is radially averaged, it is limited to giving information about the number of neighbours for each site as well as their distance from the site. However, we would like to know more precisely the location of the water molecules around the solute and their role in solvation, requiring a detailed three-dimensional picture.

tion. This has been done for pure water[49] by defining a molecular spherical polar coordinate frame about a reference water molecule and calculating two dimensional pair distribution functions such as $g(r, \theta)$. An alternative method is the calculation of $g(r_1, r_2)$, the probability of finding a solvent site at distance r_1 from solute site 1 and distance r_2 from solute site 2 relative to a random distribution, although this requires more care in normalisation. This approach was adopted by Immel and Lichtenthaler[50] in the analysis of the solution structure of sucrose. It is able to give explicit information on "bridging" water molecules which are probably involved in simultaneous hydrogen bonding to both solute sites. In doing an analysis of this type, it is important not to average over different conformations of the solute, as the probabilities will then be averaged over all conformations and will be extremely hard to interpret.

A more detailed picture of the solvent structure may be obtained by the calculation of a three dimensional probability distribution function for a particular molecular conformation (the above comments about averaging apply equally here).[49,51–54] The procedure can be summarized as:

1. Frames are selected from a trajectory according to their match to a specific conformation, in order to remove conformational averaging effects.

2. Each frame is translated and rotated to obtain a fit with a reference structure, in order to eliminate interference from translational and rotational diffusion.

3. For each solvent molecule, the site of interest (usually the oxygen of water) is "binned" into a three-dimensional grid defined with respect to the reference coordinate set. Either the actual distribution of atoms can be binned as in the one- and two-dimensional cases, or a Gaussian function of the form given in equation 2.24 is assigned to each nucleus (representing electron density) and this is binned. In 2.24 Z is the number of electrons on the atom, while a is chosen so that the function drops to 10% of its maximum value at the Van der Waals radius. In principle, the electronic distribution can always be recovered from the nuclear one by integrating the gaussian over the nuclear distribution function.

$$g(\vec{r}) = Z \times (\frac{a}{\pi})^{\frac{3}{2}} \times e^{-a \times r^2} \tag{2.24}$$

4. The density is normalized analogously to the one- and two-dimensional cases.

Hydrogen Bonding

Many of the well-known anomalous properties of water are due to its strong network of hydrogen bonds, which also give it a very ordered structure. Carbohydrates, having multiple hydroxyl groups, are able to participate in this network, explaining the

solubility of small saccharides. Insight into the nature of solvation can be gained by considering how the solvent disrupts vacuum hydrogen bonded-structures, how many hydrogen bonds are formed by the solute and their lifetime. The influence of bridging hydrogen bonds in which two solute donor/acceptors bond to the same solvent molecule.

The ideal criterion to use in hydrogen bond analysis would of course be the total non-bonded energy between the hydrogen bonded partners. This is because it is the energy which actually defines whether there is a hydrogen bonded interaction or not, and the various geometric considerations which are usually applied do not allow for distortions of the bond.[55] The geometric methods therefore often give a reduced hydrogen bond lifetime (0.05-0.3ps), and possibly a reduced average number of such bonds relative to what would be obtained by a full energy analysis (which typically have a lifetime of 6-20ps). Although this has been used to good effect when the hydrogen bond is between two small molecules such as water, it is more difficult to compute for a large solute such as a di- or polysaccharide.[56] Such a calculation would include in the energy calculation the effect of remote groups which have little to do with that hydrogen bond. An alternative method has been proposed[55] which uses a redefinition of the energy function for analysis purposes: the charge on the carbon of the C-OH pair is adjusted to make it neutral, and the hydrogen bond is defined as existing when a rolling average of the pair energy between the C-OH and the water molecule drops below a critical value, E_{HB}. However, this method really requires the energy to be written out at every step, which makes it awkward for routine use.

In all simulations analysed here, hydrogen bond analysis was done based on a distance criterion of 2.4Å for the acceptor-hydrogen distance and an angle criterion of 120^o. For the analysis of intramolecular hydrogen bonding, the angle criterion was relaxed to 100^o to include the strained bonds between the secondary glucosyl alcohols. Hydrogen bond lifetimes were evaluated as a simple average between breakages. Although this gives shorter lifetimes than an energetic definition, it can nonetheless be used as a basis for comparison. Since structures are only written every 50 fs in the dynamics simulations, some information on "lifetimes" has already been lost.

The hydrogen bonds from each hydroxyl in solution can be divided into those made to the solute, those made to water, and those involving a "bridging" water molecule between two solute hydroxyls. The most common kind of bridge found in carbohydrates is that between adjacent hydroxyl groups, although interresidue dihedrals have been observed.[50] Some possible ways in which water can bridge in carbohydrates are shown in figure 2.6.

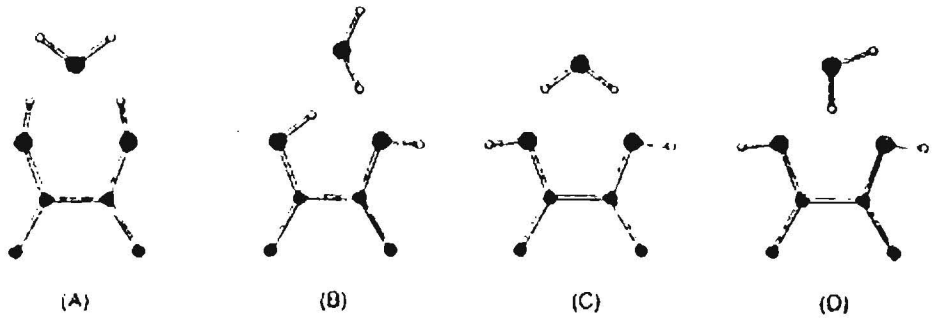


Figure 2.6: Four possible orientations of a water molecule bridging two adjacent hydroxyls in a carbohydrate.[56]

2.3.2 Correlation Functions and Transport Properties

In order to extract properties such as diffusion, shear viscosity, rotational times and similar transport properties, from an MD simulation, one needs a method of relating events at different times. For example, in the calculation of diffusion coefficients, one is interested in the rate at which molecules move through the solution. A time correlation function gives an indication of how "correlated" events at different times are. This allows the calculation of transport properties, while the power spectrum of certain time correlation functions is directly related to experimental spectra.

Correlation Functions

The correlation of two quantities x and y is often measured by the correlation function $\langle xy \rangle$ or $\langle x \cdot y \rangle$ for vector quantities. If the quantities are uncorrelated and have means of zero, the correlation function will average to zero too, while if there is a correlation, the function will be either positive or negative. Normalization is possible by dividing by the product of the mean values for x and y.

An extension of the concept of a correlation function is to consider x and y to be measured at different times, usually for the same physical quantity. Thus the *time correlation function* of a dihedral angle ϕ would be defined by equation 2.25, where the average is over all the time origins possible for each value of t. Note that in the case of the dihedral, one has to subtract the mean (i.e. look at the fluctuations) in order to get meaningful correlations.

$$C_\phi(t) = \langle \Delta\phi(t)\Delta\phi(0) \rangle \tag{2.25}$$

Thus the correlation of ϕ with the value at a later time may be examined; a schematic example is given in figure 2.7. If the decay is assumed exponential, with the general form $Ae^{-\frac{1/\tau}{t}}$, then the time τ is known as the correlation time. The time correlation function

gives an indication of how quickly the variable becomes uncorrelated (the *correlation time*) as well as any regular motion (manifested as later peaks in the function). Further insight into the frequency of such regular motion may be obtained by calculating the power spectrum of the correlation function by fourier transformation:

$$I(\omega) = \int_0^\infty \cos\omega t \langle \Delta\phi(t)\Delta\phi(0)\rangle dt \tag{2.26}$$

This spectrum will contain peaks corresponding to the regular motions in the time correlation function, giving an indication of what motions are responsible for the decay, and their frequency. A schematic of a typical power spectrum is given in figure 2.8.

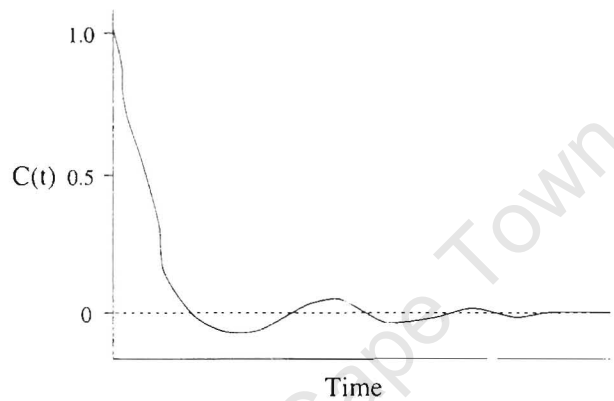


Figure 2.7: Schematic Time Correlation function

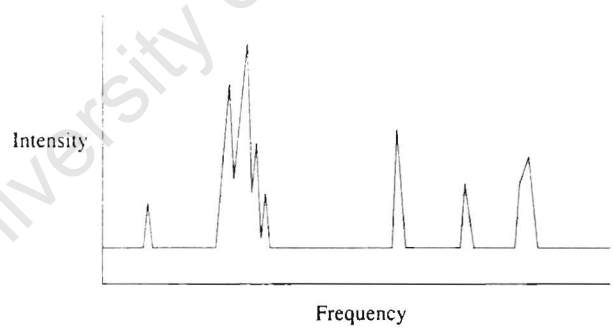


Figure 2.8: Schematic Power Spectrum

Internal and Overall Orientational Relaxation

NMR relaxation parameters depend on correlation functions of internuclear vectors. The motion of such vectors within the laboratory reference frame may be approximately divided into a global reorientation of the molecule as a whole and internal fluctuations such as the movement of side chains in a protein. These motions are also of general interest in polymer science. Approximate correlation times for these two processes may be calculated from simulation using appropriately defined correlation functions. The

RMS fit autocorrelation function is defined by $C(t) = \langle RMS(0,t) \rangle$ where the function $RMS(t_a, t_b)$ is the best RMS fit between the structures at time t_a and t_b. The decay of this function indicates the rate at which the internal degrees of freedom of a molecule become uncorrelated. Various functions could be used to evaluate the tumbling rate of a molecule. That which is relevant to NMR relaxation is $C(t) = \langle P_2(\vec{\mu}_{mol}(0) \cdot \vec{\mu}_{mol}(t)) \rangle$.

Transport Properties

Time correlation functions are not only useful for analysing conformational events; they can also be used to calculate transport properties such as diffusion and viscosity and physical properties such as the shear strain. One useful relation is the self-diffusion coefficient, D_s, defined by:

$$D_s = \frac{1}{3} \int_0^\infty \langle \vec{v}(t) \cdot \vec{v}(0) \rangle dt \tag{2.27}$$

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

$$2tD_s = \frac{1}{3} \langle |\vec{r}(t) - \vec{r}(0)|^2 \rangle \tag{2.28}$$

where $\vec{r}(t)$ is the displacement of the particle.

2.3.3 Conformational Analysis

Pucker Analysis

The closed pyranosyl form of glucose is a six-membered ring, which is able to adopt the same conformations as cyclohexane, such as the boat, chair and twist. These may be conveniently described by the generalised puckering parameters proposed by Cremer and Pople for cyclic systems.[57]

For six-membered rings, the Cremer and Pople puckering parameters are a magnitude Q, and two angles θ and ϕ which essentially define the ring conformation. Figure 2.9 illustrates the relation between the value of the puckering parameters and the familiar cyclohexaniod conformations.

Experimental evidence shows that the only significant glucopyranosyl conformer in solution is the $^1C_4$ chair. Analysis of simulation results with the Cremer and Pople protocol allows the simulated conformation to be checked.

Cluster Analysis

After a simulation of carbohydrates, it is necessary to group the various conformers visited according to structural similarity. Once a description of the predominant conformers has been obtained, it is easier to classify the results and to focus the analysis on

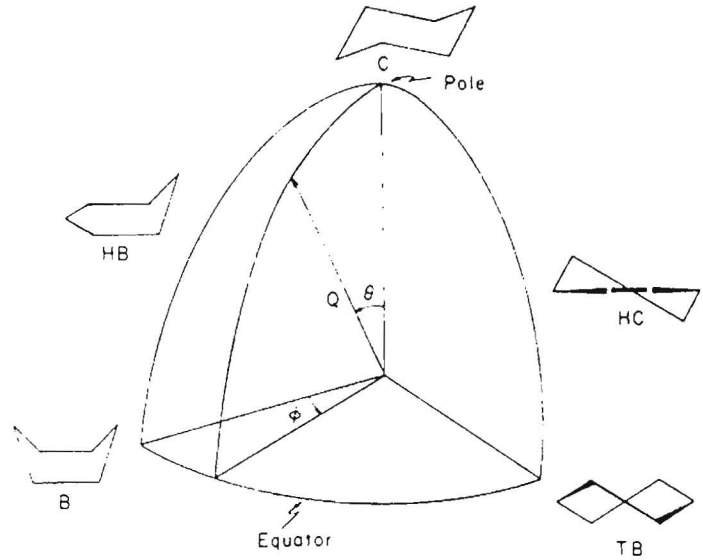


Figure 2.9: Cremer and Pople description of puckering as applied to six-membered rings[57]

the important structures. CHARMM provides an adaptive analogue pattern recognition scheme, ART2',[58,59] which is able to group structures based on a vector *pattern* of properties. The algorithm may be described as follows. The *centre* of a cluster is defined as the vector which is the average of all the vector patterns in the cluster. The *distance* between any two cluster members is usually defined as the simple rms difference of the two vectors. The *threshold distance* is the maximum allowed distance of a pattern from the centre of its cluster, while Err_{max} is a convergence criterion. The iterative cluster refinement then proceeds according to this scheme:

Initialization: The first pattern becomes the centre of the first cluster

Initial Cluster Assignment: For each new pattern i, find the closest cluster center j to i; if j is within the threshold distance of i, assign i to cluster j; else, make i the centre of a new cluster. Continue, until all patterns have been assigned a cluster.

Iterative Refinement: Take the previous set of cluster centres as a start and reassign all the patterns in a stepwise fashion as in 2.3.3 to form a new set of clusters; if the difference between this new set of centres and the previous is less than Err_{max}, then the assignment has converged and the algorithm is terminated. Otherwise this step is repeated.

A good measure of clustering is the proportion of cluster members that only lie within one cluster. If the data cluster well, it should be possible to group the elements fairly exclusively, so that most data will only belong to one cluster. The choice of cluster

radius is important, too: if it is too big, there will be overlap between clusters, while if too small, artificial clusters may form, which will also exhibit overlap.

2.4 Models for Water

Water is a liquid whose anomalous behaviour is well known (see review by Zhu[60]). The strong, highly directional hydrogen bonded network within the liquid give it a very ordered structure. X-ray[61] and neutron[48,62–64] diffraction data provide experimental insights into its structure, allowing the radial pair distribution function $g_{AB}(r)$ (§2.3.1) for nuclei A and B to be directly calculated. The experimental $g_{OO}(r)$ shows unusually strong second and third solvation shells for a liquid (figure 2.5). This is thought to be largely due to local tetrahedral structuring about each water molecule.

Given this complexity, the need to correctly treat hydrogen bonds and the fact that the light hydrogen nuclei ought to be treated by quantum methods, one would expect any successful water model to be fairly sophisticated. However, it turns out that some of the simplest models are in fact the best: the TIP3P[65] and SPCE/E[66] models are the most popular for biomolecular simulations and consist of rigid molecules with point charges on each atom. The parameters for these and two other models are shown in table 2.1. Of these, the "three-site" models (SPC, SPC/E, TIP3P) have a charge on each atom, while the "four-site" model (TIP4P) has a charge (q(M)) displaced from the oxygen by r(OM) along the H-O-H angle bisector. Besides the Coulombic potential, the remaining non-bonded interactions are represented by means of a Lennard-Jones potential placed on the oxygens, with none on the hydrogens. However, a set of small Van der Waals parameters has been introduced on the hydrogens in the CHARMM implementation[67,68] ($\epsilon_{min} = -0.046$, $\sigma_{min} = 0.449$) in order to avoid massively high energies on close approach of other atoms to the water hydrogens. The bonds are constrained by the SHAKE algorithm[69] in CHARMM, so they are not absolutely rigid.

Although previous attempts had been made to model water (such as four-site Bernal-Fowler[71] and five-site ST2[72] water), the SPC and TIP families were the first to be parameterized by fitting with experimental results, and they reproduce experimental properties such as density, heat capacity and pair distribution functions very well. More recent attempts have been made to develop flexible[73] and polarizable[74] models, but their increased demands on computational resources and small improvement in the reproduction of common experimental properties have limited their application in simulations.

Of the popular models, TIP4P and SPC/E give the best reproduction of solvent structuring, with the closest correspondence of their pdf's to experiment. This is thought to be because they favour a more tetrahedral local structuring of the water.[75] Indeed,

	SPC	SPC/E	TIP3P	TIP4P
r(OH), Å	1.0	1.0	0.9572	0.9572
∠ HOH, deg	109.47	109.47	104.52	104.52
A $\times 10^{-3}$, $Å^{12}mol^{-1}$	629.4	629.4	582.0	600.0
C, $Å^{12}mol^{-1}$,	625.5	625.5	595.0	610.0
q(O), C	-0.82	-0.8476	-0.834	0.0
q(H), C	0.41	0.4238	0.417	0.52
q(M), C	0.0	0.0	0.0	-1.04
r(OM), Å	0.0	0.0	0.0	0.15

Table 2.1: Common rigid point charge water models: SPC,[70] SPC/E,[66] TIP3P and TIP4P[65]

the recent calculation of a three-dimensional distribution function for SPC/E has shown that the average number of nearest neighbours occupying tetrahedral sites is in fact exactly 4.0.[49] Of the two families, it is generally found that the SPC-type models reproduce structural and diffusional characteristics best (although TIP4P is good here too), while the lack of structure in TIP seems to give it elevated self-diffusion rates. On the other hand, the TIP-type models reproduce best the experimental results over a range of pressure and temperature.[75]

The SPC/E model has also been shown to accurately reproduce the experimental density of water at standard simulation temperatures.

The TIP3P water model was used for all simulations in this work, on the grounds that the CHARMM protein and carbohydrate force fields have been parameterized using this model. Given that the structuring of TIP3P is known to be rather poor, it is important to know how much this affects its solvation of carbohydrate solutes. Liu and Brady have recently[52] compared the TIP3P and SPC/E models in a study of water structuring around mboxβ-D-xylopyranose. It was found that while the SPC/E showed significantly more localized structuring, both models gave the same qualitative results and agreed particularly on the position of the first solvation shell.

2.5 Models for Carbohydrates

There is now a great number of force fields for carbohydrates, using different functional forms, parameter sets and parameterization data. The force fields considered here all use a similar set of force field equations, and are in some sense inherited from each other. The first force field under consideration is that developed for CHARMM by Ha et al.[22]

(hereafter referred to by the authors' initials as the HGFB force field) using vibrational and crystallographic information for α-D-glucose and optimised for use with the TIP3P water model. Naturally, the force field gives a good reproduction of vibrational data, and the vacuum adiabatic (ϕ, ψ) energy map (see §4) for maltose has the crystal structure in a minimum.[23] However, it does not discriminate between α and β linkages, which are known to prefer different minima about the linkage dihedrals, a phenomenon known as the exo-anomeric effect, and the anomeric linkage oxygen is treated the same as the pyranosyl ring oxygen. This limits its applicability to oligosaccharides and glycoproteins.

Besides the two dihedrals ϕ and ψ describing an $\alpha(1\rightarrow4)$ linkage, the other principal degrees of freedom are the secondary alcohols and more particularly the primary alcohol. The conformation of the latter is conventionally given by a specifier for each of the dihedrals O5-C5-C6-O6 and C4-C5-C6-O6, according to whether each is *gauche* or *trans*. So, for O5-C5-C6-O6 *gauche* and C4-C5-C6-O6 *trans*, the conformation is described as *gt*. Using this notation, the favoured orientation for the primary alcohol in vacuum for the HGFB force field is *tg*, which is the same as that in the crystal structure.[76] However, while simulations in water with this force field allow frequent rotation about the C5-C6 bond, they show a preference for *tg*, which is generally not the experimentally predicted conformer, especially in solution. Solid state data for 101 glucopyranose structures gave a ratio *gg:gt:tg* of 60:40:0.[77] NMR $^3J_{HH}$ coupling and NOE results indicate approximately the same relative populations for D-*gluco*-sugars, with *gt* being slightly more favoured than in the crystal structure.[17] Thus this force field correctly modelled the primary alcohol rotations, which are believed experimentally to exchange every 10^{-10} to 10^{-9}s from ^1H relaxation,[17] but failed to predict the correct distribution of conformers. The rotation of secondary alcohols was observed to occur about once every 30ps, for which there is no experimental benchmark.

The original HGFB parameters have since been incorporated into the AMBER force field by Homans.[78] Specific atom types were created for α and β anomeric carbon, oxygen and hydrogen (i.e. C1, O1 and H1) and linkage-specific parameters were introduced. Bond lengths between the glycosidic carbon and oxygen and between the glycosidic oxygen and aglyconic carbon were fitted from crystal structure data. Several angles involving the anomeric centre were also adjusted to fit the crystal data. An extensive attempt was made to develop dihedral parameters for ϕ using *ab initio* calculations on dimethoxymethane as a model for glycosidic bonds, while ψ was left unparameterized since the non-bonded interactions were assumed to dominate. The charges were left unchanged from the HGFB values.

Liang[25] has revised the Brady force field, incorporating some of the Homans parameters and adjusting the primary alcohol dihedrals to favour the *gg* and *gt* conforma-

tions, and adapting it for the SPC/E water model. The resulting CHARMM force field (named PHLB after the authors' initials) gives a reasonable adiabatic map and vacuum behaviour, while also disfavouring the *tg* conformation. However, the rotation around the C5-C6 bond is greatly hindered, such that transitions between minima are hardly ever seen, which conflicts with the aforementioned NMR results. This has subsequently been corrected using potential of mean force calculations on the primary alcohol.[79] Since this is the best of the well-tested force fields, it has been used in all the simulations in this work.

The most recent force field is that proposed by Momany and Willet.[80] This too is based on the Homans force field, but makes a somewhat different set of changes. The biggest change has been in the charges, which have been taken from Woods *et al.*,[81] especially in the primary alcohol, in which the charges have been scaled down, which seems to allow for increased rotation about the C5-C6 bond.

Chapter 3

NMR Relaxation Theory and Experiments

In addition to providing detailed primary and three-dimensional conformational information about the macromolecular structure of proteins, NMR is able to give some information about the local dynamics around a given heteronucleus. The longitudinal and transverse relaxation times T_1 and T_2 and steady-state nOe, η, for a given heteronucleus X are, in most cases, only dependent on the motional characteristics of the vector X-H between X and its attached proton. The most frequently used methodology for extracting motional information from relaxation parameters is the model-free formalism of Lipari and Szabo[21,82] (below), which assigns to each nucleus an *order parameter* S^2 related to the motional restriction of the X-H vector and an *effective correlation time* τ_e which describes the rate of motion of the vector. As will be shown below, it is possible to calculate both the experimental measurables and the order parameters directly from a molecular dynamics simulation. These can then be used as a comparison with experiment.

3.1 Measurement of NMR Relaxation Parameters

Since the separation of nuclear energy eigenstates is very small in NMR, the time taken for the restoration of equilibrium after a radio-frequency pulse is applied is sufficiently long to be measurable. This is in contrast to other forms of spectroscopy, in which relaxation may be too rapid to be observed, such as infra-red and UV-visible spectroscopy. A variety of relaxation phenomena may be measured, such as longitudinal and transverse relaxation, and the nuclear Overhauser effect.

In the classical Bloch description of NMR, the effect of the static magnetic field pointing along the +Z axis manifests itself as a net nuclear magnetic moment aligned

along the +Z direction. This is of course the resultant of the magnetic moments of all the nuclei in the sample, with a slight balance in favour of the lower energy state aligned with the field (in the case of spin $\frac{1}{2}$ nuclei). The relaxation of the Z component of the bulk magnetisation vector (for each nucleus) after a population inversion (such as after a π_x pulse) is known as the longitudinal (or spin-lattice) relaxation of that nucleus. This is due purely to the actual return of the nuclear energy populations to equilibrium, by whatever relaxation mechanisms are operating, since the Z component arises only as the sum of the two nuclear spin states. The name "spin-lattice" refers to the fact that this relaxation occurs as a result of the dissipation of energy from the nuclei to the environment.

The change in the Z-component of magnetisation (figure 3.1) after population inversion is usually described by an exponential decay with time constant $\frac{1}{T_1}$, that is:

$$\frac{d(I_z(t) - I_{z0})}{dt} = -\frac{t}{T_1} \tag{3.1}$$

Since the longitudinal relaxation rate T_1 is related to the local motion of the nucleus in question, it is possible to draw conclusions from relaxation measurements alone as to the extent and rate of atomic motion (§3.2), for each observable nucleus.

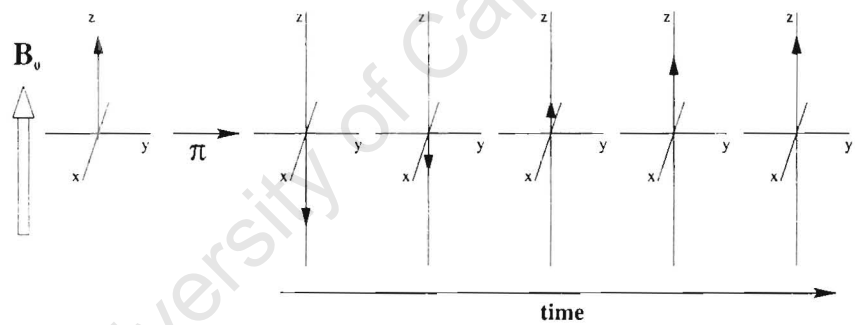


Figure 3.1: Evolution of bulk magnetisation as a result of longitudinal relaxation after a π_x pulse

A separate relaxation parameter is the rate of decay of the magnitude of the magnetisation in the xy-plane after a $\frac{\pi}{2}$ pulse. Since the net magnetisation in the xy plane after a pulse is nonzero, it may be expected that the xy-component of the net nuclear magnetic moment will decay towards zero as the result of an approach to equilibrium. This is usually approximated by an exponential decay similar to T_1; in this case it is called transverse (spin-spin) relaxation and denoted T_2:

$$\frac{d(I_y(t))}{dt} = -\frac{t}{T_2} \tag{3.2}$$

The value of T_2 is however also affected by inhomogeneties in the static magnetic field, as nuclei from different parts of the sample (experiencing different fields) will precess at different rates, thus causing a loss of phase coherence in the xy plane (and a smaller T_2). Obviously, this instrument- dependent T_2 is not of much use except as an indication of instrument quality, so the experimental techniques used must eliminate the effects of inhomogeneities in the magnetic field. Usually, this is achieved by spin-locking the magnetisation along the x-axis, which prevents the magnetisation from spreading too far.

3.1.1 Measurement of T_1

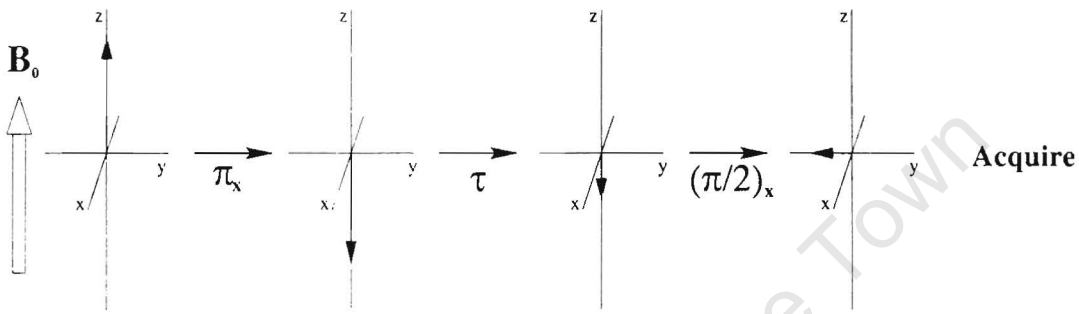


Figure 3.2: Inversion-recovery pulse sequence for measuring longitudinal relaxation

The most frequently used method for determining longitudinal relaxation is the inversion-recovery experiment.[83,84] In this pulse sequence, the sample is subjected to an initial π_x pulse, and allowed to relax for a period τ before the bulk magnetisation is rotated onto the x-axis by means of a second $(\frac{\pi}{2})_x$ pulse, where the free induction decay is acquired. The experiment is repeated for a set of times τ, a higher sampling density being used for the shorter times over which most of the relaxation in fact takes place. The T_1 values are then usually least-squares fitted to the data using an exponential function.

3.2 Physical Basis for NMR Relaxation

Since molecular motion does not directly affect the orientation of nuclear spins in an applied magnetic field, it is the local environment (or "lattice") which determines longitudinal relaxation. The magnetic energy of the nucleus is composed of a dominant part imposed by the external magnet onto which are superimposed the tiny local effects of neighbouring nuclei and the electronic environment. It is the fluctuation in the smaller local energy components that allows the nucleus to relax, effectively changing the dominant energy from the orientation of the spin in the external field. The two most

common mechanisms by which a nucleus can exchange energy with its environment in a macromolecule are:

1. dipole-dipole interactions with neighbouring nuclei or electrons

2. chemical shift anisotropy (CSA) effects in which the rotation of the molecule causes fluctuations in the energy of interaction with its electronic environment, for example due to ring currents in aromatic systems.

The CSA mechanism does not play a role in the relaxation of ^{13}C nuclei in carbohydrates, because the chemical shift does not have a strong dependence on the molecular orientation. There being no unpaired electrons the dipolar interaction is restricted to neighbouring nuclei, and of these, only the directly attached nucleus has a significant effect on the rate of relaxation. It is further assumed that the concentration of paramagnetic compounds such as dioxygen is sufficiently low that interaction with unpaired electron spins may also be neglected.

It can be shown that the dipolar contribution to each of the abovementioned relaxation parameters is related to the motion of the internuclear X-H vector, in the case that X has no significant dipolar interactions apart from with its attached proton H. If the X-H vector is denoted $\vec{\mu}_{LF}$, then the relaxation is described by the correlation function given in equation 3.3. The relaxation parameters are a function of the fourier transform, or *spectral density* $J(\omega)$ of this correlation function. The spectral density is given in equation 3.4.

$$C(t) = \frac{1}{5}\langle P_2(\vec{\mu}_{LF}(0)\cdot\mu_{LF}(t))\rangle \tag{3.3}$$

$$J(\omega) = 2\int_0^\infty C(t)\ cos\omega t\ dt \tag{3.4}$$

Having thus defined the spectral density, it is possible to express the various relaxation parameters as in equations 3.5 and 3.6. The parameter D, sometimes referred to as the "dipolar coupling constant" is given by $D = (\frac{\mu_0}{4\pi})\gamma_X\gamma_H\hbar\langle r_{XH}^{-3}\rangle$, with γ_X and γ_H being the respective magnetogyric ratios of the heteronucleus (carbon) and the attached proton, r_{XH} is the distance between them, μ_0 is the permeability constant and $\hbar$ is Planck's constant divided by 2π.

$$T_1^{-1} = \frac{1}{4}D^2[J(\omega_H-\omega_X)+3J(\omega_X)+6J(\omega_H+\omega_X)] \tag{3.5}$$

$$T_2^{-1} = \frac{1}{8}D^2[4J(0)+J(\omega_H-\omega_X)+3J(\omega_X)+6J(\omega_H)+6J(\omega_H+\omega_X)] \tag{3.6}$$

3.3 Calculation of Dipolar Relaxation Parameters from Simulation

In calculating relaxation parameters from a simulation, a practical difficulty arises if the correlation function $C(t)$ on which the spectral density is based has not decayed to zero in the time over which it can be reliably calculated (approximately $\frac{1}{3}$ of the total simulation time). This is usually due to the slow decay of the correlation function resulting from molecular tumbling. Since the internal motions in which we are interested are much more rapid, their contribution to the overall decay of $C(t)$ will decay to a steady state within the simulation period. Thus, a frequently adopted approximation is to factor the correlation function into contributions from internal motion and molecular tumbling, assuming these are statistically independent, yielding an expression of the form given in equation 3.7.

$$C(t) = C_O(t)C_I(t) \tag{3.7}$$

In equation 3.7, the overall correlation $C_O(t)$ is represented by an exponential decay with constant $\frac{1}{\tau_M}$, where τ_M is the rotational correlation time of the molecule. The internal correlation time $C_I(t)$ is calculated using equation 3.3, but with each trajectory frame least squares fitted to a reference coordinate set, to remove the effect of molecular tumbling. In effect, this is the correlation as observed from a reference frame rigidly attached to the molecule. In CHARMM, the correlation function is represented by the product of the internal and overall correlation functions up to the maximum time t_{max} to which the correlation function is calculated, and by the overall correlation function multiplied by a constant only from then on (equation 3.8).

$$C(t) = \begin{cases} C_I(t)\ e^{-\frac{t}{\tau_M}} & \text{for } t \leqslant t_{max} \\ C_I(t_{max})\ e^{-\frac{t}{\tau_M}} & \text{for } t > t_{max} \end{cases} \tag{3.8}$$

The rotational correlation time for a simulation may be evaluated by considering the correlation function in equation 3.3 for a vector rigidly attached to the molecular frame. The correlation time may then be estimated by a least squares fit of an exponential decay to the initial portion of the correlation function.

3.4 Model Free Formalism

The information directly available from the NMR relaxation parameters is the spectral density (or equivalently, the correlation function), via equations 3.5 and 3.6. Since it is not possible to extract the detailed (atomic scale) dynamics from the spectral density,

the maximum dynamical information that can be extracted from relaxation experiments is in fact that implied by the spectral density. In order to describe the actual dynamics, the spectral density must be interpreted in terms of a certain model of the motion. The models used initially included "restricted diffusion" of the X-H vector (for example, within a cone) and "jump models" in which the vector jumps between a certain number of sites. In these models, a spectral density function derived from the model would be fitted to the experimental data. The dynamics could then be analyzed in terms of the fitted model parameters. The danger of this approach is that more than one model may fit the data, and there is a risk of overinterpreting the data in terms of a particular model.

The *Model-Free formalism* of Lipari and Szabo[21,82] takes a purely mathematical approach to the data, to avoid introducing any artificial model. The correlation function is formally factored as in equation 3.7, using the correlation function for the molecular tumbling given in equation 3.9 below:

$$C_O(t) = \frac{1}{5}\, e^{-\frac{t}{\tau_M}} \tag{3.9}$$

The correlation function $C_I(t)$ for internal motions will now not usually decay to zero, since the motion within the molecular reference frame is restricted, and the internuclear C-H vector cannot take on all possible orientations. Thus $C_I(t)$ usually reaches a steady nonzero limit after a short time. Calling this steady state limit $\mathcal{S}^2$, the internal correlation function may be approximated by an exponential decay with correlation time τ_e, as in equation 3.10. The limit $\mathcal{S}^2$, known as the *generalised order parameter*, is an indication of the extent of spatial restriction of the motion: if the motion is unrestricted, it will be zero (the converse is not true), or equivalently if it is non-zero, the motion must be restricted in some way, dependent on its value. The time τ_e is a measure of the rate of the internal motions.

$$C_I(t) = \mathcal{S}^2 + (1 - \mathcal{S}^2)\, e^{-\frac{t}{\tau_e}} \tag{3.10}$$

Having made these approximations to the overall and internal correlation times, it can be shown[21] that the spectral density may be given by equation 3.11. This spectral density expression contains constants representing general assumptions about the form of the correlation functions but does not invoke any specific model. For this reason, the approach is known as model-free. Experimental results consisting of relaxation data at various frequencies are then least squares fitted to this model using equations 3.5 and 3.6, to obtain the parameter τ_M characterising the molecular motion and the parameters τ_e and $\mathcal{S}^2$ characterising the motion for each nucleus.

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_M}{1+(\tau_M\omega)^2} + \frac{(1-\mathcal{S}^2)\tau}{1+(\tau\omega)^2}\right) \quad (3.11)$$

$$\tau^{-1} = \tau_M^{-1} + \tau_e^{-1} \quad (3.12)$$

The application of these equations to experimental data requires some care to ensure that the best-fit parameters in fact reflect the "true" parameters that could be obtained if the exact time evolution of the C-H vectors were known, and are not artefacts arising from insensitivity in the fitting. This question has been considered by the authors of the model:[21] by applying the technique to data from simple and well-defined motional models, they proposed three categories for estimating the validity of results.

1. $\tau_e << \tau_M$: In this case, the fitted τ_e and $\mathcal{S}^2$ are accurate to within a few percent.
2. $\frac{\tau_e}{\tau_M} < 0.1$: $\mathcal{S}^2$, τ_e are accurate to within a few percent.
3. $\mathcal{S}^2 > 0.3$: $\mathcal{S}^2$ and τ_e are "fairly accurate", with no restriction on either τ_e or τ_M.

Usually an estimate of τ_M is made initially, either independently (preferred) from the hydrodynamic radius as determined from light scattering or PFG NMR experiments, or if such data is not available, from the actual relaxation data used in fitting the order parameters. The other two parameters are then least-squares fit to the relaxation data, using the globally determined τ_M.

3.5 Calculation of Generalised Order Parameters from Simulation

The generalised order parameters $\mathcal{S}^2$ may also be calculated directly from a molecular dynamics simulation. Using the addition theorem for spherical harmonics in conjunction with the correlation function in equation 3.3, it may be shown[21] that the order parameters are given by equation 3.13, where $Y_{lm}(\theta,\phi)$ are the spherical harmonics and θ_{mol} and ϕ_{mol} are the spherical polar angles for the C-H vector in the molecular coordinate frame.

$$\mathcal{S}^2 = \lim_{t\to\infty} C(t) = \frac{5}{4\pi}\sum_{-2}^{2} |\langle Y_{2m}(\theta_{mol},\phi_{mol})\rangle|^2 \quad (3.13)$$

In practice, the angles ϑ_{mol} and ϕ_{mol} would be calculated by least squares rotation of each trajectory frame onto a set of reference coordinates, and measuring the angles in the rotated set. This provides a "molecular" reference frame.

Chapter 4

Maltose ($\alpha(1\rightarrow4)$-linkage)

Carbohydrate monomers are essentially rigid, since only one of the chair conformations ($^4C_1$) is experimentally observed, so the flexibility of polysaccharides must be principally attributable to the interresidue glycosidic linkages. As the predominant interresidue join in most oligosaccharides and polysaccharides is an $\alpha(1\rightarrow4)$-glycoside, this has been the subject of many previous investigations.[24,57,86,87]

β-Maltose (α-D-glucopyranosyl-(1→4)-β-D-glucopyranose) (figure 4.1) is the simplest prototype of an $\alpha(1\rightarrow4)$-linkage, and in addition is one of the most abundant and best experimentally characterised disaccharides. It is one of comparatively few sugars for which a crystal structure has been obtained;[14] in addition, a neutron-diffraction refinement,[88] optical rotation[89] and nOe studies[87] are available for it. Being the obvious first step in the theoretical treatment of $\alpha(1\rightarrow4)$-linked polysaccharides, and a benchmark in the testing of new force fields has made maltose the subject of a large body of published work.[24,57,86,90] The main objective of the maltose simulations done here will not be an in-depth analysis of dynamical behaviour, as this has been done before, but rather to create a data set suitable for the calculation of NMR relaxation properties using the same protocol as for the later simulations, and to verify that the simulation results agree with previous work.

The atom naming convention which will be applied is simply the standard glucose atom numbering coupled with residue number, where the non-reducing residue is 1, and the reducing residue 2. Using this nomenclature, the interresidue dihedral angles, Φ and Ψ have been chosen as:

Φ :1H1-1C1-1O1-2C4
Ψ :1C1-1O1-2C4-2H4

This is in accordance with those used previously by Brady et. al.[24]

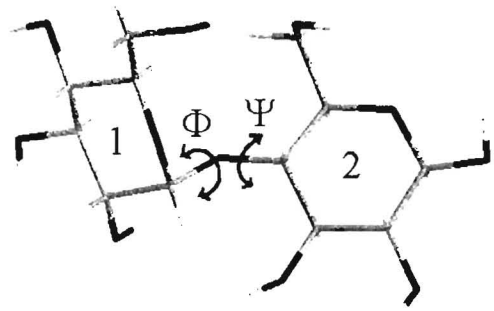


Figure 4.1: Residue numbering and torsion angle name conventions for β-Maltose.

4.1 Simulation Details

4.1.1 Adiabatic Map and Simulation Conditions

A preliminary indication of the conformational space available to maltose may be obtained from a vacuum adiabatic map. Such a map has previously been calculated for maltose using the present force field[90] and is shown in figure 4.2. The dominant feature is a broad low energy region in the centre of the map comprising a large well at $(\Phi, \Psi) = (-50, -40)$ and a smaller one at $(20, 20)$. The high energy regions in the map are primarily caused by steric clashes between the hydroxyls adjacent to the glycosidic linkage. Since the basic form of the adiabatic map is determined by this sterically congested environment, it generally has the same low energy region in the centre for all force fields,[17,22,23,25,78,79] the differences being in the details of this central area. The β-maltose crystal structure[13] with $\Phi = 0^o$ and $\Psi = -13^o$ lies within this low energy region. An interresidue hydrogen bond between O2 of residue 1 and O3 of residue 2 is observed in this structure. The wells in the adiabatic map would be expected to be most populated in a vacuum simulation and most likely also in a solution simulation. However, any dynamics simulation includes entropy effects, which could alter the favoured regions of Φ-Ψ space from that in the map (which is effectively calculated at 0 Kelvin).

The map minimum corresponding to the $(0, 0)$ conformation was used as the starting point for all simulations, since it forms the saddle point between the two main energy wells. A vacuum simulation was done for 1 ns in the microcanonical ensemble using the leapfrog verlet integrator. The temperature remained at 300K after equilibration. For

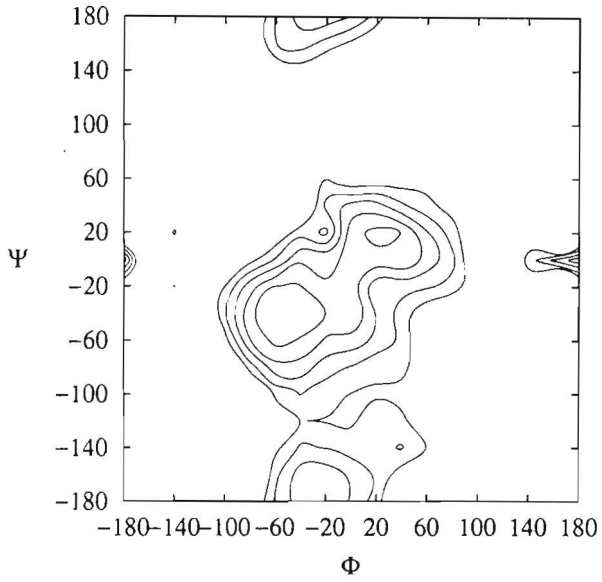


Figure 4.2: Adiabatic map for maltose calculated using the CHARMM force field:[25] contours at 2 kcal intervals

the solution simulation, the maltose solute was solvated using 488 explicit TIP3P water molecules in a cubic cell of side 24.6433Å, to give a density of $1.0037 g/cm^3$, and run for 1.5 ns with the temperature stable at 300K. Interresidue dihedrals were constrained to $(0, 0)$ during heating and equilibration.

4.2 Basic Trajectory Analysis

The vacuum simulation showed little variation in the important dihedral angles. The glycosidic (Φ, Ψ) pair (figure 4.3) immediately rotated to the larger and lower energy well at $(-50, -40)$ in the adiabatic map. There appear to be two discrete conformations within this well, one at about $(-50, -40)$ and the other at approximately $(-15, -15)$. A similar behaviour was seen in the earlier HGFB force field, where this large well was clearly split into two smaller wells, designated A and B in the vacuum potential for that force field. Following this nomenclature, the population at $(-50, -40)$ will be named *A*, that at $(-15, -15)$ *B* and the subsidiary minimum at $(20, 20)$ will be referred to as *C*. The vacuum simulation stays mainly in the main A well, making only a few brief transitions to the B conformer and to the smaller well at C. The primary alcohols (figure 4.3) made no transitions from the *gg* rotamer; however this "stiffness" is inherent in the PHLB force field. The restriction of the dynamics to regions of the adiabatic map close to the minima, and the infrequent transitions observed are typical of vacuum runs. This can be attributed to the need to completely break a number of hydrogen bonds (those forming a continuous ring or "crown" between adjacent hydroxyls on each ring,

for example) in a concerted fashion in order to cause a change in conformation, since the favourable orientation of each hydroxyl depends on the orientations of its neighbours to maximize hydrogen bonding

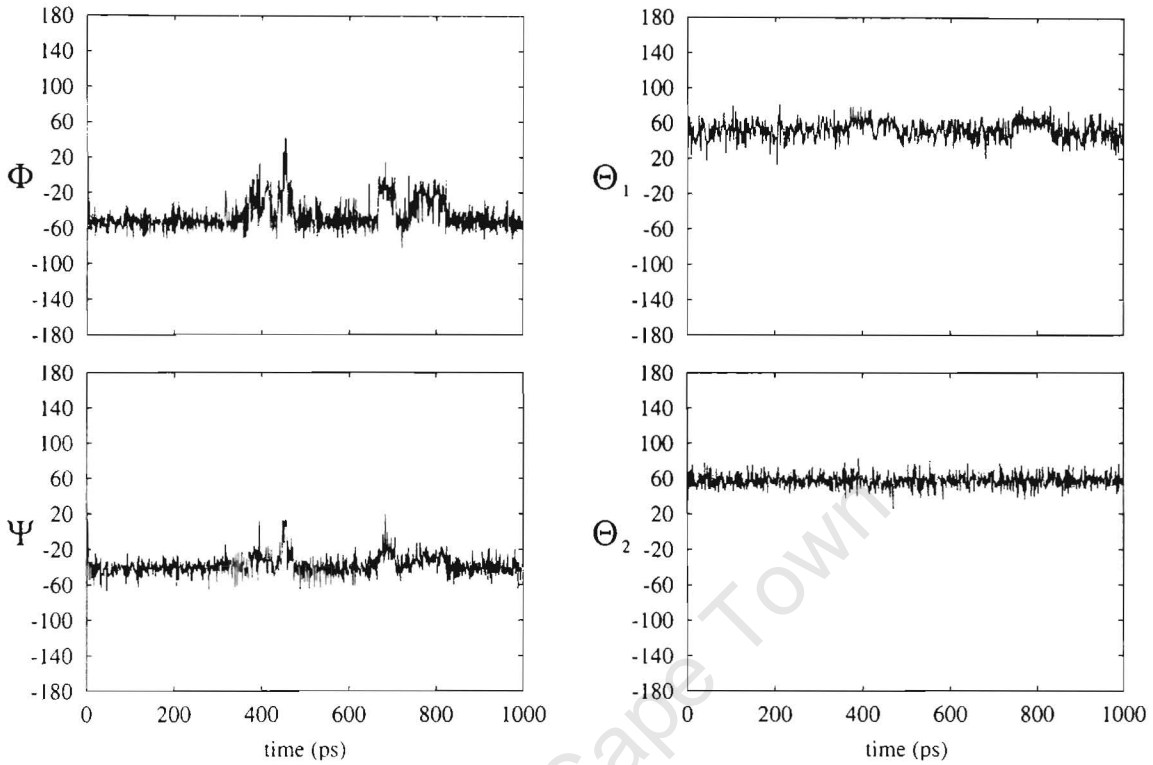


Figure 4.3: Time series for glycosidic and primary alcohol dihedral angles for the maltose vacuum simulation

The inclusion of water affected the behaviour of maltose substantially (time series in figure 4.4). Following an initial transition from $(0, 0)$ to the C well the saccharide sampled the larger well as in the vacuum case. However, the frequency of subsequent transitions in Φ and Ψ between the A and B conformers is much higher than for the vacuum run. This behaviour exemplifies the influence of the solvent: by providing alternative hydrogen bonding partners for the alcohols, it allows them to move more independently of each other, reducing the barrier to conformational transitions. This effect can be seen in the large range of Φ-Ψ space explored: Φ varies between -70^o and 0^o in solution, but for most of the vacuum simulation it fluctuates between -70^o and -30^o. A similar phenomenon is observed for Ψ, which samples between -60^o and 0^o in solution, but mainly between -60^o and -20^o in vacuum. Although this effect did not cause the small C well to be sampled, it can certainly be seen in the increased region of the larger well explored.

The primary alcohols did not make any transitions from the *gg* conformer in solution.

To get a visual impression of the conformational regions sampled, the trajectories

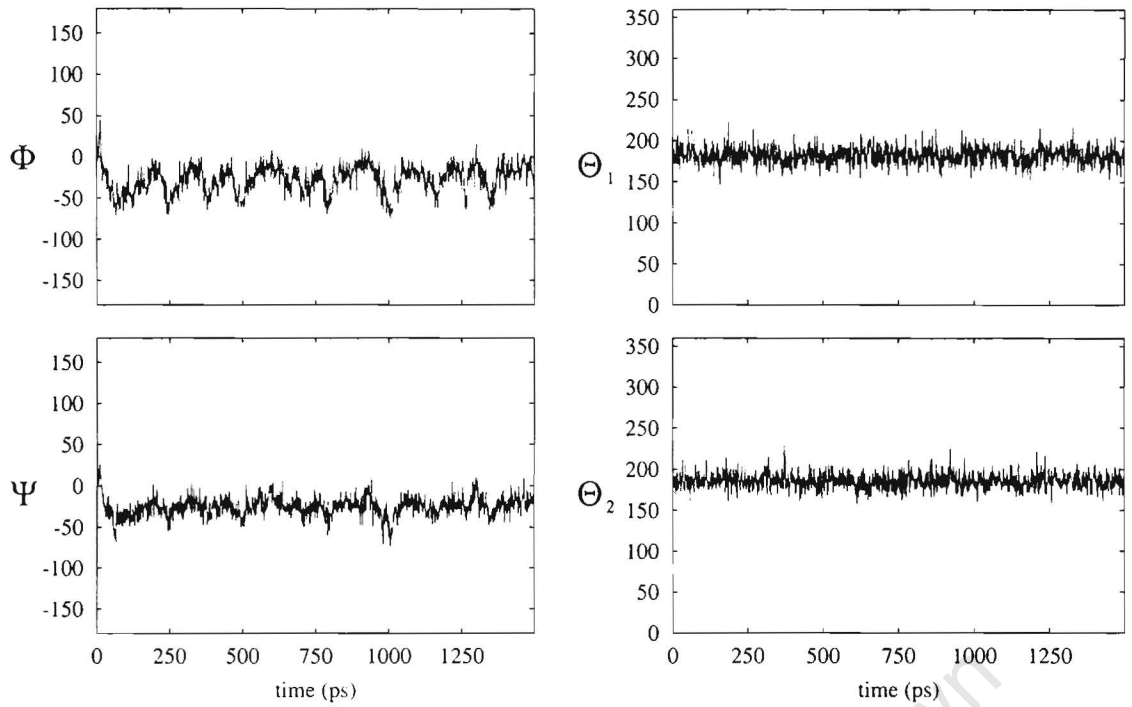


Figure 4.4: Time series for glycosidic and primary alcohol dihedral angles for the maltose solution simulation

of the glycosidic dihedrals have been projected onto the adiabatic map shown earlier, in figure 4.5 . The vacuum run can be seen to closely trace the low energy regions in the adiabatic map, which was after all calculated in vacuum. However, the dynamics does include some entropic effects, which the map itself does not. Although the two subminima in the large well are not very distinct, there is nonetheless evidence that both ends of the well are being sampled. The solution run, while at first sight similar, has some important differences. Firstly, while the general pattern of the vacuum well is followed, there is a clear shift of the population toward the centre of the map. It also appears that, of the two subminima, B is possibly favoured over A, and certainly more so than in the vacuum case. The reasons for this shift will require closer analysis, but the overall effect of the solvent may be described as an increased exploration of conformational space, as well as shifting the favoured well populations from those sampled in vacuum.

4.3 Cluster Analysis

The vacuum and solution runs each showed the existence of two distinct populations within the larger well at approximately $(-15, -15)$ and $(-50, -40)$. However, it is somewhat arbitrary to estimate the position and extent of these two clusters by inpection

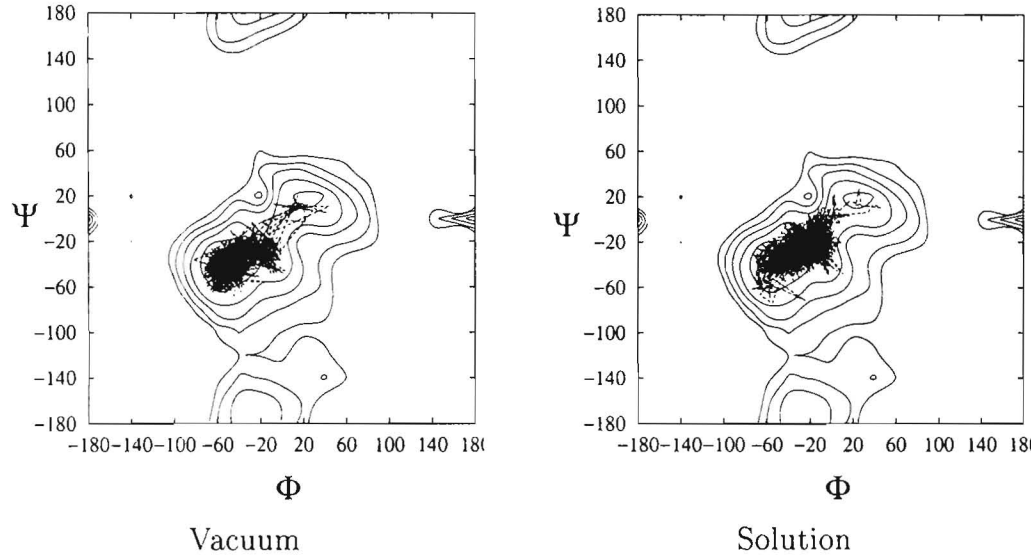


Figure 4.5: Dynamics trajectories for vacuum and solution simulations of maltose superimposed on the adiabatic (Φ, Ψ) plot.

of time series alone. Using cluster analysis, both the position and size of each cluster may be analyzed by a more rigorous algorithm.

The largest clusters arising from cluster analyses of the solution and vacuum simulations are shown in table 4.1. The clustering algorithm for the vacuum run converged quickly and gave two well-defined clusters corresponding to the two abovementioned wells. The central positions of the clusters are slightly shifted from what one would expect from looking at the time series only, especially the B well, which is now at $(-19.4, -25.9)$.

Simulation	cluster	# members	std. dev.	Φ	Ψ	Θ_1	Θ_2
Vacuum	1	15499	0.868e+1	-52.6	-40.9	50.7	58.0
	2	4051	0.114e+2	-19.4	-25.9	57.9	56.7
Solution	1	13220	0.117E+2	-26.9	-25.7	-169.8	-170.9
	2	5382	0.871E+1	-16.9	-20.2	-169.8	172.6
	3	4478	0.119E+2	-26.3	-25.4	174.3	-170.7
	4	3769	0.819E+1	-45.2	-32.1	-169.6	172.8

Table 4.1: Cluster analysis for vacuum and solution simulations of maltose.

The structures saved from the solution run did not cluster so easily. The maximum error in the algorithm was exceeded several times, due to the close overlap of the clusters. The overlap also manifested itself in the final clustering, since the cumulative number of structures in each cluster was substantially larger than the actual number assigned to

the cluster. However, although most of the clusters correspond to the B well population (the main differences being in the primary alcohols), the fourth cluster belongs to the A well conformation. The value of the clustering algorithm in locating the centre of each population was again evident, as the actual positions of the clusters are not exactly where they were expected.

The clustering analysis addresses the issue of the relative minimum well populations. The visual cue from the contour projections that the vacuum simulation favoured the A well appears to be correct; the largest cluster in vacuum indeed corresponds to this combination of dihedrals. For the solution run, the top three clusters all match the B well position on the adiabatic map.

The finding that there exist distinct populations in vacuum and in solution invites an explanation of the cause of this phenomenon. The structures which match the cluster centre dihedrals most closely are shown in figure 4.6, along with the hydrogen bonds present.

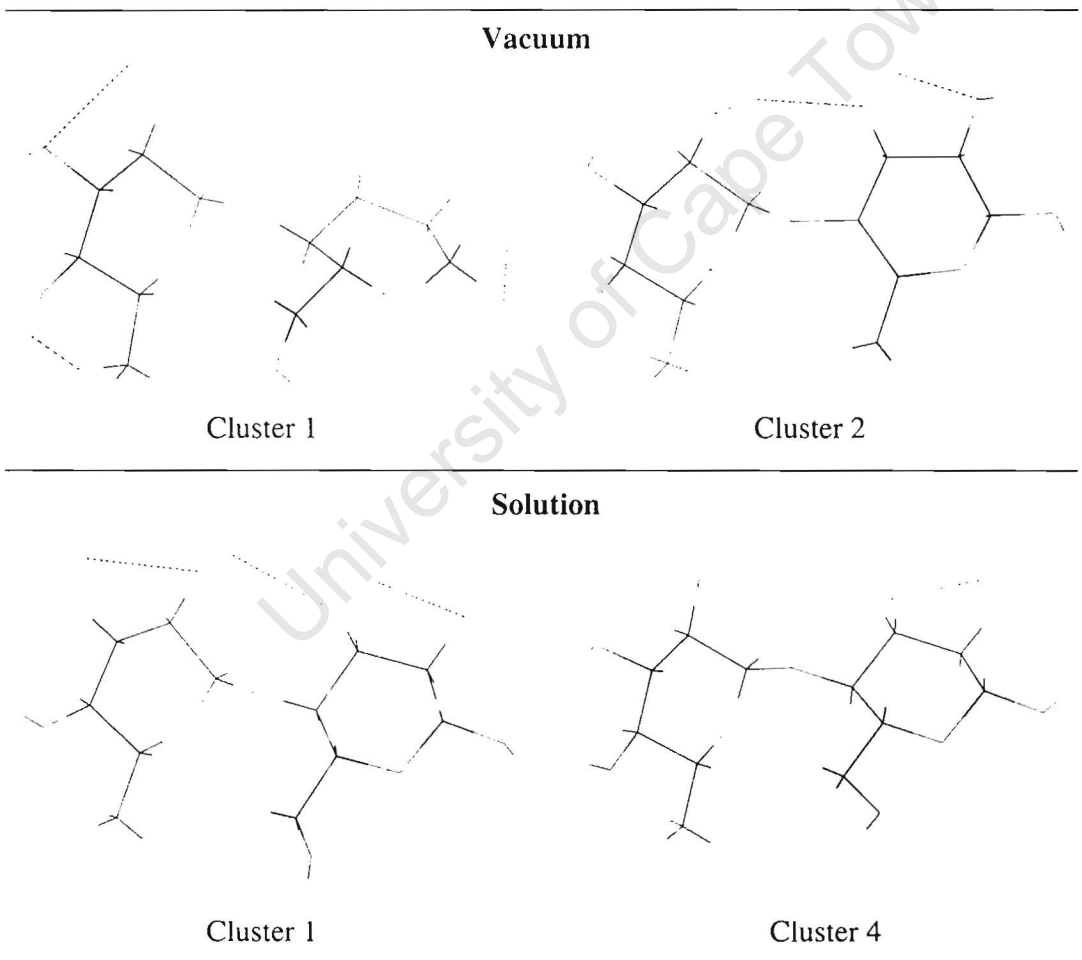


Figure 4.6: Maltose structures lying closest to each cluster centre

The distinction between the clusters is clearly the formation of the hydrogen bond

between O2 on residue 1 and O3 on residue 2. The clusters corresponding to the lower A well (cluster 1 in vacuum and cluster 4 in solution) have no interresidue hydrogen bonds, while those in the B well (cluster 2 in vacuum and cluster 1 in solution) possess the abovementioned hydrogen bond between the residues. This agrees with previous published findings using the HGFB force field,[56] in which the B and C wells were found to have this hydrogen bond, while the A well did not. The balance between the A and B wells can be seen as a trade-off between the extra steric strain induced by the B well and the hydrogen bond which forms in it. However, the favoured populations established by the short simulations here are not very conclusive, especially in view of the fact that only one starting conformation was used. A previous study in fact found a reversal of the populations in some simulations.[53] Nonetheless, the fact that there are distinct conformations, resulting from the presence or absence of an interresidue hydrogen bond is conclusively established.

4.4 Hydrogen Bond Analysis

The interaction of carbohydrate solutes with water is frequently described in terms of the hydrogen bonds formed between the carbohydrate hydroxyls and the water molecules. The directed nature of these bonds is believed to be responsible for the strong solvation effects observed in water.

Ideally, the hydrogen bonding would be separately analysed for vacuum and water; however due to the frequent nature of the transitions in solution and their short duration in vacuum, it is not possible to find a long enough section of the simulation to analyse hydrogen bond lifetimes for each conformer. Instead, the entire solution and vacuum runs have been analysed in their entirety. The average number of hydrogen bonds and their average lifetimes in picoseconds have been listed in table 4.2 for each sugar oxygen. For the vacuum runs, only internal hydrogen bonds could be studied, while for the solution runs, hydrogen bonds to water and hydrogen bonded water bridges could also be extracted.

Solvation reduces the number of intramolecular hydrogen bonds for most of the sugar oxygens, presumably due to competing interactions with water. This is evident in figure 4.6 in which the hydroxyl groups in solution are seen to be more or less randomly oriented, while those in vacuum are constrained to a tight crown running in the same direction around the sugar ring. A notable exception to this trend are the hydroxyl oxygens involved in the interresidue O2-O3 hydrogen bond. These form a larger average number of hydrogen bonds in solution than in vacuum. The persistence of this hydrogen bond in solution has been noted in previous studies[56] with other force fields. This results from this hydrogen bond being more favoured in the solution simulation than in the vacuum one. The pyranosyl oxygens and the glycosidic oxygen have a greatly reduced hydrogen bonding potential, compared to the more polarised hydroxyl groups.

In terms of the total number of hydrogen bonds formed with water, most of the hydroxyl groups form between 1.5 and 2.4 hydrogen bonds each, with the obvious exceptions being those involved in the interresidue hydrogen bond. No significant water bridges are observed, the average number formed being about 0.15, which can be attributed to bridges between neighbouring hydroxyls.

The nature of the interresidue hydrogen bond may be studied in more detail using the time series for the O2 to O3 distance and the relevant acceptor- donor-acceptor angles, which are plotted in figure 4.7. The vacuum hydrogen bonds are evidently longer lasting than those in solution and undergo sharper transitions. Presumably, this is because of the high energy barrier to transitions in vacuum. In solution, the barrier could be lowered by interaction with water. This effect manifests itself in both the frequency and gradual change observed for the breaking and formation of this hydrogen bond in

	Solution			Vacuum
Hydroxyl	To Water	Bridging	Internal	Internal
1 O1	0.05 (0.10)	0.01 (0.06)	0.07 (0.07)	0.02 (0.07)
1 OH2	0.71 (0.35)	0.11 (0.13)	0.78 (0.22)	0.66 (0.09)
1 OH3	1.38 (0.48)	0.21 (0.15)	0.34 (0.07)	0.87 (0.07)
1 OH4	1.76 (0.63)	0.12 (0.13)	0.13 (0.06)	0.41 (0.06)
1 O5	0.17 (0.13)	0.07 (0.08)	0.07 (0.08)	0.10 (0.06)
1 OH6	2.10 (0.84)	0.15 (0.12)	0.06 (0.04)	0.16 (0.06)
2 OH1	1.60 (0.86)	0.27 (0.13)	0.09 (0.07)	0.33 (0.06)
2 OH2	0.94 (0.22)	0.13 (0.13)	0.21 (0.07)	0.80 (0.07)
2 OH3	0.46 (0.28)	0.06 (0.11)	0.72 (0.20)	0.60 (0.15)
2 O5	0.49 (0.17)	0.26 (0.09)	0.03 (0.08)	0.12 (0.05)
2 OH6	2.18 (0.83)	0.16 (0.14)	0.03 (0.04)	0.13 (0.03)

Table 4.2: Hydrogen Bonding Statistics for vacuum and solution runs

solution. Another observation of interest is that in vacuum, each hydroxyl acts as a donor and acceptor (at different times), while in solution HO3 on residue 2 is always the donor. This may be an artefact of the force field used.

4.5 Water structuring about Maltose

The hydrogen bonded interactions between water and the sugar solute cause the surrounding water to adopt a higher degree of short range structure than would be the case in the bulk solvent. Radially averaged pair distribution functions between each sugar oxygen and the water oxygens, $g(O_w)$, provide approximate pictures of how the water structures about each oxygen site. Such PDF's are illustrated for two maltose oxygens in figure 4.8, together with the integrated number of water molecules for each. The O3 oxygen is typical of most in having a strong first peak at 2.82Å, which on integration over a sphere of 3.5Å about the oxygen, results in a first shell of around 3.2 waters for each hydroxyl. This is much larger than the number of hydrogen bonded interactions calculated above, so there is clearly significant local water density which is not participating in hydrogen bonds to the hydroxyls. Molecules in bulk water show the same trend, in having around 4.4 nearest neighbours under the first peak of the PDF (depending on the water model used), while forming on average just over 3 hydrogen bonds at one time. The O5 oxygen has a more distant first peak in its PDF, corresponding to a weaker interaction with the water; the total integrated number of waters up to 3.5Å is

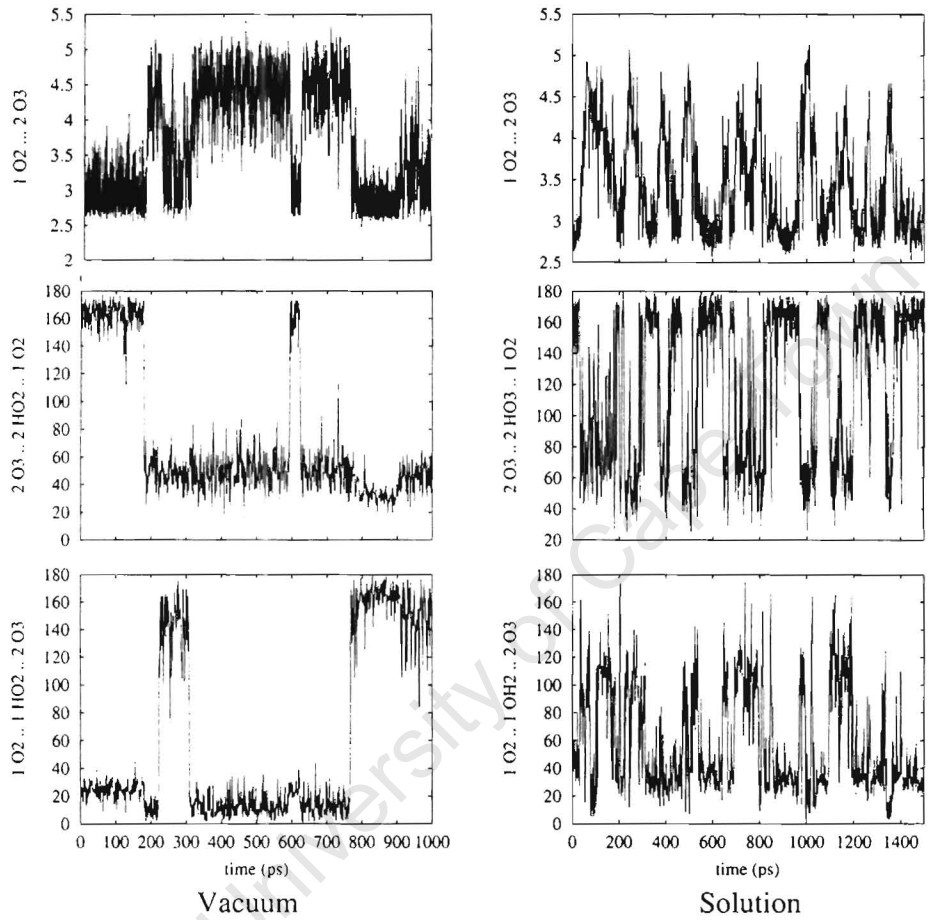
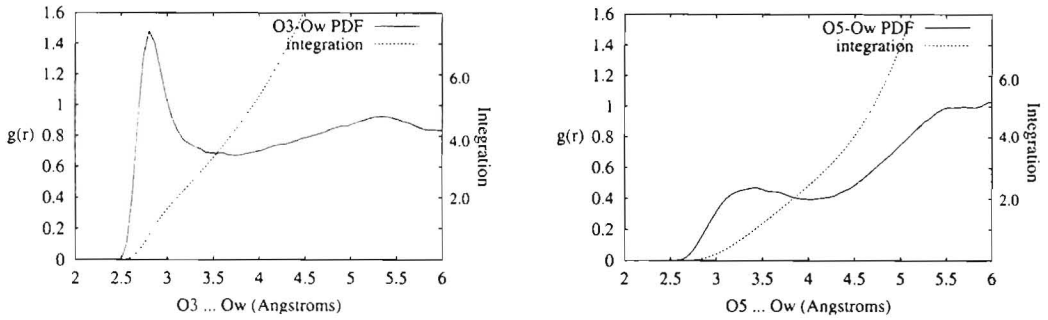


Figure 4.7: Variation of geometric hydrogen bond parameters for the 1 O2 - 2 O3 hydrogen bond in vacuum and solution

only 1.15.



1 O3 $g(O_w)$ 1 O5 $g(O_w)$

Figure 4.8: One dimensional PDF's between selected maltose oxygens and water oxygen.

The radial averaging of the one-dimensional pair-distribution function may be removed by calculating the full three-dimensional distribution of water around the solute, sometimes known as the "spatial distribution function", or SDF. Clusters 1 and 4 obtained for maltose in solution were analysed separately, firstly to avoid blurring of the oxygen density and secondly to account for the fact that they have a different number of hydroxyl groups available for hydrogen bonding with the solvent. Clusters 2 and 3 were very similar to 1, differing mainly in primary alcohol orientation, and so were not considered.

The three dimensional solvent distributions for clusters 1 and 4 are shown in figures 4.9 and 4.10 respectively, together with a plot showing the location of the maxima of solvent density about the solute. Peaks of water density in these plots follow a general pattern of being located between the hydroxyl groups and close to the plane of each ring, with the exception of those structured about the primary alcohol. Most of the water density peaks in the plots are within hydrogen bonding distance of the hydroxyls, although they cannot be identified as such because of the absence of water hydrogens. The distribution of solvent density about each maximum is fairly spatially restricted, despite the flexibility about the Φ and Ψ dihedrals observed earlier. An explanation could be that because of its small size, the rotations about the glycosidic dihedrals do not cause a large overall change in the molecular conformation.

Remarkably, the two clusters, which differ in the presence of the O2 to O3 hydrogen bond in cluster 1, do not structure the water very differently. Aside from a slight increase in water density near O2 of residue 1, they are almost exactly the same. This evidence suggests that the transitions between the A well (cluster 4) and the B well (cluster 1) do not require a substantial alteration of the water structure and that the difference between the two populations in solution is primarily intramolecular in nature.

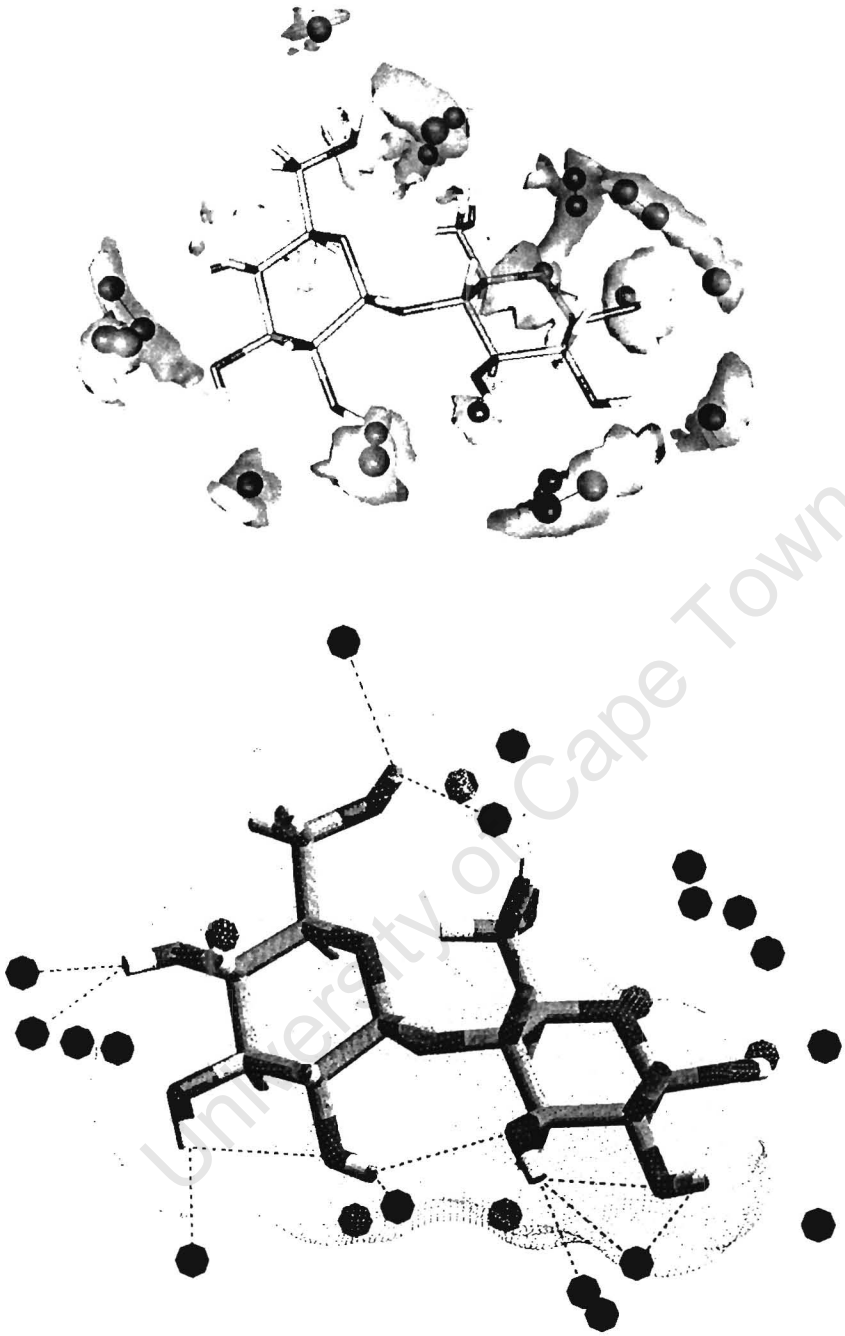


Figure 4.9: Three dimensional contour plot of water density around maltose analysed using cluster 1, with $\Phi \approx -26.9$ and $\Psi \approx -25.7$) (above); Points of maximum water density around the cluster centre structure with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

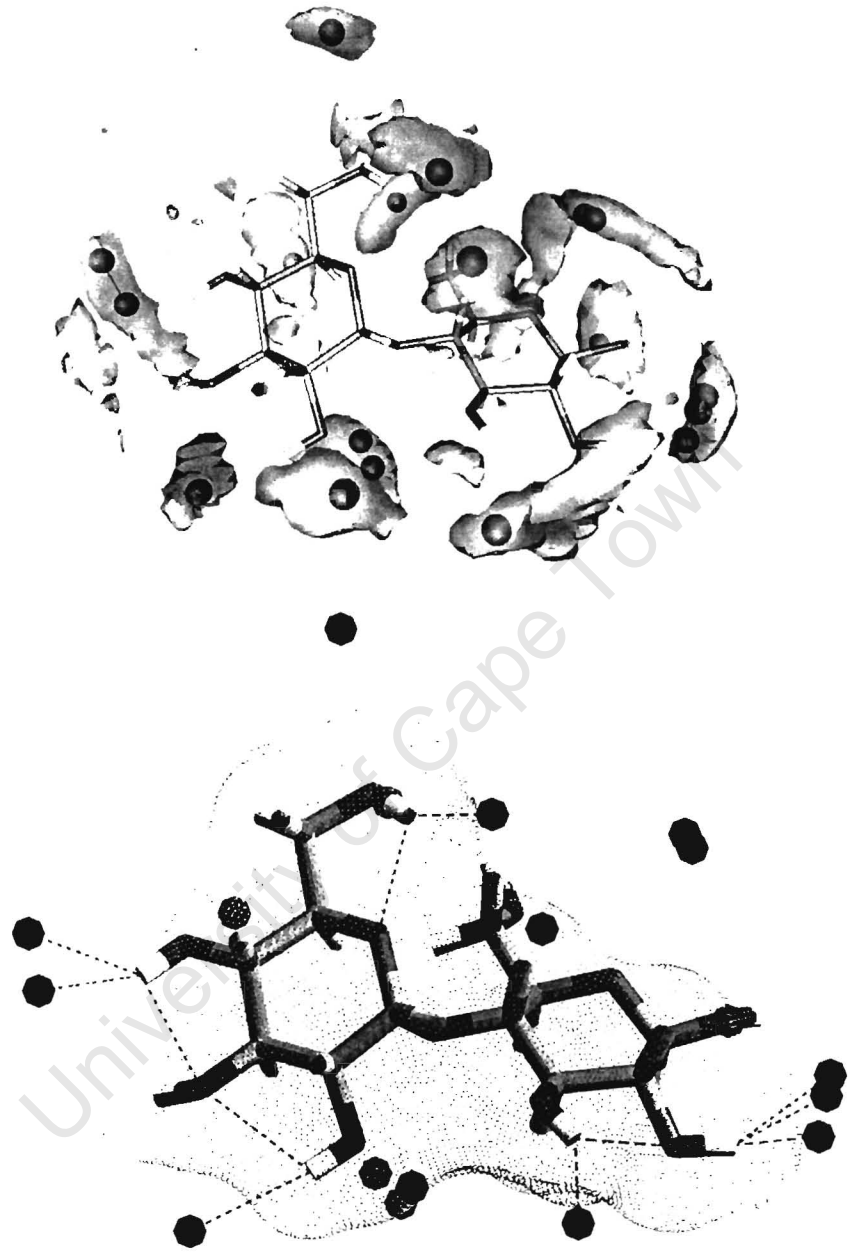


Figure 4.10: Three dimensional contour plot of water density around maltose analysed using cluster 4, with $\Phi \approx -45.2$ and $\Psi \approx -32.1$) (above); Points of maximum water density around the cluster centre structure with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

4.6 Internal relaxation behaviour

As a model for larger carbohydrate polymers, we are not only interested in the regions of conformational space explored by maltose, but also in its flexibility, or the rate at which the molecule moves. This can be numerically assessed through the rate of decay of the correlation function of the various molecular degrees of freedom, such as the dihedral angles. The rate of loss of correlation gives an indication of the rate at which the parameter under consideration loses the "memory" of its value at a previous time.

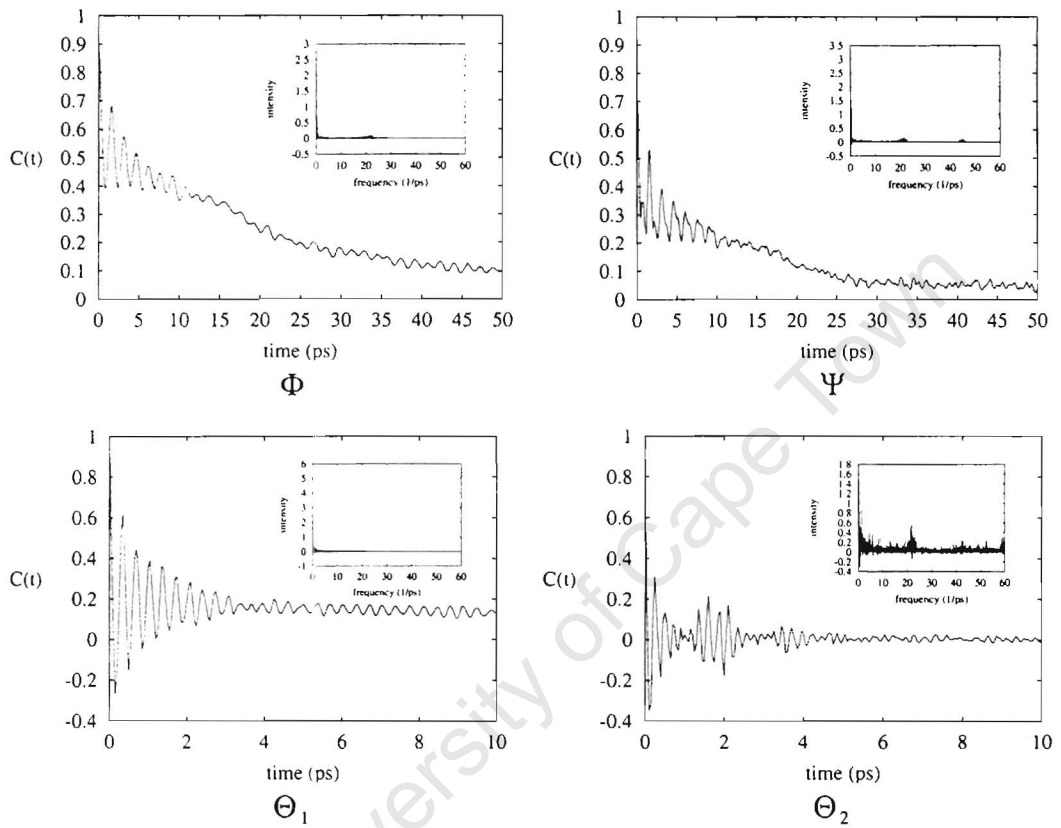


Figure 4.11: Time correlation functions of the fluctuations in dihedral angles for maltose in vacuum, and the corresponding power spectra (insets)

The correlation functions for the dihedral angle fluctuations of maltose in vacuum are shown in figure 4.11. It is not possible to look at the correlations of the dihedral angles themselves, as many of these would not decay to zero, so their fluctuations from the mean have been used instead. The Φ correlation function decays to zero after about 50 ps, and Ψ after about 35 ps, while the primary alcohols take about 5 ps to lose all correlation. The difference in decay times for the two is simply a result of the weight of the attached groups: the primary alcohol, having a lighter group attached, is subject to much more rapid fluctuations. The power spectra of the correlation functions, inset in each plot in figure 4.11 show that the glycosidic dihedral correlation functions comprise

only a few frequencies, mainly at the lower end of the frequency range, while the more rapidly moving primary alcohols have many more peaks, covering the entire frequency spectrum.

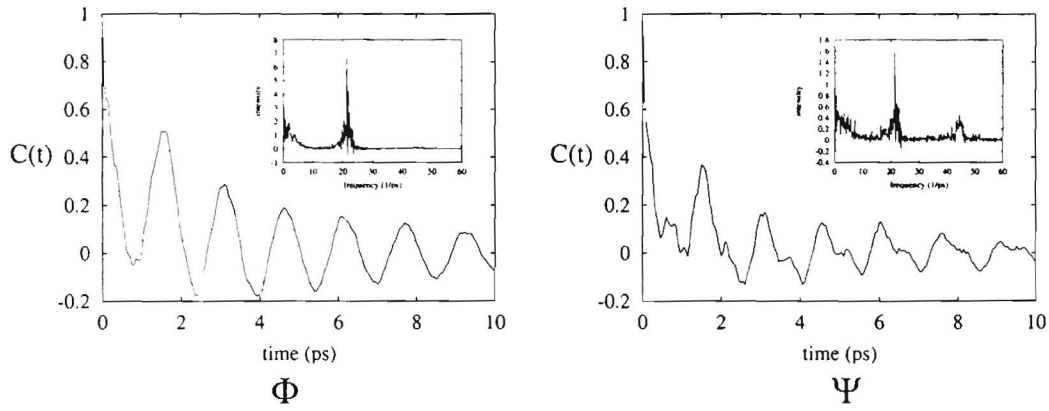


Figure 4.12: Time correlation functions in vacuum using only the first 300ps of data, for which there are no transitions.

After the differences in the overall decay time for the maltose dihedrals in vacuum, the second important feature of the dihedral autocorrelation functions is the higher frequency component superimposed on the general decay. These extremely rapid motions have been observed before by Brady and Schmidt,[56] caused by undamped motion in vacuum. An important distinction between that study and the present one, is that the overall correlation time for the glycosidic dihedrals in the Brady simulation was of the same order as that for the primary alcohols in the present one. This can be rationalized, since the simulation here makes a few, brief interwell transitions, while theirs remains in the well in which it was started, probably due to the short length of their simulation (80 ps). Thus the correlation functions for Φ and Ψ in figure 4.11 are the convolution of a rapid motion due to fluctuations within each well, and a slower decay representing interwell transitions. The effect of the transitions on the correlation function can be eliminated by only using the first part (0 to 300 ps) of the vacuum trajectory, since this segment contains no transitions. The result of this is shown in figure 4.12, which resembles the previously published data. Note that the higher frequency peaks at around $20ps^{-1}$ are now the dominant ones in the power spectrum.

Correlation functions for the maltose dihedral fluctuations in solution are shown in figure 4.13. These have similar overall decay times to the analogous solution cases, although slightly longer in the case of the glycosidic dihedrals. The most striking feature in all cases is that the high frequency fluctuations present in the vacuum runs have all but disappeared. This also manifests itself in the spectral densities for Φ and Ψ, where there are only a few very low frequency peaks and no significant ones above $5ps^{-1}$. The

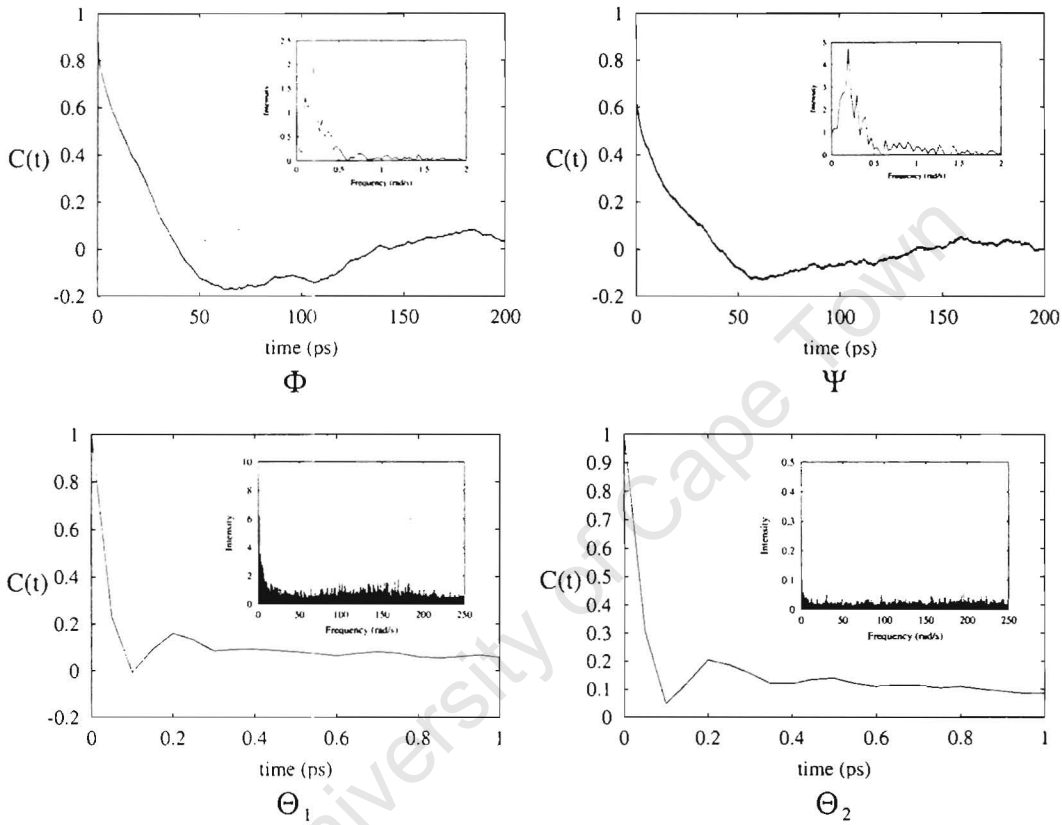


Figure 4.13: Time correlation functions of the fluctuations in dihedral angles for maltose in solution, and the corresponding power spectra (insets)

effect of the solvent is to damp the motions of the solute about the glycosidic linkage, such that only the low frequency components attributable to conformational transitions are observed. Note that the rate of transitions still appears to be greater in solution, and the solution simulations encompass more conformational space, but the small, rapid fluctuations within each well have been damped out in the presence of solvent.

The decay in correlation of dihedral angle fluctuations informs us of the flexibility of the individual dihedrals. To get a picture of the overall molecular relaxation requires a simultaneous consideration of all the molecular degrees of freedom. The molecular relaxation will be described in terms of the RMS fit autocorrelation function and the P_2 autocorrelation function (§2.3.2) for an intramolecular vector (from 1 C1 to 2 C4 in this case). These functions are shown for vacuum and solution in figure 4.14.

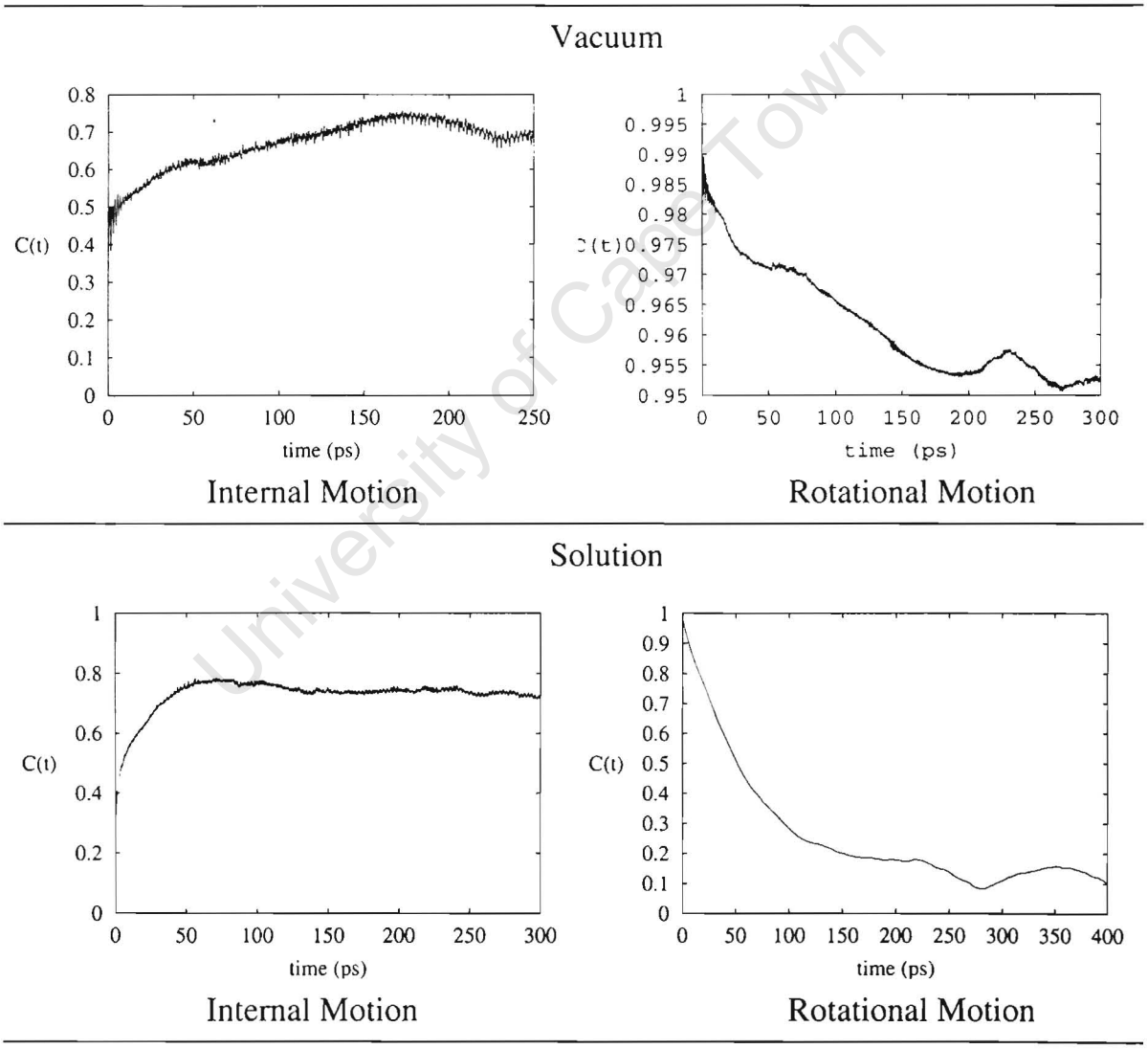


Figure 4.14: RMS fit autocorrelation functions (internal motion) and P_2 rotational autocorrelation functions for maltose in vacuum and solution.

The effect of water on the internal RMS fit autocorrelation functions is to slightly slow their initial decay and to damp out the high frequency fluctuations. RMS autocorrelation decay in solution is complete after 50 ps, which is indicative of a fairly rigid molecule. In the vacuum case the decay is extremely rapid over the first few picoseconds, but then slows as a result of the low frequency of conformational transitions in vacuum.

Rotational motion in vacuum, as represented by the P_2 autocorrelation function, is obviously much slower than in solution, due to the Brownian effects of the water molecules. More significant is the decay time of about 150 ps for the rotational motion in solution. The rotational correlation time, given by the value of the correlation function at $\frac{1}{e}$ is approximately 76 ps. This time is of relevance when considering NMR relaxation.

4.7 Calculation of NMR Properties and Experimental Comparison

Molecular Dynamics simulation data provide a level of atomistic detail which allows many experimental measurables to be accurately calculated. This provides a mechanism for checking the simulation and force field for correct behaviour, which is especially important if the force field will be used for future theoretical predictions. This can only be realistically done for solution simulations, as the internal motion is far too rapid in vacuum, and rotational motion is constant and isotropic rather than diffusional in nature.

T_1 relaxation times may be directly calculated from the maltose solution simulation using the method outlined in section 3.3, provided that a rotational tumbling time, τ_M is available for the molecule. This was estimated from the decay of the P_2 autocorrelation function for the vector across the glycosidic linkage (1 C1 to 2 C4) shown in figure 4.14. A least squares fit of an exponential decay to the first 100 ps of this function gave a rotational autocorrelation time of approximately 76 ps. Using this in the calculation of ^{13}C T_1 relaxation at the frequency at 250 MHz (the frequency at which the published data were collected) gave the values listed in table 4.3. Shown for comparison are the published experimentally determined values for maltose at this frequency.[91] Note that only the values for the β anomer are given, since that was the anomer simulated.

The calculated T_1 values are generally slightly higher than the experimental ones, possibly due to the tumbling being slightly slower than it should. This may be because the estimation of the tumbling time is limited in accuracy due to the small number of actual rotations observed. However, the general trend in the values is the same. All of the carbons in a particular ring tend to have similar T_1 values, reflecting the fact that the ring is quite rigid, and so the motions of each of its members are similar. The T_1's

for the second ring are generally higher than those for the first, both experimentally and as calculated from simulation. The cause of this is not clear; it does not seem to arise as a result of a more restricted motion in the second ring, as the generalised order parameters will show. However, it does show that the position of the rings relative to linkage has an important effect on their dynamics. An increased T_1 was observed by Benesi and Brant[91] for the reducing residues of a series of $\alpha(1\rightarrow4)$ linked oligosaccharides (but not for other linkages). The authors proposed this to be due to the anomeric exchange in this residue. However, our simulation model, which does not include anomeric equilibrium (the terminal residue being fixed in the β anomer) is able to reproduce this trend satisfactorily. The T_1 values for the C6's are much lower than the others as a simple consequence of these carbons having two attached protons. A final point is that the carbons adjacent to the linkage (1 C1 and 2 C4) have the largest relaxation times, since they are the most motionally restricted, being close to the centre of the molecule. Although this is not reflected in the corresponding set of experimental results, this trend is observed in T_1 experiments on other, related glucans.[91]

Residue	Carbon	T_1 (sim)	T_1 (exp)
1	C1	1.01	0.57
	C2	0.60	0.58
	C3	0.68	0.56
	C4	0.69	0.53
	C5	0.65	0.56
	C6	0.37	0.33
2	C1	0.78	0.71
	C2	0.78	0.68
	C3	0.80	0.69
	C4	0.84	0.63
	C5	0.70	0.68
	C6	0.37	0.30

Table 4.3: Experimental and simulated longitudinal relaxation times at 250 MHz for maltose. The experimental relaxation data are taken from Benesi and Brant.[91]

A popular approach to comparing experimental and simulated relaxation data is through the use of generalised order parameters, since their calculation from the simulation is not complicated by the need to accurately compute correlation functions. The information contained in the order parameter relates purely to the restriction of motion of the C-H vector, and does not include the rate of internal motion, or molecular tum-

bling, as does the T_1 relaxation time; this makes it simpler to interpret. However, the fitting of order parameters to experimental data for small oligosaccharides is difficult due to the similar magnitude of τ_e and τ_M. Previous studies have assumed internal motion to be much more rapid than the overall tumbling time,[92] effectively neglecting the second term of the model-free formalism; however, this has more recently been shown from molecular dynamics studies to be a poor approximation,[27] since the values for τ_e and τ_M were shown from the simulations to be similar in magnitude, as is found to be the case for maltose. An alternative approximation, setting τ_e to zero[93,94] is similarly dubious.

Order parameters have been calculated from the maltose simulation using the method described in section 3.5 and from experiment using the data of Benesi and Brant.[91] The estimation of parameters from experimental data was done using the program *MODELFREE*.[95,96] Data fitting was initially attempted with simultaneous optimization of all parameters, except for τ_M, which was estimated from simulation. However, for all the carbons, the fits converged to physically unrealistic values. By setting τ_e to zero and τ_M to 0.1 ns, it was possible to get reasonable values for the order parameters, which were comparable to those from simulation. However, the assumptions used in estimating the experimental parameters this way would suggest that caution should be used in their interpretation. Table 4.4 shows the results of these calculations.

The values of the order parameters are all quite close to 1, suggesting that the range of motion is limited for the maltose molecule, which was shown earlier in the simulation analysis. The same relation between the order parameters of the two rings is observed as for the T_1 values, namely that the order parameters within a ring are quite similar, although the overall difference between the rings is small. The order parameters for the C6 residues are slightly reduced relative to the rest, as the C6-H vectors move independently of the main ring.

Residue	Carbon	S^2 (sim)	S^2 (exp)
1	C1	0.88	0.89
	C2	0.93	0.88
	C3	0.92	0.92
	C4	0.92	0.99
	C5	0.92	0.92
	C6	0.83	0.77
2	C1	0.84	0.73
	C2	0.84	0.75
	C3	0.84	0.74
	C4	0.86	0.82
	C5	0.85	0.75
	C6	0.80	0.86

Table 4.4: Experimental and simulated generalized order parameters for maltose. The experimental relaxation data are taken from Benesi and Brant[91] and the model-free parameters fitted assuming $\tau_e = 0$;[93] no errors could be calculated, since none were reported with the data. Order parameters for the C6 carbons were calculated using $2T_1$, assuming each proton contributes equally to relaxation.

Chapter 5

Isomaltose ($\alpha(1\rightarrow6)$-linkage)

Due to the apparently simpler nature of the $\alpha(1\rightarrow4)$-linkage and of $\alpha(1\rightarrow4)$-linked carbohydrate polymers such as amylose, initial theoretical interest has centered on understanding these, rather than on $\alpha(1\rightarrow6)$linked saccharides.[24,57,86,90] However, since amylopectin, which contains both $\alpha(1\rightarrow4)$- and $\alpha(1\rightarrow6)$-links, is the major constituent of starch and is chiefly responsible for branching in polysaccharides, it is necessary to map the energy surface and dynamics of the $\alpha(1\rightarrow6)$glycoside. In contrast to the $\alpha(1\rightarrow4)$case, in which an abundance of experimental data including several crystal structures exists, the first crystal structure for an $\alpha(1\rightarrow6)$-linked carbohydrate has only recently been published, that of panose[12,13] (§6). Thus it is all the more imperative to develop theoretical models to understand its properties.

The simplest model for the $\alpha(1\rightarrow6)$-linkage is isomaltose, (α-D-glucopyranosyl-$(1\rightarrow6)$-D-glucopyranose), shown in figure 5.1, which also indicates the conventions used for the naming of residues and glycosidic torsion angles used. For the purposes of this chapter, the following definitions will be used for the glycosidic dihedrals:

Φ :1O5-1C1-1O1-2C6
Ψ :1C1-1O1-2C6-2C5
Ω :1O1-2C6-2C5-2C4

In addition, the orientation of the primary alcohol on the non-reducing residue will be denoted by Θ and defined by the C4-C5-C6-O6 torsion angle.

The $\alpha(1\rightarrow6)$-linkage is generally believed to be more flexible than the $\alpha(1\rightarrow4)$, due to the greater number of bonds separating the residues, and this is thought to be the reason that amylopectin is more water-soluble than amylose.

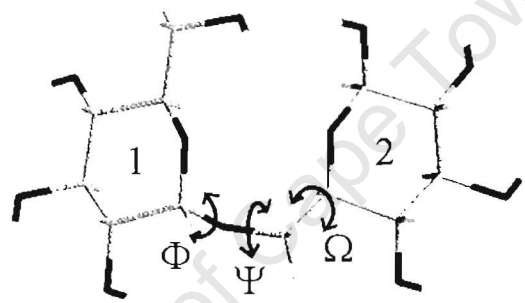


Figure 5.1: Residue numbering and torsion angle name conventions for Isomaltose

5.1 Relaxed Adiabatic Map of Isomaltose

The first step in exploring the potential energy of isomaltose is the compilation of an adiabatic minimum potential map. In addition to the three glycosidic dihedrals, isomaltose has four secondary alcohols per residue and an hydroxymethyl group which each have torsional flexibility. Since each secondary alcohol in principle has 3 feasible gauche arrangements and each primary alcohol has two dihedrals, each with three gauche rotamers, the combinations of these, together with the glycosidic linkages, define a crude conformational space of 3^{15} or about 14×10^6. Clearly an exhaustive search of these conformations with current computers would be impossible. Thus this presents a classic problem of multivariate minimization.

5.1.1 Methodology

Previous approaches to conformational searching in disaccharides have attempted to reduce the number of conformations available, by making the (very reasonable) assumption that the secondary alcohols in each glucosyl ring favour hydrogen bonding to each other so as to form a "crown" of hydrogen bonds going either clockwise or anticlockwise around the ring,[23,50] and adopting various minimization methods to explore this reduced conformational space. An alternative to this approach is a method known as "simulated annealing",[97] which involves the use of molecular dynamics at a high temperature as a means of overcoming energy barriers to find the lowest minimum. This is the method that was chosen for Isomaltose, using the protocol described by Naidoo et. al:[98]

1. **Initial grid search.** The (Φ, Ψ, Ω) space of isomaltose was divided into a grid by sampling at 20^o increments of each angle, first in Φ, then in Ψ and Ω. Effectively, a stack of Ψ-Ω maps were generated, one for each value of Ψ. At each point in this map, the carbohydrate was subjected to simulated annealing according to the following method the glycosidic dihedrals being harmonically constrained to their reference grid values and the endocyclic torsion angles being constrained to prevent "ring-flipping" to $^1C_4$ and twist structures:

    - **Heating:** The molecule was "heated" from 200K to 900K in increments of 100K per picosecond over 9ps by rescaling the atomic velocities. It was then maintained at 900K for 3ps of equilibration in order that it explore the available conformational surface.
    - **Cooling:** The molecule was "cooled" from 900K to 300K over 20ps, in decrements of 60K every 2ps. The rate of cooling was deliberately slower in order

to prevent the molecule becoming trapped in false minima, while still forcing it into lower energy regions.

- **Dynamic Quenching:** After 2ps the temperature was dropped to 250K, and after a further 2ps to 200K, at which it was run for 11ps, to allow the molecule to explore its newfound potential well by exchanging kinetic and potential energy.
- **Minimization:** A final 1000 step conjugate gradients minimization was performed to find the local minimum.

2. **Refinement of local minima.** Although the simulated annealing of the first step would have a good chance of finding low minima, in order to explore the important low energy regions of the map more comprehensively, a process of refinement was conducted at selected minima: for each slice in Φ, five minima were chosen, making 90 minima in all. Each was subjected to 30 cycles of further simulated annealing as described in step 1, and the lowest resulting conformation was chosen for each.

3. **Smoothing to obtain final map.** The minima from step 2 were used as trial structures in the formulation of a final "smoothed" map, in which all the original grid points would be refined. At every (Φ, Ψ, Ω) grid point, six starting structures were considered: the minimum for that point from step 1, and the five refined structures with the same Φ value from step 2, with their glycosidic dihedrals edited to the actual grid point values. The lowest energy structure obtained after minimizing all the trial structures was used as the point in the final map.

The resulting energy map was binned into a three-dimensional data structure for analysis.

5.1.2 Features of Map

Despite the large volume of conformational space available due to the extra glycosidic torsion angle, all of the significant local minima in the adiabatic map for the $\alpha(1\rightarrow6)$-linkage turn out to be located in a single low energy region. The complete map is represented as a stack of nine Ψ-Ω maps in figure 5.2, in which the low energy regions are darker, and can be identified in the centre of each map. The minimum energy region has been subdivided into several local energy minima: each point in the three-dimensional map was analysed for being a local minimum, and six clear low-lying wells were found, all within 2 kcal/mol of the global minimum. These have been labelled A to F, as listed in table 5.1.

Φ	Ψ	Ω	**Label**
80	180	180	A
80	180	60	B
80	240	200	C
80	100	40	D
80	80	200	E
140	180	60	F

Table 5.1: Minima from Isomaltose Adiabatic Map

The wells have been illustrated by three representative 2-dimensional slices in figures 5.3, 5.4 and 5.5. The network of minima can essentially be described as two sets having Ω approximately equal to 60^o and 200^o respectively, and in which almost all of the low energy values have Φ restricted to a narrow range about 80^o.

Newman projections down the three interresidue bonds are shown in figure 5.6. The Φ dihedral has only one 1,4-clash on transition between rotamers as does Ψ (since O1 has only C6 attached), while Ω has three (since both termini of the bond have three substituents). This difference is most likely the main reason for the high energy barrier between the two sets of wells {A,C,E} and {B,D,F} relative to the energy penalties for transitions to wells within each set of minima. A simple consideration of 1,4 steric clashes using the Newman projections in figure 5.6 reveals that the Φ dihedral would favour the g^+ and t rotamer over the g^- orientation. Consideration of the three dimensional structure of the t rotamer reveals a steric clash between the C6 methylene group in the linkage and O2 on residue one. Hence, the $g+$ rotamer, with a value of about $70-80^o$ is favoured in almost all the minima. Likewise, most of the minima have their Ψ values close to the trans value predicted by a simple analysis based on the Newman

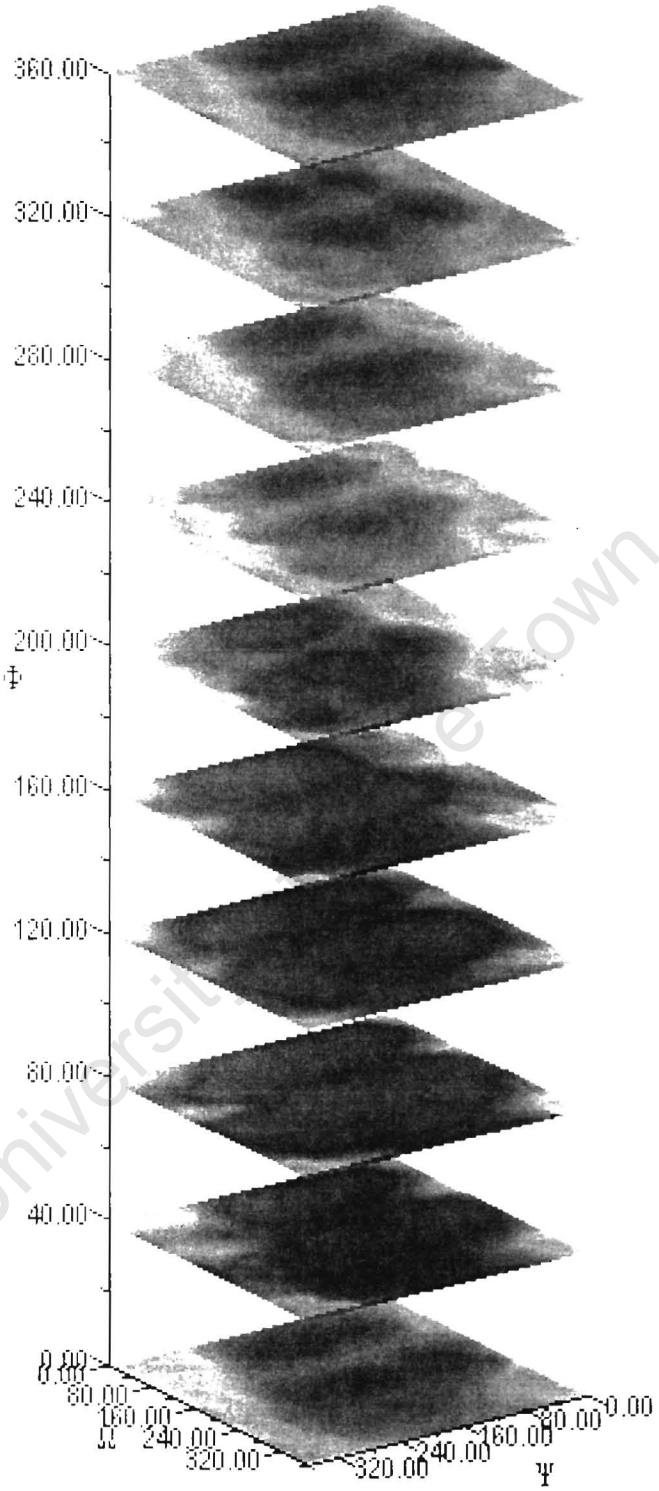


Figure 5.2: Stacked representation of Ψ-Ω maps at 9 values of Φ. Darker areas indicate low energy regions.

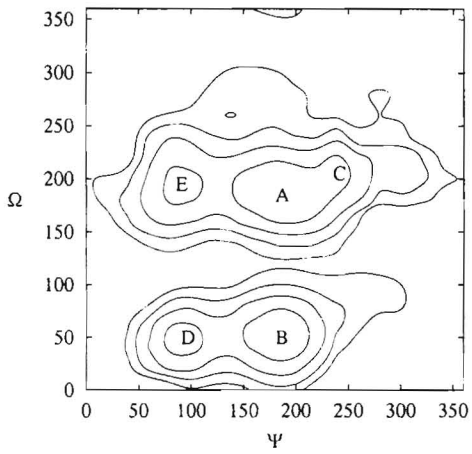


Figure 5.3: Slice through isomaltose adiabatic map for $\Phi = 80$

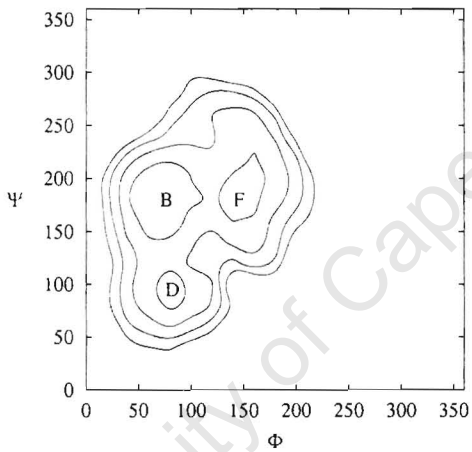


Figure 5.4: Slice through isomaltose adiabatic map for $\Omega = 60$

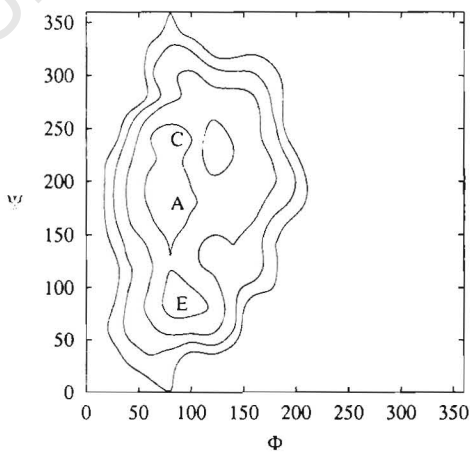


Figure 5.5: Slice through isomaltose adiabatic map for $\Omega = 200$

projection. The favoured values for Ω are the same *gg* and *gt* rotamers preferred by the free hydroxymethyl group.

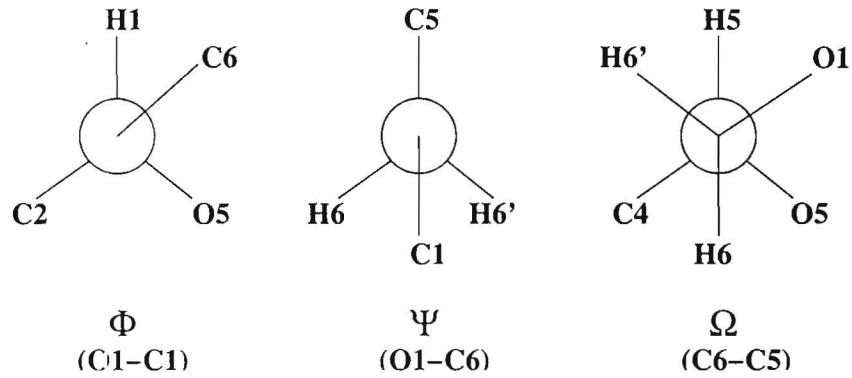


Figure 5.6: Projections of Φ, Ψ, Ω rotamers

The molecular structures corresponding to the conformational minima A-F differ principally in the values of the three glycosidic dihedrals Φ, Ψ, Ω, with the notable exception of well C. All of wells A,B,D,E,F share the same properties for each residue:

1. The secondary alcohols form a clockwise crown in ring 1 (seen from above the ring), and an anticlockwise crown in ring 2.

2. The hydroxymethyl group is *gg* in all, and the C5-C6-O6-H6 torsion is in the g^+ rotamer.

Well C is the exception, having a different conformation for ring 1, where the secondary alcohols form an anticlockwise crown, the hydroxymethyl is *gt* and the C5-C6-O6-H6 torsion is g^-. These changes are most likely necessary to accommodate the interresidue hydrogen bond between O6 of residue 1 and H1-O1 of residue 2 (see figure 5.7), the only such hydrogen bond found in all of the minima. This could also be the reason that this structure forms a distinct minimum in the adiabatic map, even though it is very close to the "global" minimum A. Despite being close in the adiabatic map, wells A and C are favoured for different reasons. Well A is low in energy mainly as a consequence of having low steric strain associated with its glycosidic dihedrals. Hydrogen bonding does not appear to play an important part in the stability of well A, since most of the other minima display a similar extent of hydrogen bonding. However, well C forms several more hydrogen bonds in exchange for a slight deformation from the sterically ideal dihedrals of well A. The lack of interaction between the rings in most of the minima may possibly explain the postulated flexibility about the $\alpha(1\rightarrow6)$-link.

Reaction path optimization[100] is a method for calculating the lowest energy, or saddle path between two states. The optimized reaction path between the four main minima

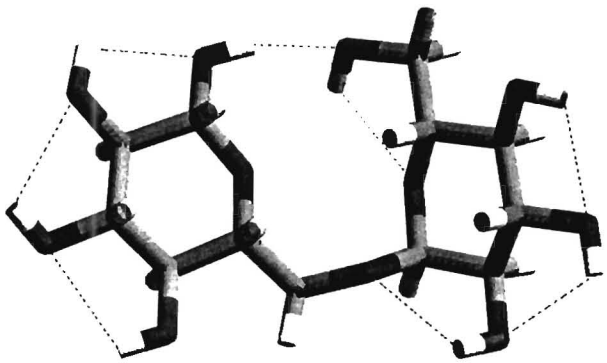


Figure 5.7: Structure of isomaltose corresponding to minimum C

(A, B, D, E) was calculated using CHARMM, following the cycle of paths (i), (ii), (iii) and (iv) shown in figure 5.8. The points between the minima in the adiabatic map were used as "seeds" in the calculation. The energy profile along the resulting reaction coordinate gives an indication of the energy barrier along the "ideal" path. Plots of the energy profiles for each transition are shown in figure 5.8. The barriers to rotation about Ψ, namely for the A to E and D to B transitions are less than 2 kcal in size. By contrast the transitions in Φ in the B to A and E to D transitions require approximately 9 kcal to overcome the intervening barrier. Thus the principal degree of freedom in the $\alpha(1\rightarrow6)$-linkage appears to be the Ψ dihedral, since rotation about Ω is energetically very restricted, and Φ is close to 80^o for all of the four major minima. Dynamics simulations at 300K in vacuum are usually able to overcome barriers of around 2 kcal, but would not have sufficient energy to cross a 9 kcal energy hill. It should also be noted that since minima A and B differ by only about 0.5 kcal in energy, both would be expected to be populated at room temperature. However, the high energy barrier between them would cause a slow rate of transition.

The structures of the four major minima and the transition states from the cyclic transition pathway are depicted in figure 5.9. The features of the transitions leading to the height of the energy barriers may be identified. The transitions in Ψ (transition states AE and DB) are seen to have only one 1,4 clash between one of the H6 hydrogens on the second residue and the C1 carbon on the first. The transitions in Ω have three 1,4 clashes: between O1 of residue 1 and O5 of residue 2, between one H6 and H5 of residue 2 and between the other H6 and C4 of residue 2. The analysis using the Newman projection appears therefore to give a valid reasoning for the relative heights of the different energy barriers. A final point is that the transition state ED is the only structure with its primary alcohol in *gt* rather than *gg*. However, this rotamer does not seem to be forming any favourable interactions, so it is possible that this is an artefact of the refinement of the map: the structures furthest from the minima would

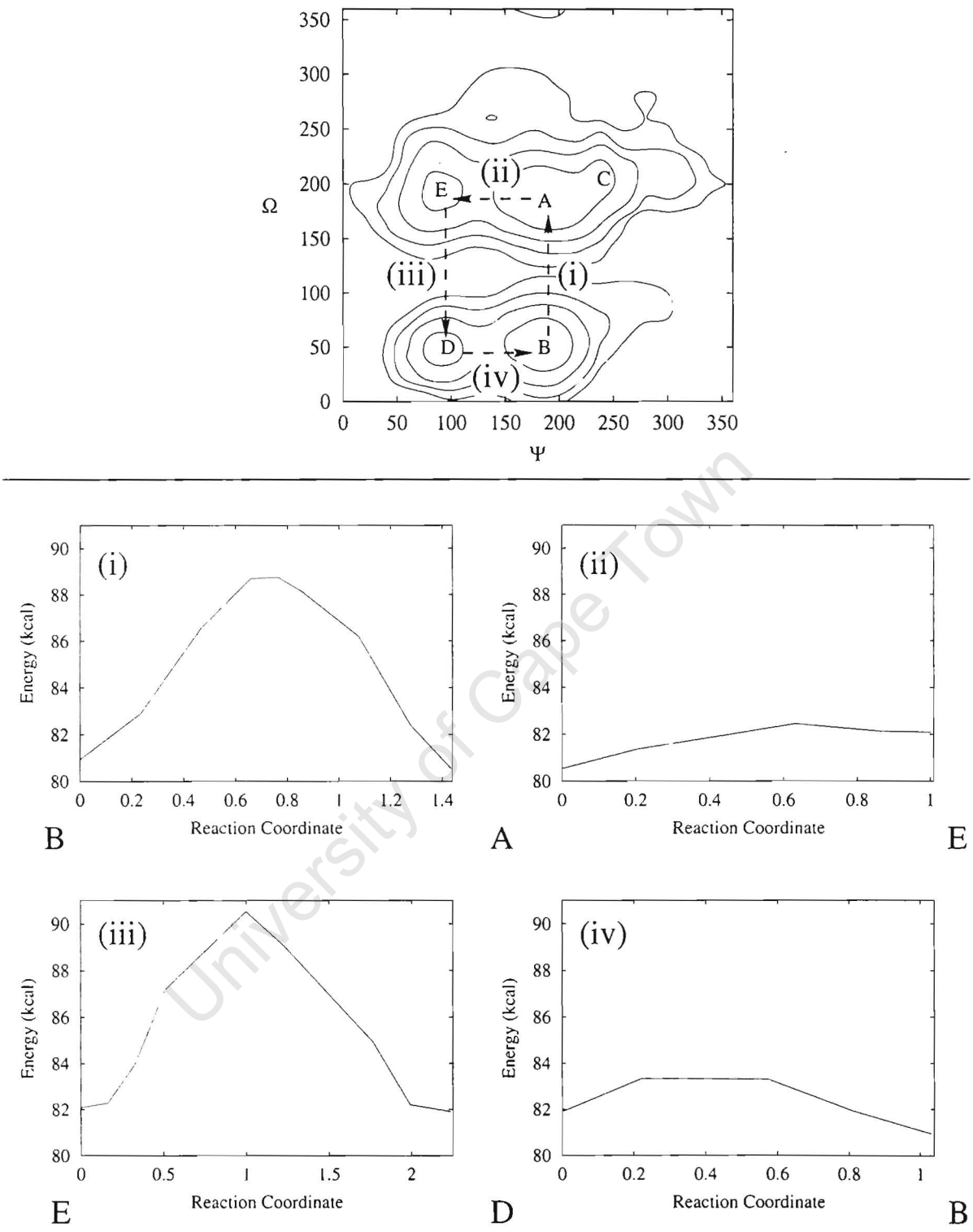


Figure 5.8: Definition of interwell transition pathways (above) in the isomaltose adiabatic map, and calculated energy profiles for each transition (below).

have benefitted the least from the third stage of refinement using optimized structures located within the minima.

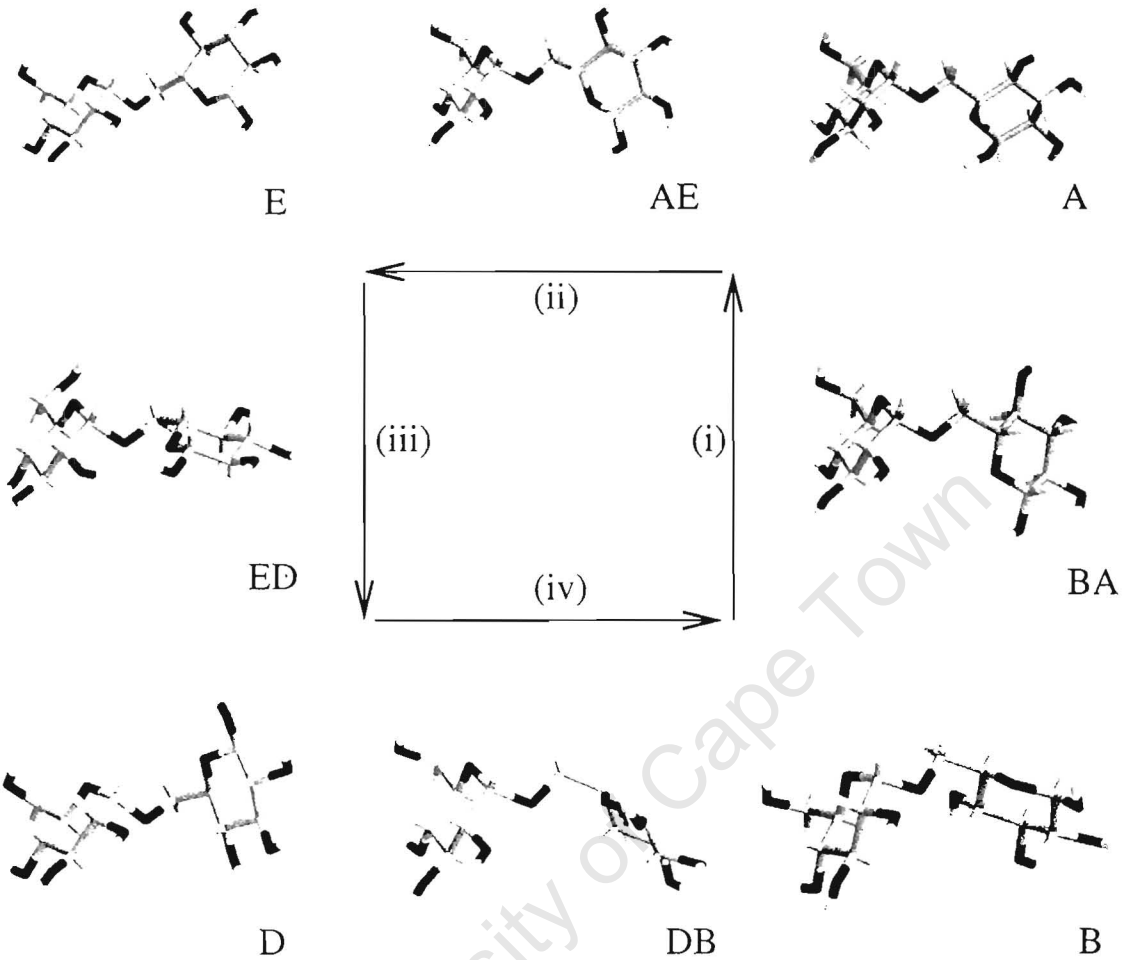


Figure 5.9: Structure of isomaltose for the four major wells A, B, D and E and the transition states (BA, AE, ED and DB) connecting them in the transition pathways.

5.2 Molecular Dynamics Simulations of Isomaltose

Since it is the dynamical behaviour of isomaltose that is being sought, a simple adiabatic map will only give an indication of the conformational regions likely to be traversed in vacuum, and an approximate idea for water. In addition it does not provide any dynamical information such as the rate of conformational transitions, or the degree of conformational flexibility. Molecular dynamics simulations are able to provide a complete time evolution of the molecule, and with the explicit inclusion of water are able to realistically model a solution environment.

5.2.1 Simulation Methodology

Based on the three-dimensional adiabatic map, the six lowest lying minima A-F were chosen as starting points for vacuum simulations, and the three lowest A-C for solution simulations. Notably, the one conformer found to have an interresidue hydrogen bond in the adiabatic minimum, C, was included in both studies. The conformation corresponding to the crystal structure of panose[12,13] is also effectively included amongst the starting structures, since it falls in well A. Thus, simulations have been started from both sides of the large 9 kcal energy barrier between the two sets $\{A, C, E\}$ and $\{B, D, F\}$ in the vacuum adiabatic map, so that all regions can potentially be explored.

Vacuum Simulation

Each of the chosen minima (wells A-F) was heated for 25ps from 50K to 300K to avoid a sudden increase in energy. Constraints were placed on the interresidue dihedrals, to keep them at the minimum values. The constrains were then released and dynamics was allowed to proceed for 1 ns using the Verlet integrator.

Water Simulations

A truncated octahedral water box of side 36.9953Å containing 846 TIP3P[66] water molecules was used to solvate each of the three lowest energy structures from the adiabatic map (wells A to C). Each solute was placed in the center of the box and all water molecules with heavy atom distances from the solute of less than 2.76 Å were deleted. To adjust the density to the experimentally found value, the box size was reduced by coordinate scaling followed by energy minimization. The system was slowly heated from 100K to 300K over 20ps and allowed to equilibrate for a further 50ps at 300K and constant volume, using leapfrog verlet with velocity reassignment to maintain the correct temperature. During heating and equilibration, the kinetic, potential and total energy were monitored to check for stable partitioning. Each of the three systems

was then subjected to a further 2ns of leapfrog verlet dynamics at constant volume and energy. It was decided to use the microcanonical (NVE) and not the canonical(NVT) ensemble in order to preserve the time history as best possible, as this is required for the calculation of correlation functions and NMR properties. Of the two commonly used methods to ensure constant temperature, velocity reassignment is obviously detrimental to the calculation of time correlation functions, while velocity scaling is known to have detrimental effects, such as the "hot solvent, cold solute" phenomenon and a transfer of random thermal velocity to collective motion about the centre of mass.[100] Observation showed that after equilibration, the kinetic energy and temperature fluctuated about a steady value, the temperature being stable at 300K.

5.2.2 Pucker Analysis

The glucopyranosyl ring, like its alkane analogue cyclohexane, is able to adopt various conformations such as the boat, chair and twist-boat. Experimental evidence indicates that for carbohydrates in solution, only the chair form is present; therefore, any simulated data should reproduce this preference.

The puckering of the rings for the molecular dynamics simulations of isomaltose was analyzed using the Cremer and Pople puckering parameters for six-membered rings[57] (see2.3.3). All rings in all the simulations gave a puckering magnitude Q, of about 0.57, which is typical for cyclohexanoid rings, a value of Θ of about 7^o and values for Φ covering the full range from 0^o to 360^o. A value for Θ of 0^o corresponds to desired chair conformation. Thus all the rings are close to the chair, with a slight constant distortion introduced by the asymmetry of the ring. The value of Φ makes little difference to the conformation at such small values of Θ, and so can be ignored.

5.2.3 Cluster Analysis

Isomaltose structures obtained from dynamics were subjected to a separate cluster analysis for each vacuum and simulation run. The clustering was done according to the interresidue dihederals (Φ, Ψ, Ω) and the free hydroxymethyl group, Θ, since these are the main determinants of conformational variation for this disaccharide. A cluster, or threshold radius (§2.3.3), was chosen to give the least cluster overlap, while at the same time not artificially separating clusters. The most populated clusters for vacuum are shown in table 5.2 and for solution in 5.3.

sim.	cluster	# members	std. dev.	Φ	Ψ	Ω	Θ
a	1	11931	0.785e+01	71.113	187.823	52.722	50.904
	2	6817	0.798e+01	70.218	173.555	52.714	51.299
b	1	12457	0.776e+01	69.648	186.851	52.600	52.80
	2	7543	0.794e+01	69.689	174.356	53.870	54.22
c	1	8305	0.871e+01	70.928	194.465	189.759	50.433
	2	4062	0.884e+01	71.071	169.884	169.097	50.225
	3	2351	0.239e+01	72.229	225.889	197.342	295.905
d	1	11488	0.814e+01	69.141	186.713	52.894	53.860
	2	2103	0.567e+01	65.761	174.042	57.141	46.130
e	1	10323	0.849e+01	70.744	194.888	189.513	50.210
	2	5517	0.798e+01	70.929	170.535	190.505	50.240
	3	2168	0.826e+01	70.567	295.159	175.801	49.430
	4	987	0.944e+01	85.119	90.051	194.104	49.750
f	1	11931	0.785e+01	71.113	187.823	52.722	50.900
	2	6817	0.798+01	70.218	173.555	52.714	51.290
	3	967	0.940e+01	150.507	194.498	55.342	49.780

Table 5.2: Largest clusters arising from isomaltose trajectories in vacuum

sim.	cluster	# members	std. dev.	Φ	Ψ	Ω	Θ
a	1	17415	0.939e+01	70.748	196.159	193.999	56.887
	2	11805	0.930e+01	68.771	167.141	192.445	56.719
	3	3541	0.864e+01	71.535	195.370	175.035	56.907
	4	2630	0.108e+02	66.764	84.236	191.676	57.921
b	1	20764	0.864e+01	71.568	189.659	54.626	56.774
	2	16354	0.926e+01	69.815	169.841	54.271	56.574
	3	2882	0.992e+01	71.074	107.249	52.475	55.858
c	1	10573	0.834e+01	70.035	194.969	192.499	189.107
	2	9369	0.880e+01	68.146	167.510	192.133	189.643
	3	5613	0.817e+01	70.471	194.283	192.178	173.614
	4	1288	0.112e+01	71.188	88.688	190.714	189.879

Table 5.3: Largest clusters arising from isomaltose trajectories in solution.

Clustering in Vacuum

The first feature of the clusters that stands out in the vacuum runs is that there is not a large amount of variation in the dihedral values sampled during the simulations, and that a large number of the structures fall within only a few clusters. The Φ torsion angle remains essentially the same at 70^o across virtually all clusters, in accordance with the preference for this angle in the adiabatic maps. The only exception to this is the vacuum run from F, which has a small cluster (corresponding to the F well) with $\Phi = 150^o$, but this almost immediately reverts to 70^o as the molecule makes the transition to the B well. The orientation of the hydroxymethyl group, given by Θ, remains constant throughout each simulation, which is a known feature of the applied force-field. It remains in the preferred *gg* rotamer for most of the simulations, with the exception of the one initiated from well C, which started in the *gt* rotamer (cluster 3). However, the primary alcohol in this simulation clearly makes a transition to *gg*, which is found in the larger two clusters (1 and 2).

The remaining Ψ and Ω torsions are in practice the only significant degrees of freedom. In accordance with the adiabatic map, which showed a high energy barrier between well B and wells A and C, it turns out that no transitions between these sets of minima were observed at all: the clusters from the run starting from minimum B have all of their Ω values at around 55^o, while those for wells A and C are both close to 180^o. In fact, the runs starting from A and C are quite closely related, since the two wells are are connected by a low energy path in the map. The major distinction is the hydroxymethyl orientation, but that may well be an artefact of the high rotational barrier in the force-field.

Few interwell transitions between neighbouring minima are observed for the vacuum simulations. Each run samples the minimum from which it was started, while runs C-F sample others as well. Runs C and E sample minimum A and runs D and F sample B.

Clustering in Solution

The clusters obtained from the solution simulations followed the same basic pattern of allowed values of Φ, Ψ and Ω as the vacuum cases. However, a somewhat greater range of dihedral angle space is sampled, especially for the runs starting from A and B. These both sample multiple wells unlike the vacuum runs started from these minima. In solution, the runs starting from minima A and C both visit well E for short periods during the simulation, while that starting from minimum B visits well D. Thus all of the major minima in the map are sampled in the solution dynamics, apart from well F, which is the highest in energy. Cluster 1 from simulation A and clusters 1 and 3 from simulation C are practically the same and represent minimum A on the map. Likewise,

cluster 2 from simulation A and cluster 2 from simulation C are the same (corresponding to a different region of minimum A), as is cluster 4 from each of A and C (corresponding to minimum E). Cluster 3 of A also falls broadly within minimum A. For run B, clusters 1 and 2 are located in minimum B, while cluster 3 is in D.

The primary alcohol favours the *gg* rotamer in solution just as in vacuum, however the run from the C well remains in the *gt* rotamer, while that in vacuum made transitions. This orientation for the C well opens the possibility of interresidue bridging hydrogen bonds in water, similar to the direct interresidue hydrogen bonds observed for this minimum in the vacuum adiabatic well.

In summary, all minima A through E were sampled in solution. Minima A, C and E were sampled in the simulations starting from A and C, while minima B and D were sampled in that starting from B. Significantly, no transitions were observed between the sets {A, C, E} and {B, D}.

5.2.4 Time dependence of molecular motion

Although cluster analysis gives a good idea of which minima are sampled, it does not indicate the rate of transitions, nor the fluctuations within each well. This can be provided by a set of time series plots for each dihedral angle, which clearly illustrate the exact history of that torsion angle.

A sample of plots for the vacuum simulation from well A and for the solution simulation from well A are shown in figures 5.10 and 5.11 respectively, while those for the other simulations are given in appendix A. The essential difference between the dynamics in vacuum and in water is that the water permits a wider variation in the dihedrals: the vacuum dynamics is tightly restricted to a particular well in the vacuum adiabatic map, while in the solution run, the dihedrals are able to explore more widely areas that were less favourable in the adiabatic map. Water also facilitates a more gradual transition between wells: transitions in vacuum (not present in the example in the text) are observed to take of the order of 1 ps from start to end, while the transitions in solution (as in figure 5.11) typically last up to 50 ps. In addition, the regular fluctuations in vacuum are also very much more rapid than in solution, although this may not be immediately clear from the time series.

The actual sampling of minima in the various wells shows that the majority of the time is spent in either well A or well B, with only short transitions to other wells.

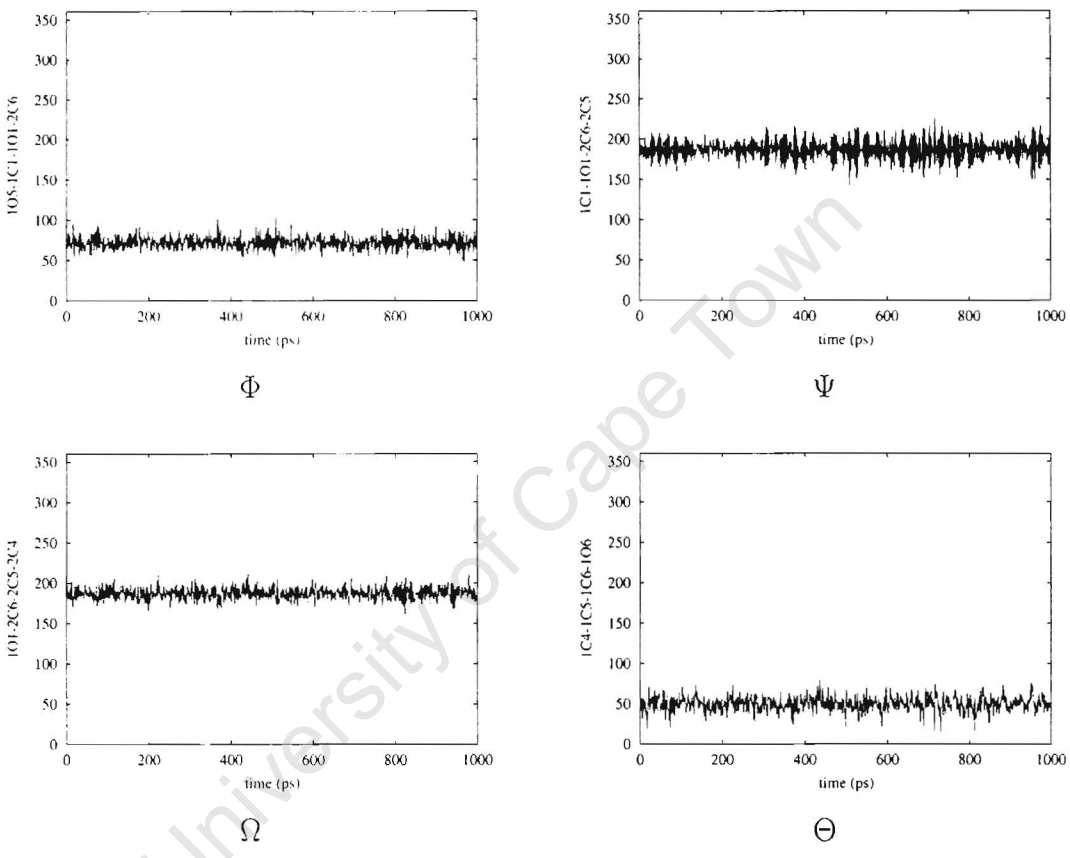


Figure 5.10: Time series for isomaltose in vacuum from minimum A

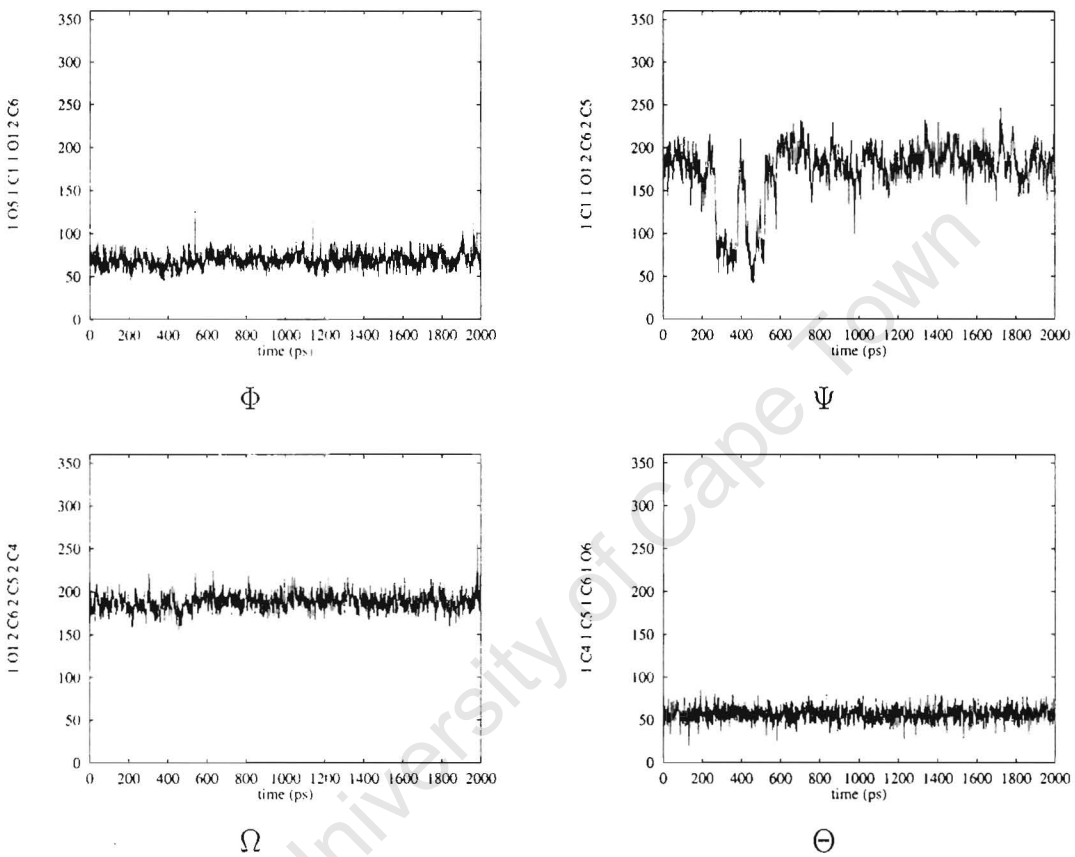


Figure 5.11: Time series for isomaltose in solution from minimum A

The sampling of the vacuum adiabatic energy map by the dynamics runs may be evaluated by projecting the dihedral trajectories onto the Φ and Ω slices used previously in defining the maps. Such maps are given in figures 5.12 to 5.17 for vacuum and 5.18 to 5.20 for solution. The principal difference between vacuum and solution runs is again that the vacuum dihedrals are confined to a much narrower range.

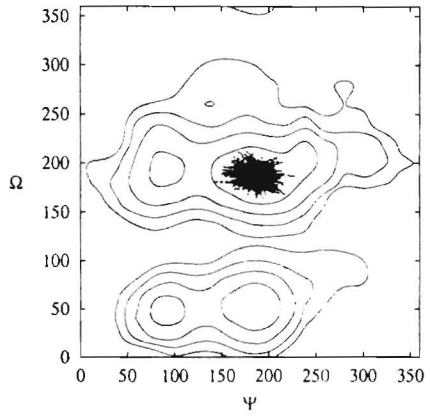
Examination of the map projections for the vacuum simulations shows that the preferred minima during dynamics lie very close to the low energy regions of the vacuum adiabatic map, as might be expected. In particular, wells A and C, which are separated by only a small energy barrier, are clearly separately sampled in figure 5.14. The transitions between all the wells consist of very few points, supporting the earlier conclusion that the transitions in vacuum are rapid.

By contrast, the sampling by the simulations in solution is more diffusely spread around the minimum energy wells, demonstrating that the water does indeed allow the dissaccharide to explore wider regions of the vacuum adiabatic map. In addition, the saddle regions between wells are sampled far more in solution than vacuum, as one would be led to conclude based on the duration of the transitions. Wells A and C, which were distinct in the vacuum simulations, are merged together in a single broad region. This explains the absence of distinct A and C conformers in the cluster analysis of the runs starting from A and C.

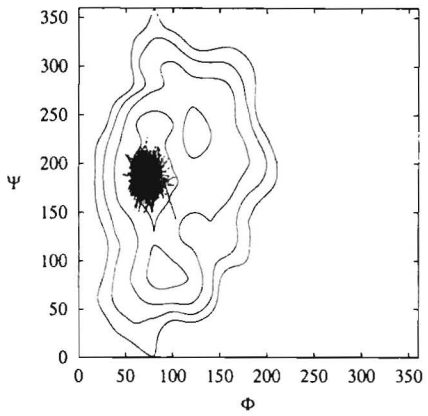
The vacuum runs starting from well A and well B do not make any transitions away from those conformations, consistent with the fact that these are the deepest energy minima. The runs starting from other conformations all make transitions into either of wells A or B, which are then sampled for most of the trajectory, with only occasional transitions being made to the higher energy minima. As noted before, there are no transitions between the two sets of wells separated by a transition in Ω, while all other possible transitions are observed.

The orientation of the secondary alcohols does not play a role in the scheme we have adopted for describing isomaltose conformations; however, the secondary alcohols certainly play a role in determining the conformation, so their distribution is also of interest. An example of the same secondary alcohol in vacuum and solution simulations is shown in figure 5.21. The effect of water is firstly to allow the hydroxyl group to significantly sample more than rotamers, and secondly to broaden each of the peaks, corresponding to the three gauche conformations. This could be interpreted as the competing hydrogen bonds from the water reducing the importance of the intramolecular "crown" of hydrogen bonds, thus facilitating transitions.

Water also has the effect of reducing the correlation between the orientation of the secondary alcohols in one ring. The time series for secondary alcohols in the same ring in

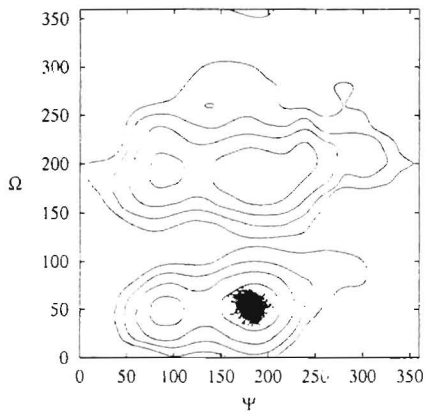


($\Phi = 80$)

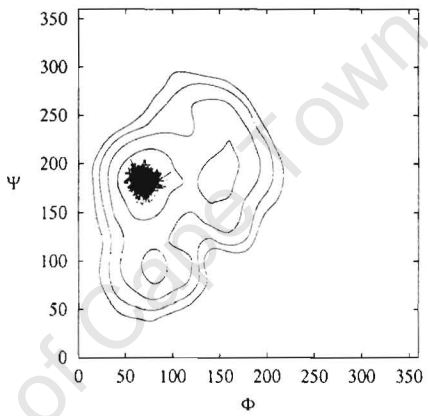


($\Omega = 200$)

Figure 5.12: Projection of MD trajectory for isomaltose onto adiabatic map slices (Vacuum run from minimum A).

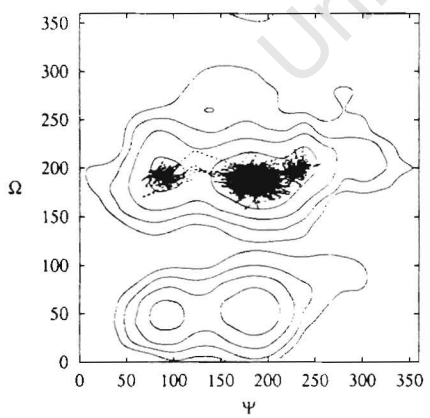


($\Phi = 80$)

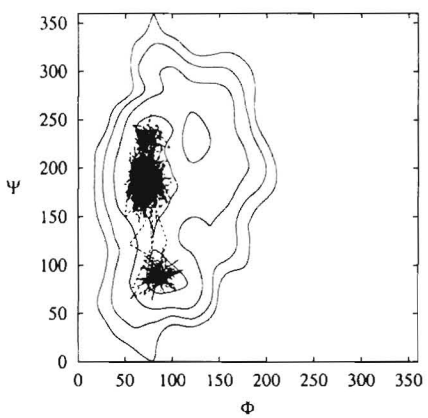


($\Omega = 60$)

Figure 5.13: Projection of MD trajectory for isomaltose onto adiabatic map slices (Vacuum run from minimum B).

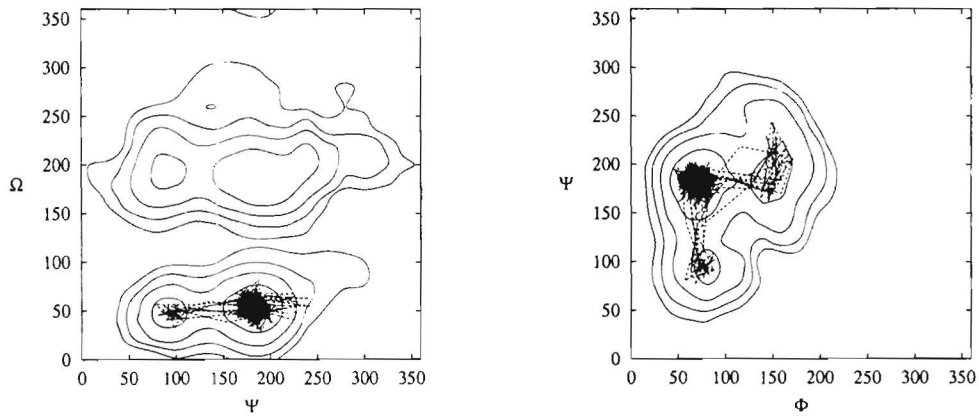


($\Phi = 80$)



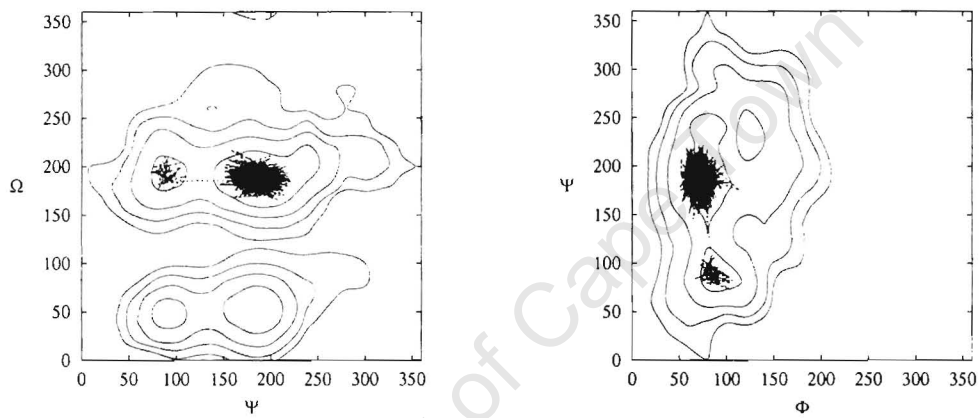
($\Omega = 200$)

Figure 5.14: Projection of MD trajectory for isomaltose onto adiabatic map slices (Vacuum run from minimum C).



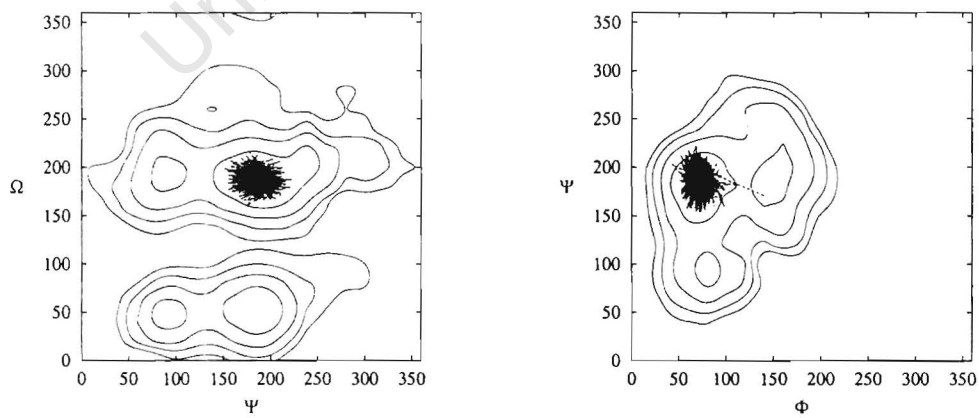
($\Phi = 80$) ($\Omega = 60$)

Figure 5.15: Projection of MD trajectory for isomaltose onto adiabatic map slices (Vacuum run from minimum D).



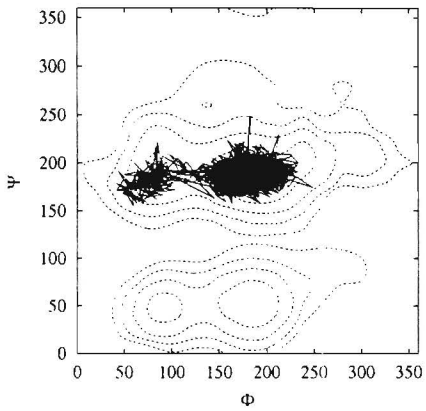
($\Phi = 80$) ($\Omega = 200$)

Figure 5.16: Projection of MD trajectory for isomaltose onto adiabatic map slices (Vacuum run from minimum E).

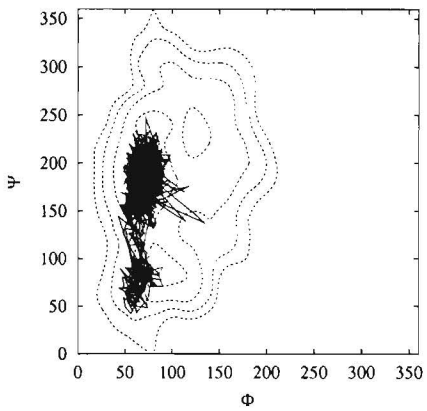


($\Phi = 80$) ($\Omega = 60$)

Figure 5.17: Projection of MD trajectory for isomaltose onto adiabatic map slices (Vacuum run from minimum F).

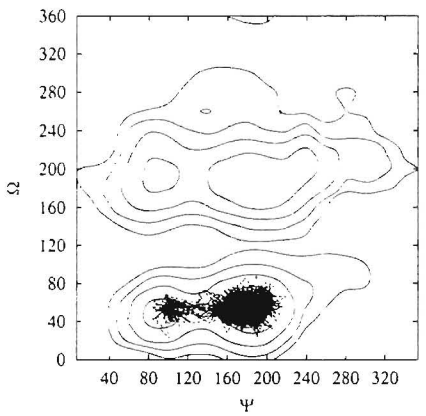


($\Phi = 80$)

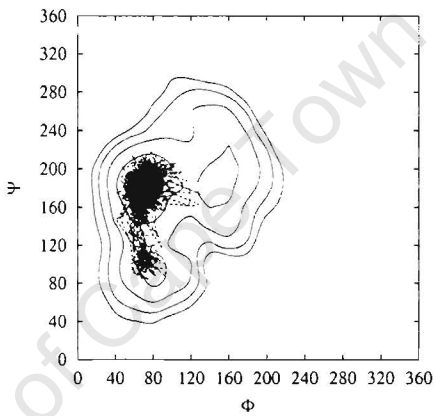


($\Omega = 200$)

Figure 5.18: Projection of MD trajectory for isomaltose onto adiabatic map slices (Solution run from minimum A).

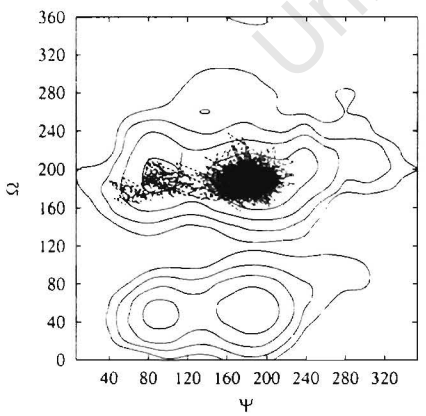


($\Phi = 80$)

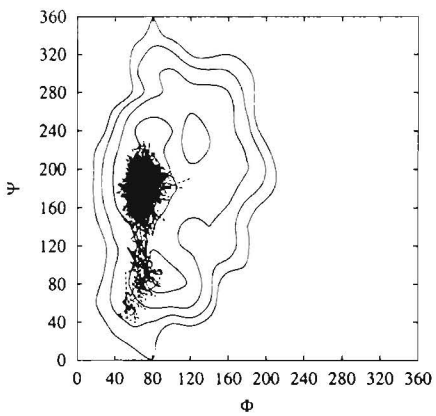


($\Omega = 200$)

Figure 5.19: Projection of MD trajectory for isomaltose onto adiabatic map slices (Solution run from minimum E).

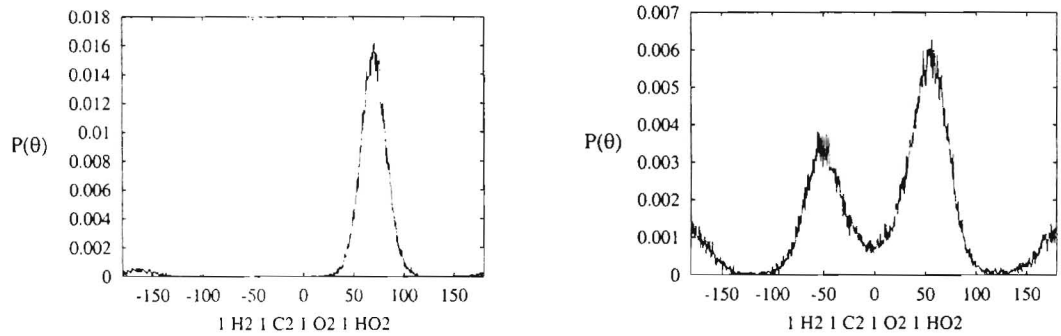


($\Phi = 80$)



($\Omega = 200$)

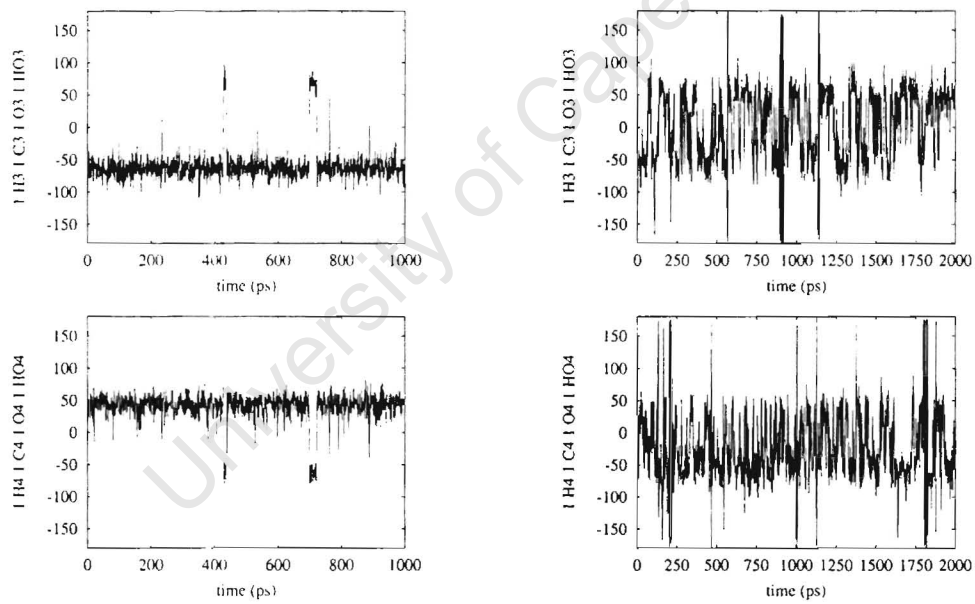
Figure 5.20: Projection of MD trajectory for isomaltose onto adiabatic map slices (Solution run from minimum C).



Vacuum Solution

Figure 5.21: Secondary alcohol rotameric distribution for a selected hydroxyl in vacuum and water

vacuum (figure 5.22) are closely correlated, since the optimal hydrogen bonding occurs when all are pointing in the same direction around the ring. In solution, the water provides an alternative hydrogen bonding partner, allowing the secondary alcohols to move more independently, as is seen in figure 5.22, although there is still clearly some residual correlation.



Vacuum Solution

Figure 5.22: Time evolution of adjacent secondary alcohols in vacuum and in water

5.2.5 Hydrogen Bonding Analysis

One of the characteristic features of water as a solvent is its capacity for forming directed "hydrogen bonds" both with itself and with solutes such as saccharides. As a result, the distribution of water around a dissolved sugar is very spatially directed. This is usually described in terms of hydrogen bonding of the solute with the solvent; previous simulations have demonstrated hydrogen bonding between the sugar hydroxyl groups and the water,[23,50,55,56,76] as well as instances of water molecules forming "bridging" hydrogen bonds between two hydroxyl groups on adjacent residues.[50] Since modern force-fields such as those used in CHARMM do not include explicit hydrogen bond energy, it is necessary to adopt a working definition, which is usually based on donor-acceptor distance as well as acceptor-donor-acceptor angle. This formulation does pose problems when calculating hydrogen bond lifetimes, however, since a transient distance change often does not cause a change in the associated energy, which is the true indicator of hydrogen bonding, but much harder to calculate. Although definitions based on a rolling average of donor-acceptor distance have been proposed,[55] they require data saved at a high frequency during the simulation, and are not generally practical.

The four minima sampled during the solution simulations of Isomaltose, namely A (merged with C), B, D and E were each represented by selected sections of simulation data, as were the six minima from the vacuum simulations. The chosen sections are detailed below:

	Vacuum		Solution	
Well	Simulation from:	Time Section	Simulation from:	Time Section
A	well A	1000-1500ps	well A	0-1000ps
B	well B	500-1000ps	well B	0-1000ps
C			well C	0-50ps
D	well B	1550-1650ps	well D	0-50ps
E	well A	300-350ps	well E	0-50ps
F			well F	0-50ps

Clearly, more statistically significant data will be obtained from the longer time sections, but it is desirable to have as much information as possible for each well. For each hydroxyl in solution, a search was done for direct hydrogen bonds to the rest of the solute, direct hydrogen bonds to water and bridging hydrogen bonds through water. In vacuum, only the intramolecular hydrogen bonds could be analysed. The criteria for a hydrogen bond were (i) an angle of 120^o to 180^o for the acceptor-donor-acceptor angle and (ii) an acceptor-donor distance of less than 2.4Å. Criterion (i) was relaxed for intramolecular hydrogen bonds to be 100^o in view of the fact that the proximity of

the participating groups permits a more strained geometry. The resulting values are recorded in table 5.4 for the vacuum simulations and tables 5.5 to 5.8 for the solution simulations. Note that both donor and acceptor hydrogen bonds have been included in the analysis where applicable.

Hydroxyl	A	B	C	D	E	F
1 O1	0.02 (0.10)	0.25 (0.11)	0.57 (0.11)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
1 OH2	0.47 (0.06)	0.69 (0.07)	0.96 (0.08)	0.49 (0.04)	0.47 (0.04)	0.54 (0.04)
1 OH3	0.75 (0.06)	0.73 (0.06)	0.76 (0.06)	0.79 (0.06)	0.75 (0.06)	0.83 (0.07)
1 OH4	0.35 (0.06)	0.33 (0.06)	0.37 (0.06)	0.34 (0.06)	0.32 (0.06)	0.34 (0.06)
1 O5	0.08 (0.05)	0.13 (0.06)	0.32 (0.08)	0.08 (0.05)	0.07 (0.05)	0.09 (0.05)
1 OH6	0.14 (0.06)	0.17 (0.06)	0.79 (0.40)	0.12 (0.06)	0.11 (0.06)	0.14 (0.06)
2 OH1	0.24 (0.03)	0.24 (0.03)	0.72 (0.39)	0.25 (0.03)	0.27 (0.03)	0.26 (0.03)
2 OH2	0.57 (0.05)	0.59 (0.06)	0.63 (0.06)	0.62 (0.06)	0.63 (0.06)	0.60 (0.06)
2 OH3	0.74 (0.06)	0.73 (0.06)	0.77 (0.06)	0.67 (0.06)	0.78 (0.06)	0.66 (0.06)
2 OH4	0.41 (0.03)	0.38 (0.19)	0.42 (0.03)	0.30 (0.03)	0.42 (0.03)	0.32 (0.03)
2 O5	0.00 (0.00)	0.00 (0.00)	0.03 (0.05)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Table 5.4: Hydrogen Bonding Statistics for Vacuum Simulations; numbers in brackets refer to average lifetimes in ps.

The hydrogen bonding pattern in vacuum is fairly similar for all sampled minima. The non-hydroxyl oxygens, namely the pyranosyl ring oxygens O5 and the linkage oxygen O1 do not form as many hydrogen bonds as the hydroxyl groups. This could be due to their lack of a donor, but is more likely the result of having a lower partial charge in the force-field (due to the lower C-O bond polarization). The trend for the remaining hydroxyls is that OH2 and OH3, being surrounded on both sides by other ring hydroxyls, form more hydrogen bonds than OH1 and OH4 which are only adjacent to one ring hydroxyl and OH6, which is somewhat distant from potential intramolecular hydrogen bonding pairs.

The major exception in the vacuum dynamics is well C, where 1 O1, 1 OH2, 1 O5, 1 OH6 and 2 OH1 are observed to participate in unusually strong hydrogen bonds. This demonstrates that the well C conformation in vacuum dynamics has similar properties to that in the adiabatic map: figure 5.7 shows the corresponding structure, in which there are hydrogen bonds between 1 O2-H2 and 1 O1, between 1 O5 and 1 O6-H6 and between 2 O1 H1 and 1 O6, exactly the groups which show unusual behaviour in the dynamics. The longer lifetimes of these "anomalous" hydrogen bonds probably reflect the fact that they have a less strained geometry than those formed between adjacent ring

hydroxyls. Although these hydrogen bonds are very significant for the conformation in well C in vacuum, they are unlikely to be as important in solution, due to competitive hydrogen bonding to water molecules, which would be entropically favoured over longer range hydrogen bonds between flexible parts of the molecule.

What is evident from the analysis of the vacuum hydrogen bonds is that the type of hydrogen bonding does not play a strong role in determining the conformation, with the exception of well C. Most of the hydrogen bonds formed are formed within the same ring, and therefore do not have much effect on the linkage torsion angles. Thus it seems as though the major determinant of conformation must be the 1,4 and Van der Waals nonbonded interactions.

Hydroxyl	To Water	Bridging	Internal
1 O1	0.26 (0.13)	0.11 (0.07)	0.03 (0.07)
1 OH2	1.67 (0.48)	0.21 (0.14)	0.23 (0.07)
1 OH3	1.54 (0.53)	0.23 (0.16)	0.31 (0.07)
1 OH4	1.80 (0.62)	0.27 (0.14)	0.12 (0.06)
1 O5	0.35 (0.16)	0.11 (0.07)	0.02 (0.07)
1 OH6	2.17 (0.65)	0.14 (0.10)	0.03 (0.06)
2 OH1	2.34 (0.70)	0.46 (0.24)	0.06 (0.06)
2 OH2	1.76 (0.48)	0.27 (0.15)	0.17 (0.06)
2 OH3	1.61 (0.52)	0.30 (0.20)	0.23 (0.06)
2 OH4	1.57 (0.57)	0.16 (0.17)	0.12 (0.06)
2 O5	0.42 (0.21)	0.22 (0.10)	0.00 (0.00)

Table 5.5: Hydrogen Bonding Statistics for Well A in solution; numbers in brackets refer to average lifetimes in ps.

The statistics for hydrogen bonding in the solution simulations of isomaltose show a lower number of hydrogen bonds formed internally than was the case in vacuum, showing that not only are additional hydrogen bonds made to the water, but that this interaction displaces some of the vacuum hydrogen bonds as well. In general, about half the number of intramolecular hydrogen bonds are made in solution as are in vacuum.

The total number of hydrogen bonds made to water follows the same trend as the total number of intramolecular ones, due to the lower capacity of the non-hydroxyl groups for forming hydrogen bonds. Most of the hydroxyls form between 1.5 and 2.4 hydrogen bonds to water, under the criterion adopted here. If the hydroxyl is thought of as a fragment of a water molecule, then one of the four potential neighbour sites is occupied by the saccharide itself, so that it can effectively make a maximum of three

Hydroxyl	To Water	Bridging	Internal
1 O1	0.17 (0.09)	0.18 (0.07)	0.00 (0.08)
1 OH2	2.13 (0.71)	0.65 (0.18)	0.23 (0.07)
1 OH3	1.83 (0.48)	0.40 (0.18)	0.38 (0.07)
1 OH4	1.83 (0.57)	0.27 (0.17)	0.16 (0.06)
1 O5	0.39 (0.16)	0.11 (0.07)	0.02 (0.09)
1 OH6	2.16 (0.62)	0.14 (0.11)	0.02 (0.08)
2 OH1	2.04 (0.51)	0.37 (0.09)	0.10 (0.07)
2 OH2	1.61 (0.46)	0.28 (0.16)	0.22 (0.06)
2 OH3	1.75 (0.46)	0.35 (0.18)	0.24 (0.06)
2 OH4	1.58 (0.63)	0.21 (0.14)	0.12 (0.06)
2 O5	0.65 (0.25)	0.56 (0.13)	0.00 (0.00)

Table 5.6: Hydrogen Bonding Statistics for Well B in solution numbers in brackets refer to average lifetimes in ps.

Hydroxyl	To Water	Bridging	Internal
1 O1	0.10 (0.16)	0.02 (0.07)	0.01 (0.09)
1 OH2	1.47 (0.83)	0.76 (0.18)	0.29 (0.10)
1 OH3	1.28 (0.44)	0.13 (0.11)	0.39 (0.08)
1 OH4	0.87 (0.43)	0.10 (0.12)	0.12 (0.07)
1 O5	0.34 (0.15)	0.13 (0.08)	0.01 (0.08)
1 OH6	1.81 (0.49)	0.15 (0.14)	0.01 (0.06)
2 OH1	2.27 (0.53)	0.87 (0.21)	0.11 (0.07)
2 OH2	0.50 (0.31)	0.13 (0.14)	0.30 (0.06)
2 OH3	0.13 (0.23)	0.40 (0.13)	0.29 (0.06)
2 OH4	0.14 (0.18)	0.05 (0.12)	0.10 (0.06)
2 O5	0.33 (0.14)	0.47 (0.12)	0.01 (0.15)

Table 5.7: Hydrogen Bonding Statistics for Well D in solution; numbers in brackets refer to average lifetimes in ps.

Hydroxyl	To Water	Bridging	Internal
1 O1	0.40 (0.16)	0.14 (0.05)	0.001 (0.05)
1 OH2	2.09 (0.54)	0.52 (0.05)	0.09 (0.05)
1 OH3	2.01 (0.44)	0.28 (0.09)	0.21 (0.06)
1 OH4	1.98 (0.86)	0.17 (0.08)	0.12 (0.06)
1 O5	0.05 (0.08)	0.02 (0.05)	0.21 (0.08)
1 OH6	2.12 (0.63)	0.03 (0.05)	0.21 (0.04)
2 OH1	2.23 (0.50)	0.49 (0.07)	0.10 (0.06)
2 OH2	1.88 (0.51)	0.18 (0.08)	0.24 (0.06)
2 OH3	1.50 (0.35)	0.19 (0.09)	0.29 (0.07)
2 OH4	1.74 (0.55)	0.13 (0.08)	0.14 (0.06)
2 O5	0.39 (0.18)	0.31 (0.06)	0.00 (0.00)

Table 5.8: Hydrogen Bonding Statistics for Well E in solution; numbers in brackets refer to average lifetimes in ps.

hydrogen bonds, two as an acceptor and one as a donor. Other studies[56] have found the total to be closer to three, but this could be due to a difference in definition; namely that an oxygen-oxygen distance of 3.4Å is used rather than the oxygen-hydrogen distance of 2.4Å. An anomaly which stands out in the total number of hydrogen bonds to water is the low number made by several of the hydroxyls in well D. There does not seem to be a simple explanation for this, apart from the fact that in conformer D the two rings lie in a "stacked" arrangement, which may possibly disrupt the normal pattern of hydrogen bonds. In general, the lifetime of the hydrogen bonds to water is several times larger than the intramolecular hydrogen bonds, presumably because the intramolecular bonds are highly strained.

Significant bridging hydrogen bonds through water are observed for all hydroxyls in all wells, averaging about 0.25. This is a phenomenon which has been previously observed[56] to arise from bridging between adjacent hydroxyls (§2.3.1). However some of the hydroxyls exhibit a stronger bridging phenomenon. In the analysis for well A, OH2 on residue 2 has an average of 0.46 bridging interactions, although it is not clear what its partner is. Well B shows strong bridging for OH2 on ring 1 and O5 on ring 2, indicating a possible interresidue bridge. Well D shows even stronger evidence of such a bridge, with an average of 0.76 bridged interactions for OH2 on residue 1 and an average of 0.87 for OH1 on residue 2. There is also some evidence of bridging hydrogen bonds in well E which has an average of 0.52 and 0.49 bridges through water to OH2 on ring 1 and OH1 on ring 2 respectively. Naturally, such bridges would be expected to stiffen

the $\alpha(1\rightarrow6)$linkage in the wells where they occur.

5.2.6 Solvent Structuring about Isomaltose

As is evident from the difference in conformational mobility between the solution and vacuum simulations of Isomaltose, the effect of the water is to increase the range of conformational space accessible to the molecule. In order to explain the plasticizing effect of the water, the solution structure of isomaltose was investigated using one dimensional pair distribution functions (PDF's, §2.3.1) and three-dimensional spatial distribution functions (SDF's, §2.3.1).

One Dimensional Pair Distribution Function (PDF)

The one-dimensional pair-distribution functions for Isomaltose were separately calculated for each section of the simulation which corresponded to a separate minimum. However, the information available from such pair distribution functions is limited, as they are radially averaged. All of the minima analysed gave almost exactly the same pattern of pair distribution functions, with each sugar hydroxyl having almost exactly the same position for its first maximum for all minima. An example from minimum A has been chosen to illustrate the general trend in figure 5.23, while the positions of the first maximum peak for each hydroxyl have been given in table 5.9.

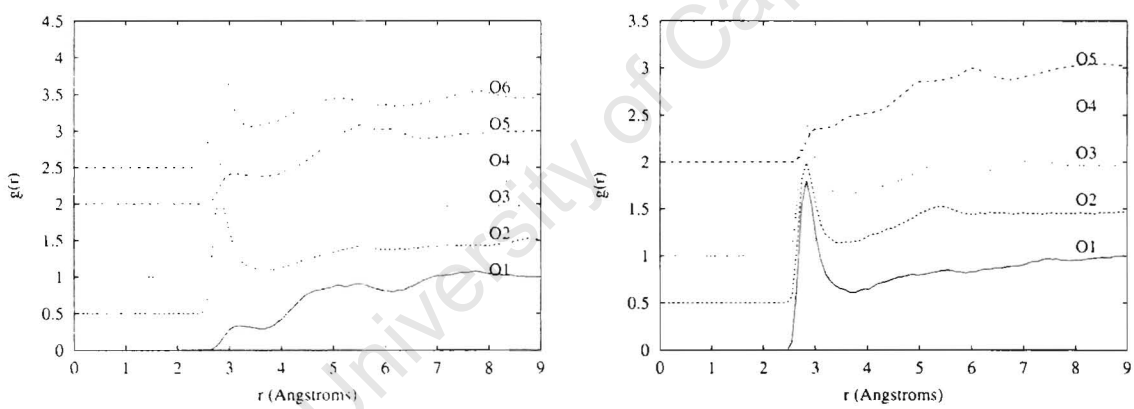


Figure 5.23: Isomaltose hydroxyl-water pair distribution functions for ring 1 (left) and ring 2 (right), offset by 0.5 per plot.

The clear trend is that all free hydroxyl groups have a first maximum peak at 2.825Å with varying height, while the ring oxygens and glycosidic oxygens structure the water much more weakly and have less well-defined peaks at greater distances (> 5Å), consistent with the findings in §5.2.5. In order to get an estimate of the "first shell" coordination number of each sugar oxygen, the pair distribution function was integrated over a sphere centered on the oxygen. The radius of the sphere was taken as 3.47Å, as this represented the first minimum in the more strongly structured pair distribu-

	Ring 1			Ring 2		
Hydroxyl	r_{max}	$g(r_{max})$	Integral	r_{max}	$g(r_{max})$	Integral
O1	7.73	1.078	1.00	2.825	1.802	3.75
O2	2.825	1.622	3.38	2.825	1.494	3.32
O3	2.825	1.567	3.55	2.825	1.420	3.06
O4	2.825	1.469	2.93	2.825	1.328	2.99
O5	5.511	1.089	1.35	8.568	1.044	1.33
O6	2.825	1.749	3.16			

Table 5.9: Radial distances of first peak in the sugar-oxygen-water pdf for Isomaltose (from dynamics in well A). Integrals of oxygen density are from 0 to 3.5Å.

tion functions. The total number of waters in the first shell obtained by this method is greater than the total number of hydrogen bonded waters obtained by the previous analysis. However, it is known that even in water, the integrated intensity of the first peak up to 3.3Å (the first minimum) is approximately 4.4,[74] while the average number of hydrogen bonds is less than 4. This has been shown in the case of water to be due to "interstitial" non-hydrogen bonded waters.[49] The total coordination number for the isomaltose oxygens follows a similar trend to the values of the first peak height, being just over 3 for most of the hydroxyls and just over 1 for the other oxygens. However, this gives limited information on where the water oxygens are actually located, rather a radial average over all the possible sites. Furthermore, no evidence of bridging hydrogen bonds may be obtained by this method.

Three Dimensional Spatial Distribution Function (SDF)

A more detailed insight into the directed nature of the water structure about Isomaltose may be obtained by calculating the three-dimensional equivalent of the PDF, the spatial distribution function. This can give direct 3-D information on the location of the time-averaged water density, in order to explain the trends observed in the hydrogen bonding analysis and to resolve the averaged peaks in the one dimensional PDF into distinct spatial positions.

Separate distributions of water were calculated for each of the minima A, B, D, and E. Since the linkage in isomaltose is rather flexible, it was necessary to select frames corresponding closely to particular conformations. A cluster corresponding to each minimum was chosen, and the structure from the trajectory corresponding most closely with the center of the cluster was extracted. The frames were selected by making a best RMS fit with the cluster center structure and choosing a cutoff RMS distance which gave a

similar number of members to the actual size of the cluster as obtained from the original cluster analysis. The clusters used for analysis were:

Well	Simulation from:	Cluster No.
A	well A	1
B	well B	1
D	well B	3
E	well A	4

Figure 5.24 to 5.27 show the water oxygen density contoured at three times above the bulk density, as well as dummy atom positions corresponding to points of peak density.

In the analysis of each cluster, it is necessary to ignore the hydroxyl hydrogens of the sugar completely, since these adopt several rotameric orientations, and the one illustrated in each case is only a single possibility. In addition, although in principle it would be possible to also contour water hydrogen density in the same fashion as the oxygen, the interpretation of this would also be ambiguous due to averaging over the various positions the water molecule might adopt (see §2.3.1). Thus only the oxygen density is available, which limits the interpretation in terms of hydrogen bonding between specific hydroxyls and specific water oxygen density peaks. It is still feasible, though, to assign possible hydrogen bonds and to make comparisons of this information with that which may be obtained from hydrogen bond analysis.

Common to all of the density distributions is the general pattern of water density around each ring. The peak of water density is generally located between the hydroxyls, supporting the contention that all of the hydroxyls participate in hydrogen bonded water bridges with their neighbours, albeit weakly. The isosurfaces are usually elongated perpendicular to the plane of the ring, with the peak densities being close the equatorial plane. The elongation can be interpreted as arising from those waters which are hydrogen bond donors, so only one of their hydrogens needs to be near the equatorial hydroxyls. The distance from these water oxygen peaks to the potentially hydrogen-bonded hydroxyl covers the full range of hydrogen bonding, from 2.4 to 3.4Å. However, the interpretation of the maximum water oxygen density is not as straightforward as a simple preference for regions in which favourable contacts may be made with the solute, such as through hydrogen bonding, as several of the water peaks are more than 3.4Å away from the nearest sugar hydroxyl. It appears therefore that the isomaltose solute has a capacity for structuring the water other than through direct hydrogen-bonded interactions. This could be either through secondary hydrogen bonding or through hydrophobic interactions.

The cluster centre representing well A in the adiabatic map has an extended conformation in which the two rings are quite distant. The potential for bridging hydrogen

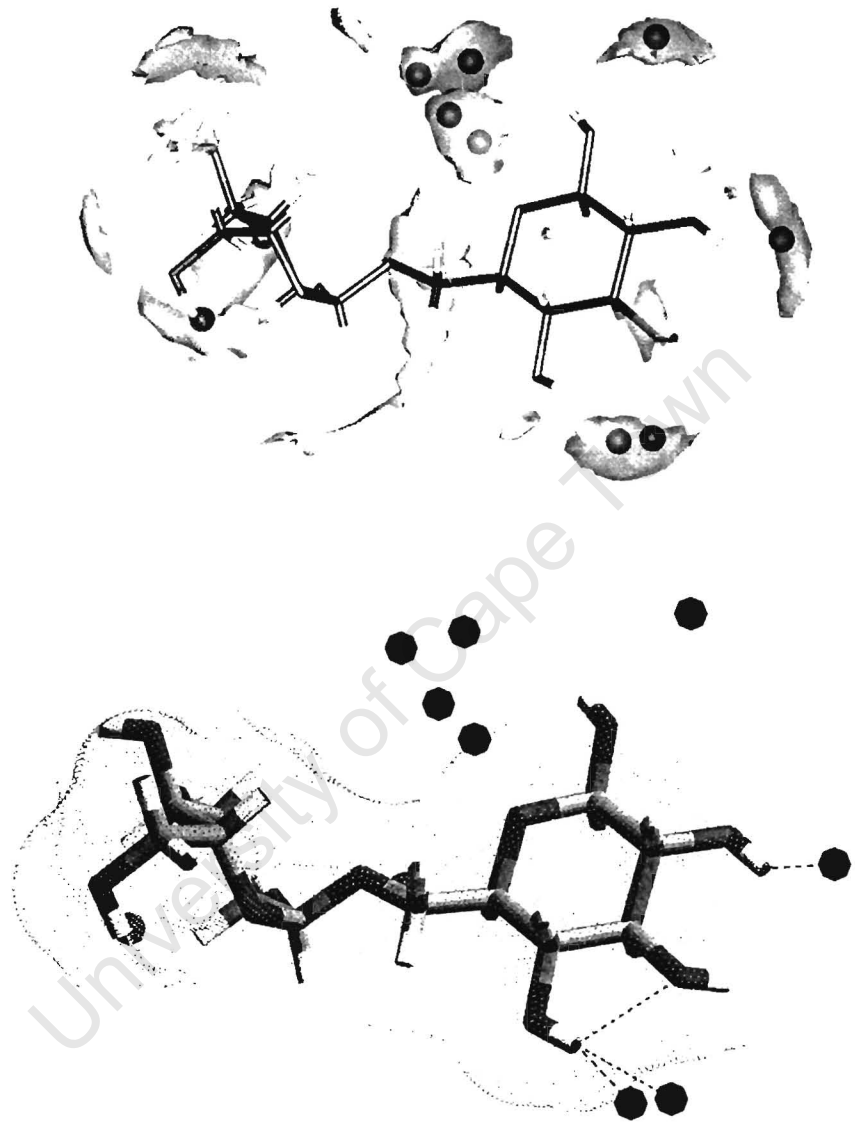


Figure 5.24: Three dimensional contour plot of water density around Isomaltose in well A at 3 times above bulk (above); Points of maximum water density around Isomaltose with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

bonds between residues is therefore reduced, as was evident in the hydrogen bond analysis. There are however several maximum density peaks lying in the slight cleft between residues, some of which are potentially bonded to one of the residues. The high density region between the rings not involved in hydrogen bonding to the solute could be caused by the disruption of water structure by the residues on either side. It could also simply be the result of the water between the rings filling the gap left by the strongly structured water peaks around the residues on either side.

The structure for well B is more compact than that of A, with the two rings lying almost parallel to each other. The bridging hydrogen bond between O2-H2 of ring 1 and O5 of ring 2, which was suggested by the hydrogen bond analysis is found to be present; the peak of oxygen density is 3.22Å from the O2 and 3.25Å away from the O5. A ridge of oxygen density similar to that observed for well A lies between the rings.

Similarly, in well D, which showed the strongest evidence of bridging hydrogen bonds in the simple hydrogen bond search, has a peak of oxygen density lying between the expected alcohols. There is a peak lying 3.27Å away from O2 of ring 1 and 2.64Å away from O1 of ring 2.

The well E structure, like that for well A, is extended, and does not favour any kind of interresidue bridging hydrogen bonds, despite some indication of quite strong bridges in the hydrogen bond search. A ridge of water density lies between the rings, as in well A.

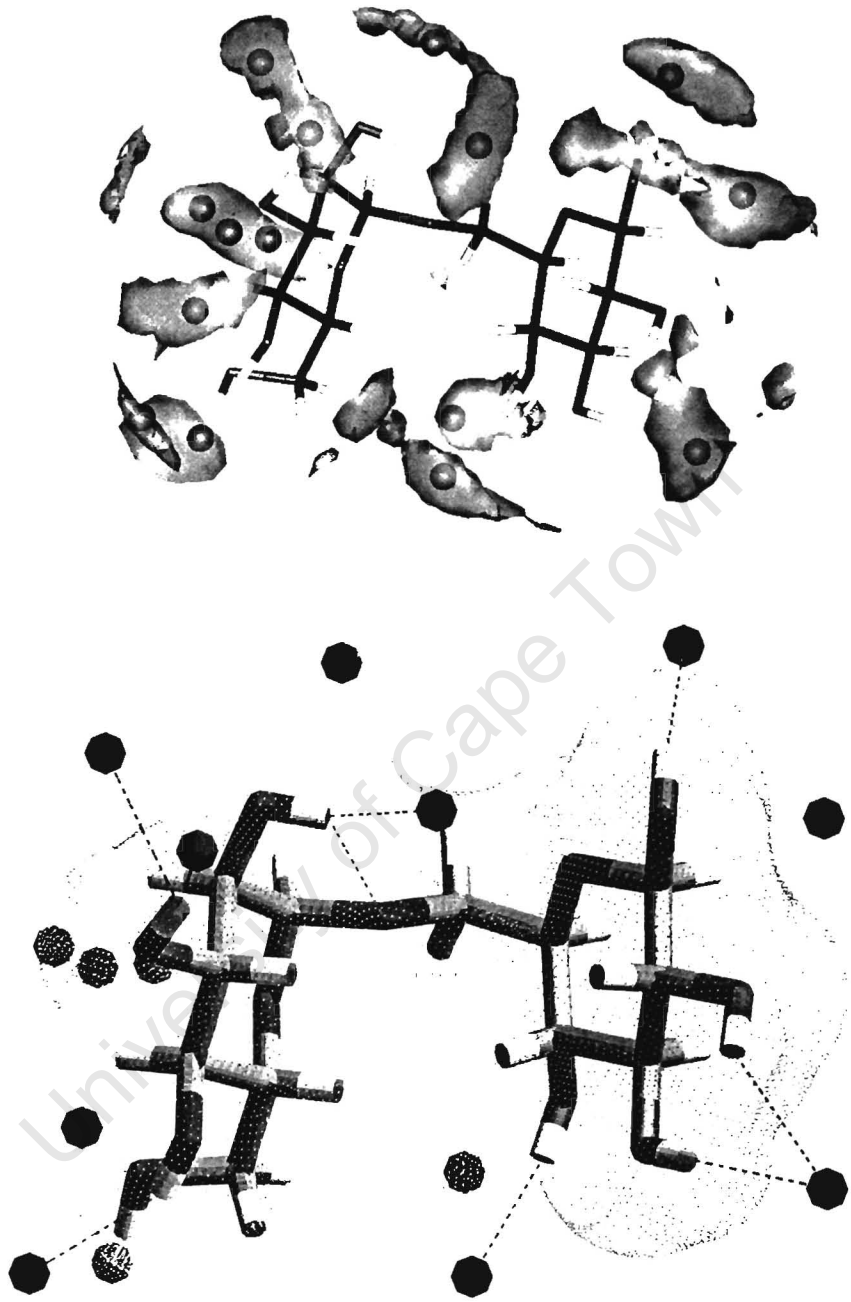


Figure 5.25: Three dimensional contour plot of water density around Isomaltose in well B at 3 times above bulk (above); Points of maximum water density around Isomaltose with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

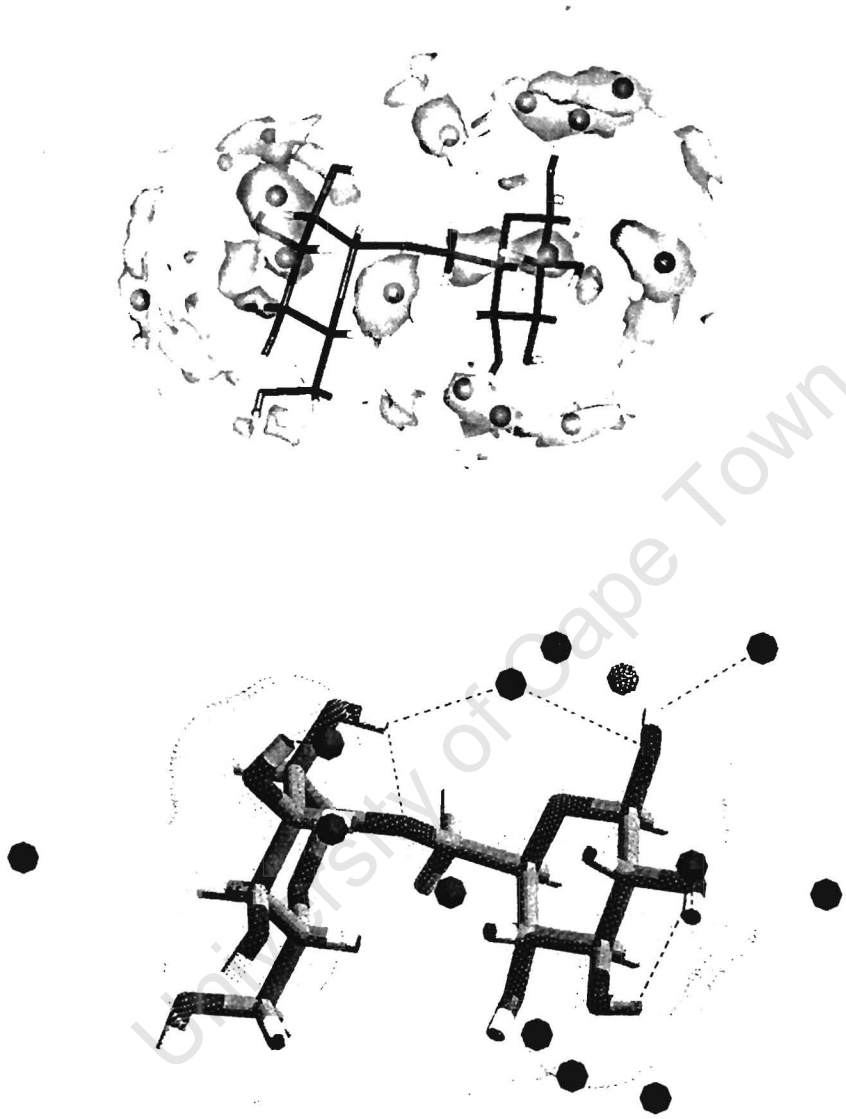


Figure 5.26: Three dimensional contour plot of water density around Isomaltose in well D at 3 times above bulk (above); Points of maximum water density around Isomaltose with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

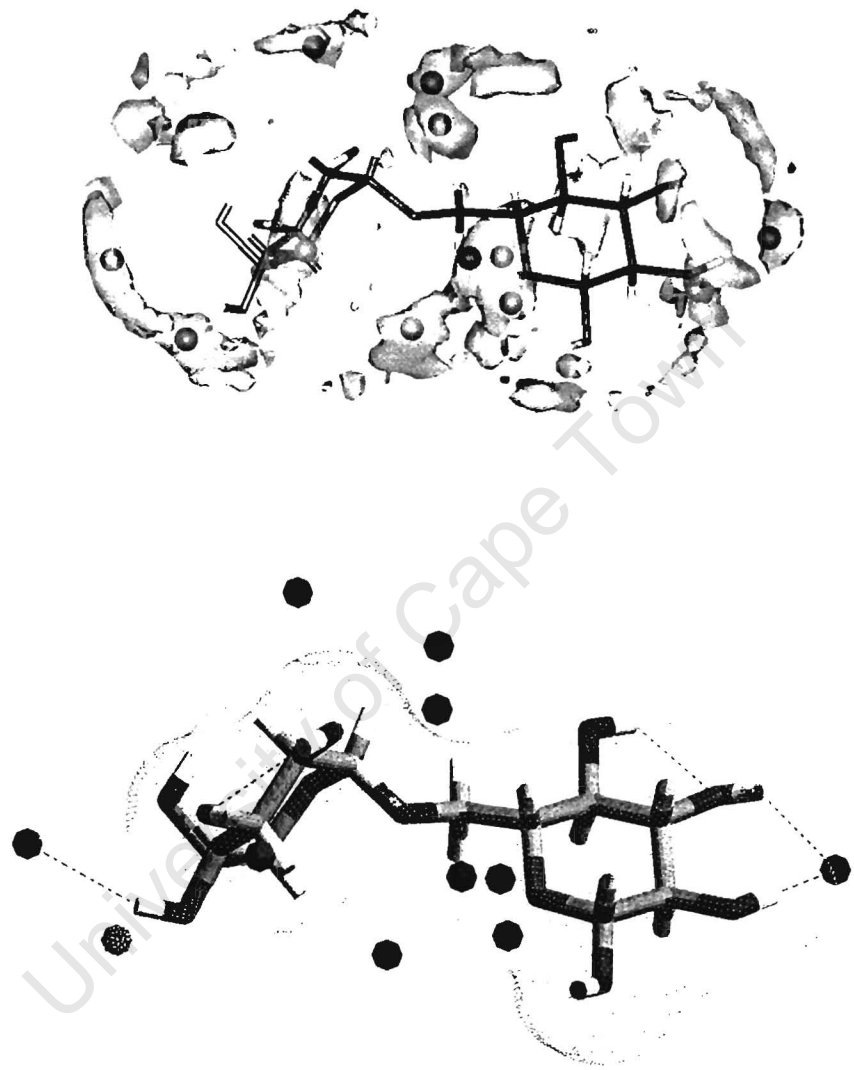


Figure 5.27: Three dimensional contour plot of water density around Isomaltose in well E at 3 times above bulk (above); Points of maximum water density around Isomaltose with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

5.3 Molecular Relaxation

5.3.1 Dihedral Angle Time Correlation Functions

Correlation plots of the fluctuations in dihedral angles $C(t) = \langle \Delta\chi(0).\Delta\chi(t) \rangle$ indicate the extent to which their motion is correlated with that at a time t away. Thus their rate of decay will give an indication of the rate of fluctuation.

The correlation plots for the dihedral angles discussed above are shown in figures 5.28 (vacuum) and 5.29 (solution). These indicate that the loss of correlation occurs much more rapidly in vacuum, with the correlation function decaying to close to zero within a few picoseconds, versus around 20 to 50 ps for the solution case. This can be interpreted as the damping effect of the water, which also removes the fluctuations observed in the correlation functions for the vacuum simulations.

In vacuum, the most rapid fluctuation (and therefore fastest loss of correlation) is found in the primary alcohol dihedral, Θ, which would be expected, given that the one bond terminus has only two attached atoms. For the other dihedrals, Ω has the fastest loss of correlation, possibly due to being confined in a very steep potential, while Ψ and Ω are similar. For the solution case, the order is similar.

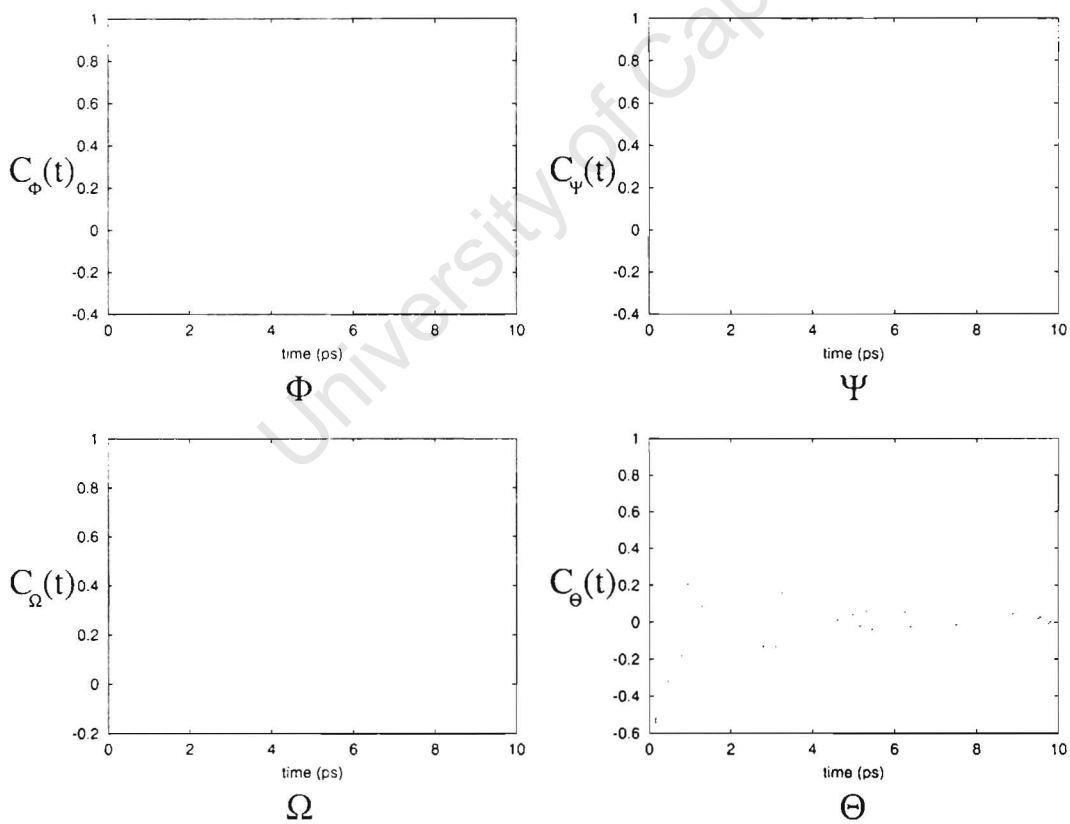


Figure 5.28: Time correlation functions for isomaltose in vacuum from minimum A

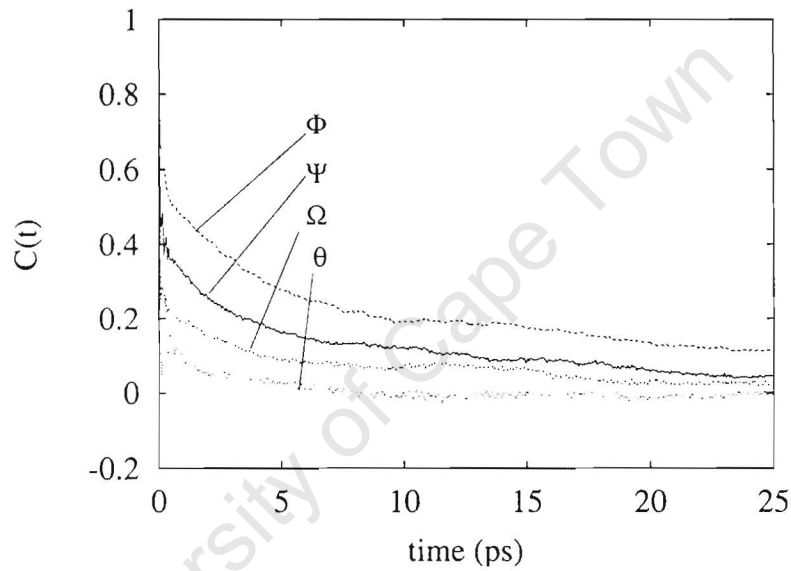


Figure 5.29: Time correlation functions for isomaltose in solution from minimum A

5.3.2 Overall Molecular Correlation Functions

An estimate of the relative contributions of internal fluctuations and overall molecular tumbling may be made using the RMS fit autocorrelation function and the P_2 autocorrelation function for a vector rigidly attached to the macromolecular backbone (§2.3.2) respectively. The RMS fit autocorrelation functions for each of the simulations are shown in figure 5.30. For each of the simulations, essentially all internal correlation has been lost over the first 50 ps.

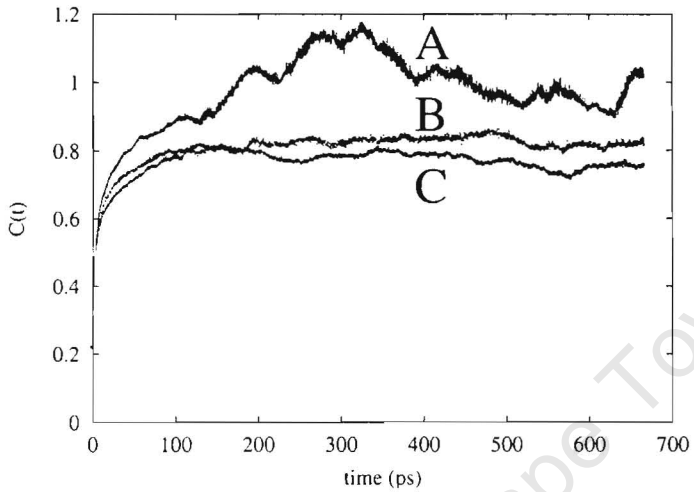


Figure 5.30: RMS Autocorrelation function for relaxation in the molecular coordinate frame.

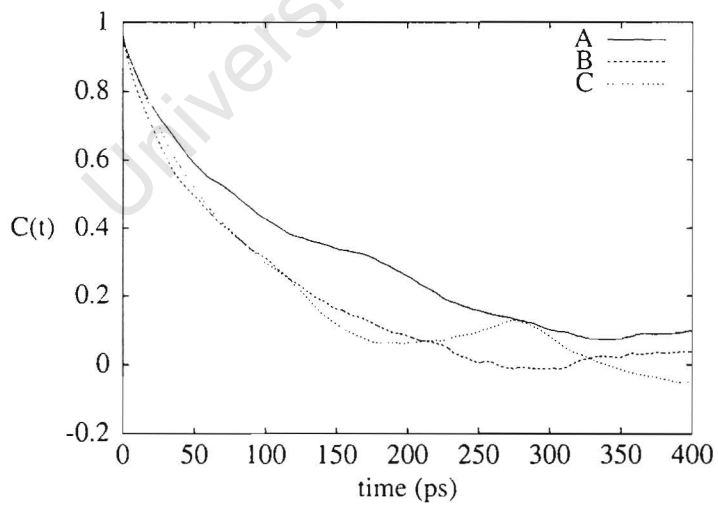


Figure 5.31: P_2 autocorrelation functions describing molecular tumbling.

The vector used for the P_2 autocorrelation describing molecular tumbling was that between C1 of residue 1 and C5 of residue 2. The decay of the overall correlation

function is also shown in figure 5.31 for each simulation. In this case, the correlation function takes about 350ps to decay to 0. Tumbling times τ_M, which are important for NMR analysis, were calculated from an exponential fit to the first 100 ps of the graph, giving values of 106 ps, 109 ps and 80 ps for the simulations from wells A, B and C respectively.

5.4 Isomaltose T_1 Relaxation Experiments

For comparison with the simulation, relaxation data at several frequencies were collected for isomaltose in D_2O. The ^{13}C assignment of isomaltose was taken from published work,[101] the only change being the reversal of the reducing terminus C1 assignment. Table 5.10 shows these assignments. All of the peaks on the reducing residue and some on the non-reducing residue exhibited splitting due to the anomeric effect. The peaks arising from the α anomer are indicated as such in table 5.10, while the rest are from the β anomer or from peaks which are not split. Note that anomeric splittings which could be observed at 300 and 400 MHz were not visible at 200 MHz, while certain other peaks which were resolved at higher frequencies appeared overlapped at 200 MHz.

Ring	Carbon	δ (ppm)	T_1 (200 MHz)	T_1 (300 MHz)	T_1 (400 MHz)
1	C1	98.76	0.47(1)	0.59(2)	0.66(1)
	C2	72.17	0.41(1)	0.50(4)	0.55(2)
	C3	73.75		0.62(3)	0.60(1)
	C4	70.18		0.54(2)	0.58(1)
	C5	72.43	0.43(1)	0.56(3)	0.58(1)
	C6	61.09	0.27(1)	0.35(1)	0.37(1)
	C1(α)	98.79		0.61(2)	0.64(1)
	C2(α)	72.13		0.56(2)	0.57(2)
	C5(α)	72.47		0.59(4)	0.57(2)
2	C1	96.86	0.49(2)	0.64(3)	0.68(1)
	C2	74.74	0.52(1)	0.67(3)	0.65(1)
	C3	76.64	0.47(2)	0.65(3)	0.67(3)
	C4	70.08	0.47(1)	0.62(2)	0.64(2)
	C5	74.98	0.45(1)	0.58(3)	0.61(2)
	C6	66.37	0.24(1)	0.32(2)	0.34(1)
	C1(α)	92.97	0.64(2)	0.64(4)	0.68(2)
	C2(α)	72.08		0.61(2)	0.66(1)
	C3(α)	73.69		0.62(3)	0.64(1)
	C4(α)	70.23		0.62(4)	0.63(3)
	C5(α)	70.69	0.47(3)	0.62(4)	0.60(1)
	C6(α)	66.45		0.38(2)	0.33(2)

Table 5.10: ^{13}C Experimental data for isomaltose: chemical shifts[101] and T_1 measurements (in seconds) at 200, 300 and 400 MHz.

A standard inversion-recovery sequence was used to obtain T_1 first order relaxation

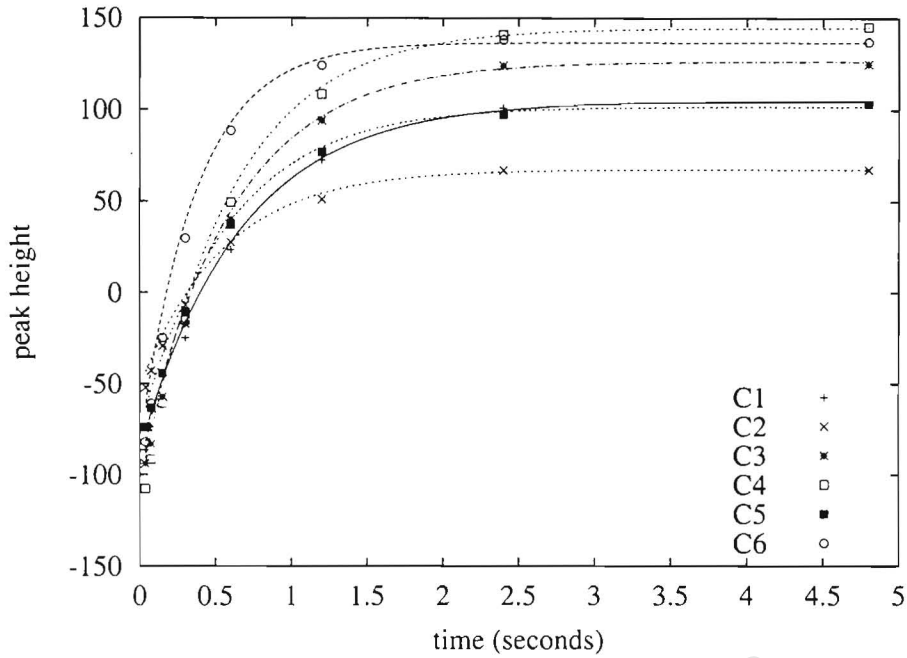


Figure 5.32: Exponential decay plots for isomaltose non-reducing residue at 400MHz.

constants for all the carbons in isomaltose. The data were collected with continuous decoupling during preparation and acquisition, to prevent cross-relaxation effects with the ^1H nuclei in the molecule. An example of exponential decay curves (for the non-reducing residue at 400 MHz) is given in figure 5.32, and a single peak decay is shown in figure 5.33. These demonstrate a good exponential fit to the data. The experiment was run at 200, 300 and 400 MHz to provide data for use in the calculation of model-free generalized order parameters, S^2; the data for all frequencies are given in table 5.10.

Figure 5.33: Decay of C6 peak on non-reducing residue; from left to right, peaks are at 0.0375, 0.075, 0.15, 0.3, 0.6, 1.2, 2.4, 4.8 seconds.

5.5 Comparison of Relaxation Parameters from Experiment and Simulation

Longitudinal relaxation times, T_1 were calculated from each of the three solution simulations using an analogous procedure to that adopted for maltose. Rotational correlation times τ_M were calculated by a least squares fit of an exponential decay in section 5.3.2. Using the values of τ_M estimated by this procedure, T_1 relaxation times at 400 MHz were estimated from each simulation. These are compared with the experimental values in table 5.11.

Ring	Carbon	T_1 (A)	T_1 (B)	T_1 (C)	T_1 (exp)
1	C1	0.55	0.52	0.68	0.66(1)
	C2	0.59	0.56	0.73	0.55(2)
	C3	0.61	0.61	0.74	0.60(1)
	C4	0.61	0.57	0.74	0.58(1)
	C5	0.60	0.59	0.72	0.58(1)
	C6	0.30	0.28	0.39	0.37(1)
2	C1	0.72	0.53	0.80	0.68(1)
	C2	0.70	0.53	0.79	0.65(1)
	C3	0.69	0.49	0.79	0.67(3)
	C4	0.69	0.32	0.81	0.64(2)
	C5	0.67	0.19	0.77	0.61(2)
	C6	0.33	0.36	0.38	0.34(1)

Table 5.11: Longitudinal relaxation times (T_1), in seconds, as calculated from the isomaltose simulations and as measured by experiment at 400 MHz.

The relaxation times in table 5.11 are generally similar for both rings in both the simulated and experimental data (unlike the maltose case, in which the reducing residue had higher T_1's). Within each ring, the T_1 relaxation times are very similar (as was the case for maltose), indicating that the main contribution to the motion of each C-H vector is in fact the motion of the ring to which it is attached. Each simulation gives a slightly different trend, as a result of the limited amount of dynamics used; this justifies the use of several simulations, rather than just one, since a more complete representation of the dynamical behaviour is thus obtained. The agreement with experiment is mostly very good, with a few simulations exhibiting anomalous results.

Generalised order parameters for isomaltose were extracted from the experimental data using the program *MODELFREE*[95,96] under the assumption that $t_m = 0.2ns$ and

$t_e = 0$ for all carbons. This assumption is required in order to get a sensible fit from the data; however, the caution expressed in §4.7 applies here as well. Order parameters were obtained from the simulations by direct calculation (§3.5). The order parameters from the simulations are similar to the experimentally derived ones; however, in light of the assumptions made in the fit, the best validation of the simulation comes from the agreement in the T1 values. This is a direct comparison with experimentally measured parameters and does not require a model or any unjustified assumptions.

Ring	Carbon	S^2 (S1)	S^2 (S2)	S^2 (S3)	S^2 (exp)
1	C1	0.81	0.87	0.85	0.83(1)
	C2	0.77	0.84	0.81	0.92(3)
	C3	0.76	0.83	0.80	0.82(1)
	C4	0.76	0.84	0.80	0.85(1)
	C5	0.77	0.85	0.82	0.95(1)
	C6	0.81	0.84	0.76	0.73(1)
2	C1	0.64	0.87	0.74	0.76(1)
	C2	0.63	0.86	0.73	0.79(1)
	C3	0.62	0.88	0.73	0.81(2)
	C4	0.63	0.89	0.74	0.90(1)
	C5	0.65	0.90	0.76	0.93(2)
	C6	0.67	0.85	0.67	0.82(1)

Table 5.12: Generalized order parameters (S^2) calculated from each of the isomaltose simulations and from experimental T_1 values.

Chapter 6

Modelling an Amylopectin branch point

Maltose and isomaltose are good models for linear polysaccharides containing $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkages, such as amylose ($\alpha(1\rightarrow4)$) and pullulan ($\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$). However, branched polymers such as amylopectin and glycogen require a branch model which includes the full branch environment, including residues on both sides of the branch point. Polymer branching generally has the effect of increasing the plasticity of the material and lowering the glass transition temperature. In the case of carbohydrates, where $\alpha(1\rightarrow4)$ linked monomers prefer to form quite rigid helices which strongly structure water, $\alpha(1\rightarrow6)$ branch points have the potential to add flexibility to the polymer, and possibly to disrupt the helical repeats. In order to study the $\alpha(1\rightarrow6)$ branch points of amylopectin, trisaccharide and tetrasaccharide model compounds were simulated using molecular dynamics. Using this approach, it is possible to comment on the relative orientation and motions of the glucopyranosyl units on each side of the branch, which cannot be done with simple disaccharides. This information can be used to postulate the function of these branch points within a highly branched polysaccharide, such as amylopectin or glycogen.

6.1 Trisaccharide Model: Panose

α-Panose (α-D-glucopyranosyl-(1→6)-α-D-glucopyranosyl-(1→4)-D-glucopyranose) is the simplest model compound representing both the $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkages, in a linear sequence (see figure 6.1 for an illustration of the ring and dihedral naming scheme). It also has the distinction of being the only $\alpha(1\rightarrow6)$ linked oligosaccharide for which a crystal structure is available,[12,13] and being readily available, is convenient for investigation using NMR.

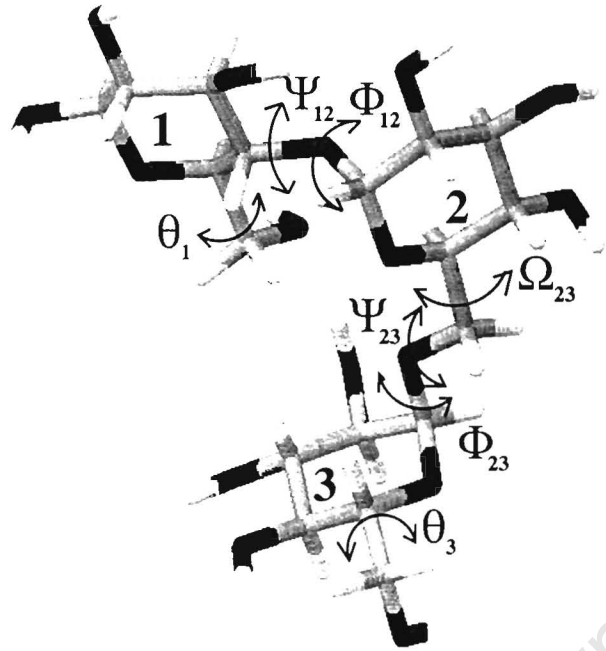


Figure 6.1: Residue numbering and torsion angle name conventions for panose.

The panose crystal structure has several features which will be of interest in comparison with solution simulations. The molecule adopts a folded conformation in which the terminal residues are close. The hydrogen bond observed across the $\alpha(1\rightarrow4)$ linkage in maltose is present between O2 of residue 2 and O3 of residue 1. There is a further intramolecular hydrogen bond in the trisaccharide between O6 of residue 1 and O2 of residue 3, which acts as a bridge in the folded conformation. This hydrogen bond is especially interesting, since it distorts the orientation of the primary alcohol of the reducing residue to an eclipsed rotamer in the crystal structure. Such an orientation is high in energy, so the bridging O2-O6 hydrogen bond must provide significant compensation. The remaining hydroxyl groups are involved in hydrogen bonds with neighbouring panose molecules in the crystal, and thus did not form the "crown" of hydrogen bonds in each ring observed in vacuum. The glycosidic dihedral angles lie within the broad low energy region of the maltose adiabatic map in the case of the $\alpha(1\rightarrow4)$ linkage and within well A of the isomaltose adiabatic map in the case of the $\alpha(1\rightarrow6)$ linkage. The conventions for dihedral angles used for panose in this chapter are given in table 6.1, and follow the same pattern as that used for maltose and isomaltose. Thus $\{\Phi_{23}, \Psi_{23}, \Omega_{23}\}$ describe the $\alpha(1\rightarrow6)$ glycosidic linkage, $\{\Phi_{12}, \Psi_{12}\}$ the $\alpha(1\rightarrow4)$ join and $\{\Theta_1, \Theta_3\}$ the orientation of the primary alcohols.

Φ_{23} :3O5-3C1-3O1-2C6

Ψ_{23} :3C1-3O1-2C6-2C5

Ω_{23} :3O1-2C6-2C5-2C4

Φ_{12} :2H1-2C1-2O1-1C4

Ψ_{12} :2C1-2O1-1C4-1H4

Θ_1 : 1C4-1C5-1C6-1O6

Θ_3 : 3C4-3C5-3C6-3O6

Table 6.1: Dihedral angle naming conventions used for panose.

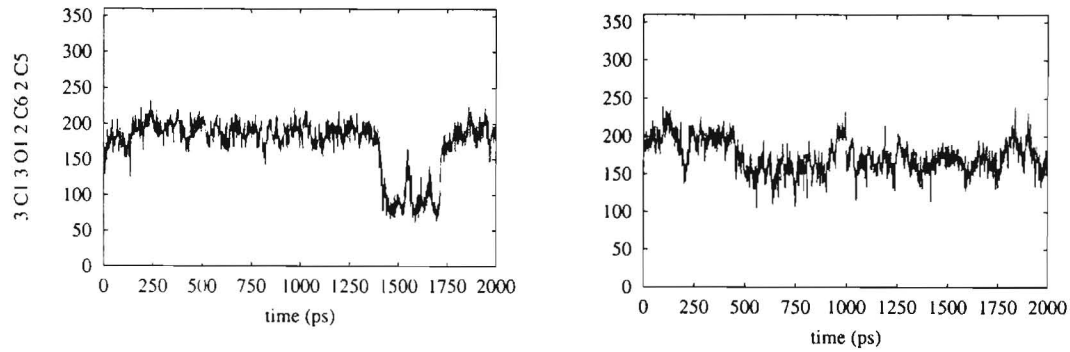
6.1.1 Molecular Dynamics Simulations

Since the difference between vacuum and solution conditions has already been established for the glucose dimers maltose and isomaltose, it was decided to only investigate the solution behaviour for the three and four residue branch models, using the same protocol as for the previous simulations (§4, 5). Two simulations were run for panose, starting from different conformations, to ensure a good coverage of conformational space. The first simulation was started from the crystal structure, and solvated in a truncated octahedral box of 815 TIP3P water molecules. The second was started from a conformation in which the $\alpha(1\rightarrow4)$ linkages were set to the larger well in the maltose vacuum map ($(\Phi, \Psi) = (-50, -40)$), the $\alpha(1\rightarrow6)$ linkage was set to the C well in the isomaltose vacuum map ($(\Phi, \Psi, \Omega) = (80, 240, 200)$) and the primary alcohols were set to the *gg* rotamer and solvated in a box of 818 TIP3P water molecules.

Each simulation was equilibrated and run for 2 ns in the microcanonical ensemble under the same conditions as the maltose and isomaltose simulations.

6.1.2 Simulation Events

The time series for both panose runs are shown in appendix B in figures B.1 and B.2 for the runs from the crystal structure and the starting structure constructed from the vacuum maps respectively. The $\alpha(1\rightarrow4)$ linkage in panose behaves similarly to the isolated one in maltose, showing frequent transitions between two populations in the third quadrant of the maltose adiabatic map, without any obvious coupling to the transitions of any of the other linkage dihedrals. The $\alpha(1\rightarrow6)$ glycosidic dihedrals also conform to the general pattern of the isolated isomaltose molecule. The Φ_{23} and Ω_{23} dihedrals make no significant transitions, while the Ψ_{23} dihedral makes a transition from the A to the E well in the run from the crystal structure, as well as several smaller transitions in the second run (figure 6.2). Ψ covers a range of approximately 150° in all simulations; this flexibility is what make the Ψ dihedral the principal degree of



Ψ_{23}, I Ψ_{23}, II

Figure 6.2: Ψ_{23} dihedral time series from first (I) and second (II) panose simulations.

conformational freedom. In retrospect, the choice of the C well as the starting point for the $\alpha(1\rightarrow6)$ linkage in the second simulation was not very good, since the isomaltose solution simulation showed that in water the A and C wells are essentially the same, since the vacuum hydrogen bonds that stabilised the C well are not present in solution.

In all the panose simulations, the primary alcohol remained in the *gg* rotamer, a known feature of the force-field.

6.1.3 Cluster Analysis

A cluster analysis was done for panose using all defined linkage and primary alcohol dihedrals. The cluster parameters from cluster analyses of each panose simulation are shown in table 6.2. The data do not cluster neatly into distinct groups, instead forming many small clusters. This is an indication that the molecule cannot be described as adopting single discrete conformations. Consequently, only the most populated clusters are shown in table 6.2. Despite the large number of groupings, clusters P1 and P2 are almost identical, as are P3 and P4, so, of these, only clusters P1 and P3 will be considered further. Cluster P5 shows a significant difference, namely $\Psi_{23} = 88.7^o$ in the $\alpha(1\rightarrow6)$ linkage and the Φ_{12} and Ψ_{12} dihedrals of the $\alpha(1\rightarrow4)$-linkage are -45.5^o and -42.7^o respectively. This corresponds to the E well of the isomaltose map, unlike all the other clusters, which lie in the large A well, while the $\alpha(1\rightarrow4)$ dihedrals for this cluster correspond to the A well of the maltose map, which is known not to form the O2 - O3 hydrogen bond. This is in contrast to the other clusters, lying closer to the B well, which would be expected to form this bond.

Much better grouping of structures was obtained from the cluster analysis of the second panose simulation. Most of the structures fell easily within two large clusters, appearing at first like clusters P1 and P3 of the first simulation. However, the dihedral

sim.	cluster	# members	std. dev.	Φ_{23}	Ψ_{23}	Ω_{23}	Φ_{12}	Ψ_{12}
I	P1	12999	0.106e+02	71.7	194.9	193.4	-28.0	-27.4
	P2	8075	0.102e+02	71.6	195.0	193.4	-31.8	-27.8
	P3	5799	0.109e+02	70.3	168.7	193.9	-30.1	-26.6
	P4	3840	0.103e+02	69.2	168.7	193.9	-32.8	-28.5
	P5	3128	0.113e+02	67.4	88.7	190.4	-45.5	-42.7
II	P6	20319	0.978e+01	65.9	159.3	191.4	-33.7	-29.1
	P7	12471	0.123e+02	72.7	197.5	194.8	-32.7	-29.4

Table 6.2: Results of clustering for two panose simulations (I: crystal structure). For each cluster center, only the linkage dihedrals are shown.

Θ_1 is in the *gg* rotamer in the second simulation, having occupied the *gt* rotamer in the first. For this reason these two clusters will be distinguished from the others in further analysis.

The structures closest to the important cluster centres for both simulations are shown in figure 6.3. The central ring has been held approximately constant in these diagrams, revealing the rotation of the terminal groups. The formation of the O2 - O3 hydrogen bond in cluster P1 causes a more linear alignment of the maltose fragment as would be the case in amylose.[53] This hydrogen bond is not seen in any of the other illustrations, although the $\alpha(1\rightarrow4)$ linkage is in a suitable conformation in clusters P3, P6 and P7. As observed above, cluster P5 is considerably different, with its $\alpha(1\rightarrow4)$ linkage in the maltose A well and its $\alpha(1\rightarrow6)$ linkage in the isomaltose E well. Ring 3 is in a quite different orientation in cluster P5, as a result of the large change in Ψ. Cluster P6, in the second simulation, is the only one depicted which has the potential for making the hydrogen bond found in the panose crystal structure, despite the fact that an alternative hydrogen bond forms between O6 of residue 1 and O1 of residue 3 in the structure chosen. The absence of any such interactions in the clusters from the first simulation is due to the *gt* orientation of the primary alcohol which geometrically distorts the hydrogen bond found in the crystal structure.

In the overall context of the polysaccharide branch being modelled, one would like to predict the effect of adding more residues to each chain terminus. In the case of clusters P1, P2, P4 and P5, the branch has folded back on itself, so addition of further monomers to the free ends would result in antiparallel strands, while in cluster P3, the strands would branch at right angles.

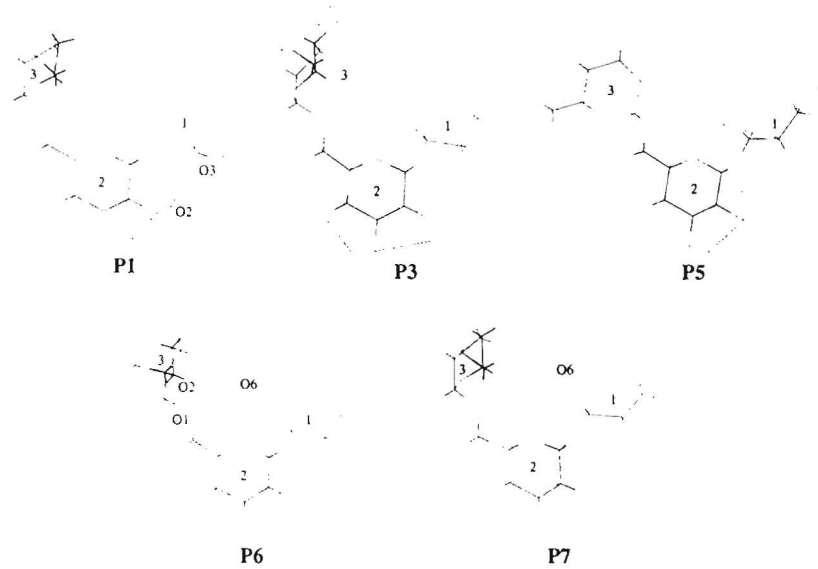


Figure 6.3: Structures closest to the cluster centres obtained from the panose simulation, starting from the crystal structure.

6.1.4 Hydrogen Bond Analysis

The foregoing discussion of the geometry of the cluster centre structures emphasizes the importance of hydrogen bonding in the conformations of panose. In addition, analysis of hydrogen bonding patterns in the dimer models maltose and isomaltose showed the importance of hydrogen bonding in solvation. In particular, hydrogen bonds were found not only to influence the individual secondary alcohols on each ring, but both direct and bridging interresidue hydrogen bonds were found to exist between rings, stabilizing conformations.

A complete hydrogen bonding analysis similar to that done for the dimer models was done for panose, by selecting sections of the trajectory data representative of certain clusters. Clearly, for flexible molecules such as these, it is not possible to choose any reasonable period of time during which only one cluster is predominantly visited, given that there are so many clusters, and that transitions between them are quite rapid. However, it is at least possible to distinguish the clusters for which there are large differences, in, for example, the $\alpha(1\rightarrow6)$ linkage.

Only the trajectory initialized from the crystal structure of panose was used. The section from 500 to 1000 ps was chosen to represent clusters P1 to P4 and that from 1450 to 1650 ps to represent cluster P5. Averaging over several clusters like this has the possible drawback of missing some subtle differences between the clusters; however, given that panose changes cluster very frequently, it may well not be valid to analyze each section separately. The results are given in appendix C. The overall trend is the

same as before, with regard to the total number of hydrogen bonds to water, to the solute and bridges to the solute. A strong intramolecular hydrogen bond is observed between 1 O3 and 2 O2 in the section corresponding to clusters P1-4, which is expected from the cluster centre geometry. In addition, a weak bridging interaction is observed for 1 O6 and 3 O2, the hydroxyls which are directly hydrogen bonded in the crystal structure. For the section corresponding to cluster 5, there are no notable direct or bridged hydrogen bonding interactions with the rest of the molecule.

The limitation of analysing hydrogen bonds over certain segments of the trajectory for a flexible molecule has already been pointed out, and for this reason an alternative method of analysis was also employed. The direct observation of oxygen-oxygen distances for potentially hydrogen bonded pairs, such as those seen in the maltose and panose crystal structures provides information on the duration and dynamics of hydrogen bonding, as well as correlation between the formation of various hydrogen bonds. The deficiency of this method is that it does not give any direct evidence of bridging hydrogen bonds.

Time series of the distances between potential hydrogen bonded pairs for panose are given in figure 6.4. The hydrogen bond partners monitored are those observed in the crystal structure of panose. The plots are shaded according to which cluster was predominant during each period of the trajectory. There is no clear relation between hydrogen bonding and cluster identity in the first panose simulation ((a) in figure 6.4). This means that in this case, the molecule has sufficient flexibility to change conformation (or cluster) and not disrupt the interresidue hydrogen bonding very much. However, in the second simulation, the larger cluster, P4, corresponds closely to a hydrogen bond between O2 on residue three and O6 on residue 1, but none between O3 on residue 1 and O2 on residue 2. The smaller cluster, P5, shows exactly the opposite behaviour. Thus, not only do the hydrogen bonds in this simulation follow the cluster identity very closely, but they appear to be mutually exclusive. This could be because residue 1 needs to move toward residue 3 in order to form the O2 - O6 hydrogen bond, but in doing so must break the O3 - O2 hydrogen bond, and vice versa.

The influence of the primary alcohol orientation on the hydrogen bonds which can be formed appears to be significant. The hydrogen bond between residues 1 and 3 never forms in the first simulation in which Θ_1 is in the *gt* rotamer, but this same interaction is dominant in the second simulation, when Θ_1 is in the *gg* rotamer. Although the current rate of primary alcohol rotation in the force-field is within the bounds of experimental measurement, it would be desirable to have more frequent transitions between rotamers, since this would obviate the need for starting configurations with different primary alcohol rotamers and allow all the simulations to sample the same conformations in the

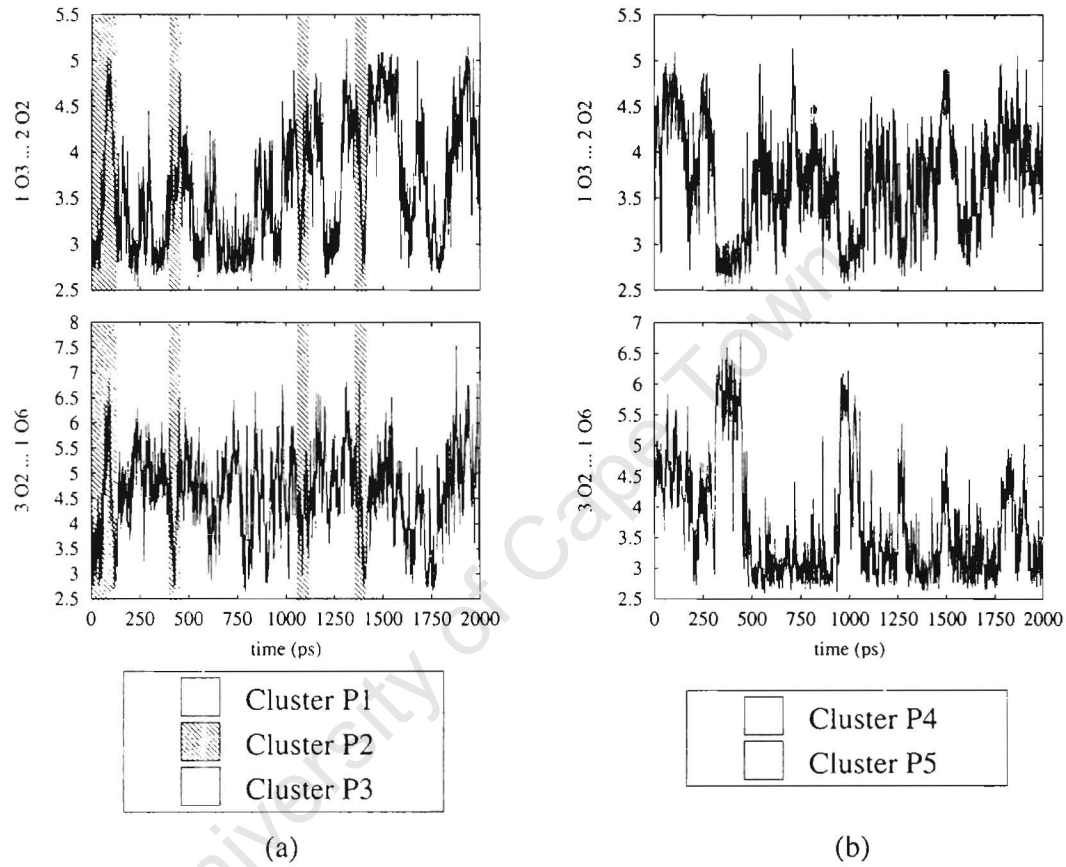


Figure 6.4: Time series of distances between potential hydrogen bonded partners for panose, simulation 1 (a) and simulation 2 (b). Shading (see key) indicates the predominant cluster at a particular time.

nanosecond regime.

6.1.5 Solvent Structure Analysis

Given the flexibility alluded to for panose above, it would require a large number of separate analyses if the solvent distribution about these molecules was described in terms of discrete conformers. Aside from the practical issues entailed in such an approach, it would be of dubious validity, since the water would not restructure rapidly enough to give a distinct distribution for each conformer, if the solute frequently changes conformation. Consequently, an approach has been adopted, in which only the conformers which show large conformational differences are separately analysed.

The clusters P1 and P5, from the first simulation and cluster P6 from the second were used for water density analysis. P1 was chosen for being the largest cluster, P5 for having the $\alpha(1\rightarrow6)$ linkage in the isomaltose E well, and P6 for being the only cluster having a direct hydrogen bond between O2 on residue 3 and O6 on residue 1. The water density distribution around these clusters is given in figures 6.5, 6.6 and 6.7.

The general pattern of water structure for all the distributions was similar to that obtained for maltose and isomaltose. Cluster P1 shows evidence of an interresidue water bridged hydrogen bond between 1 O6 and 3 O2 (tube of water density between the rings in figure 6.5); measurement of the oxygen-oxygen distances confirms that this is plausible. The illustration of cluster P5 (figure 6.6) contains a lot more noise as a consequence of the lower number of frames binned; nonetheless, it also shows some evidence of the 1 O6 - 3 O2 bridge, the water density peak in question being located further out in the cleft between rings 1 and 3 than that for cluster P1. Cluster P6, which has a direct, rather than bridged hydrogen bond between the terminal residues has very localised water density, especially in the cleft of the molecule. This is most likely the result of the rigidity of the molecule caused by the direct hydrogen bond.

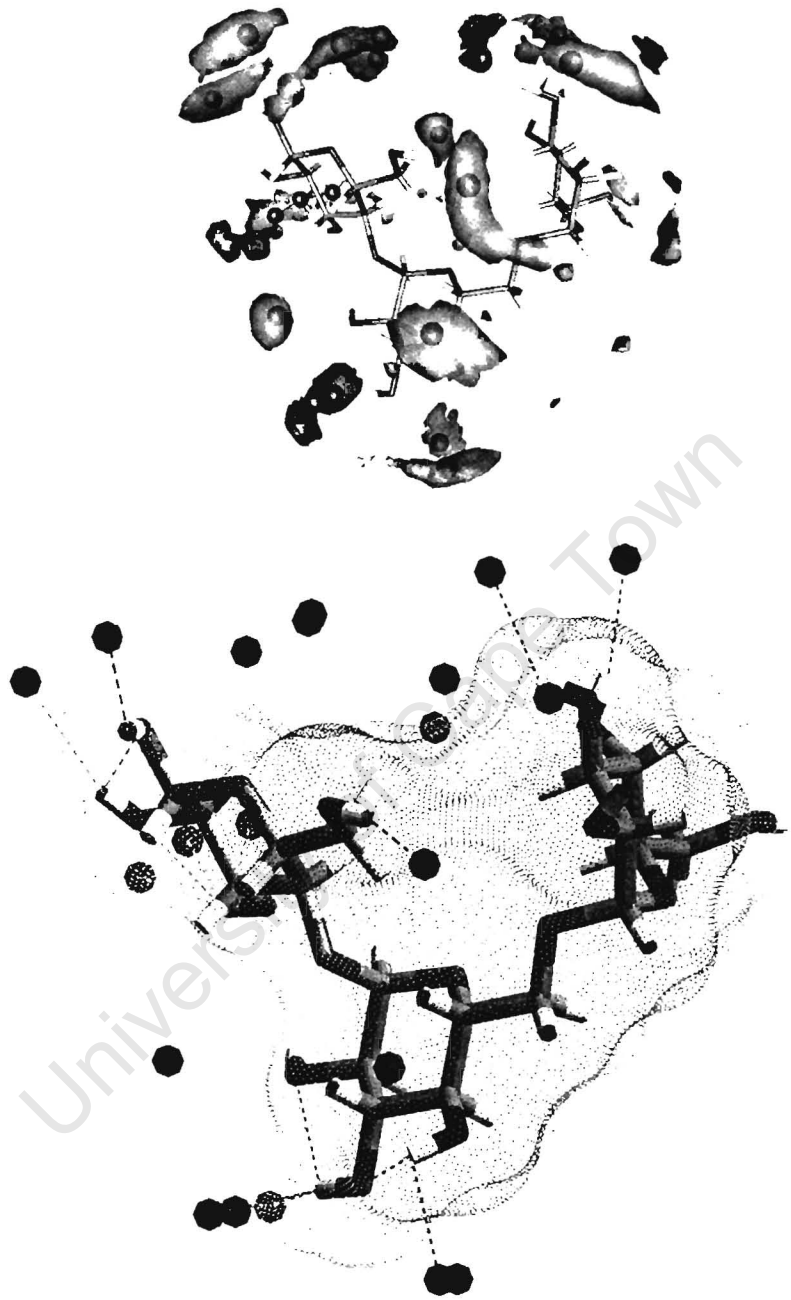


Figure 6.5: Three dimensional contour plot of water density around panose analysed using cluster P1 (above); Points of maximum water density around the cluster centre structure with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

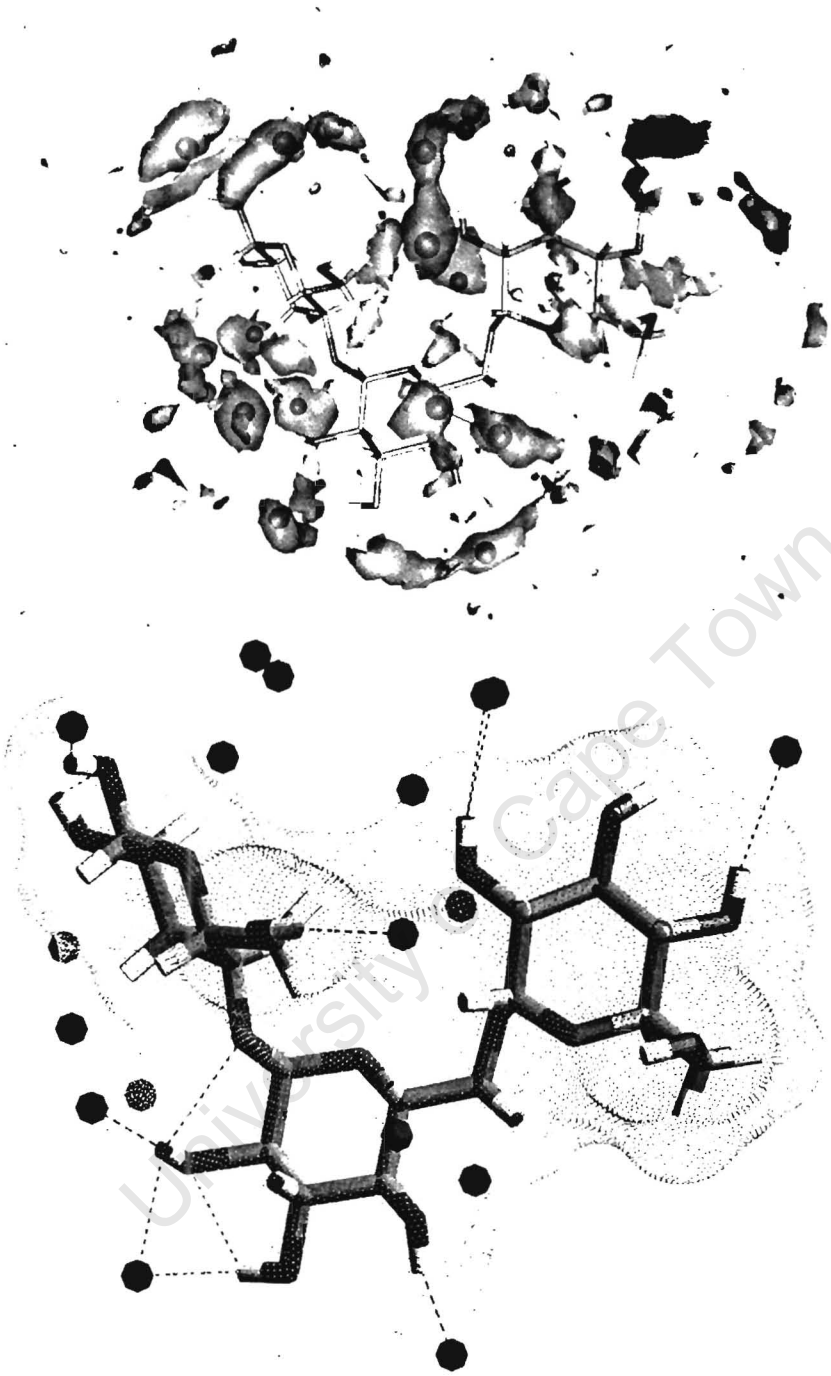


Figure 6.6: Three dimensional contour plot of water density around panose analysed using cluster P5 (above); Points of maximum water density around the cluster centre structure with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

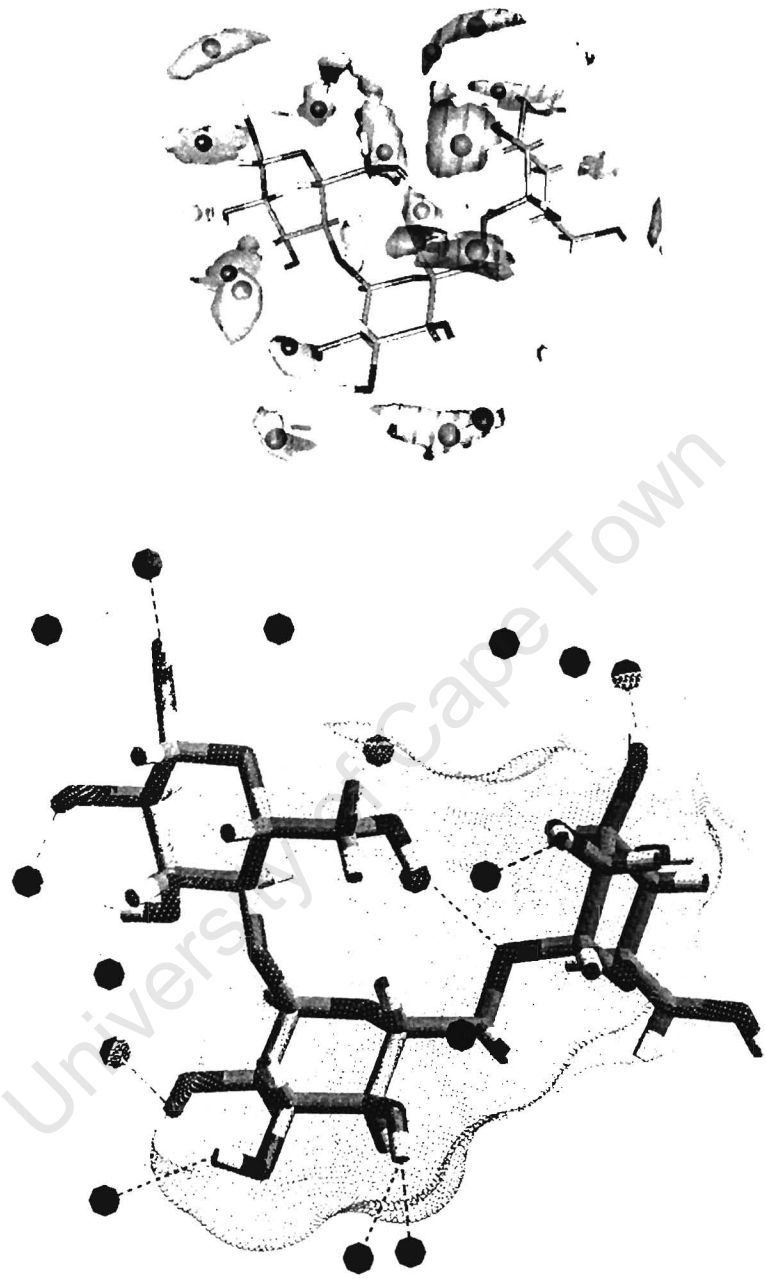


Figure 6.7: Three dimensional contour plot of water density around panose analysed using cluster P6 (above); Points of maximum water density around the cluster centre structure with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

6.1.6 NMR Relaxation parameters from Simulation and Experiment

As a check on the simulation of panose, experimental T_1 measurements were made for comparison with values calculated from simulation, in the same fashion as for isomaltose. The assignment of the panose peaks was accomplished by a set of two-dimensional NMR experiments.

Spectral Assignments

Since the carbon-13 assignments for panose have not yet been fully published, it was necessary to make an initial assignment of proton and carbon-13 NMR chemical shifts before relaxation experiments could be performed. Typically for an oligosaccharide, the proton NMR spectrum of panose has an downfield set of peaks corresponding to the H-1 resonances and a poorly resolved set of downfield peaks from the remaining protons (figure 6.8). Although standard two-dimensional methods such as COSY and TOCSY may be of use in such a situation, in practice the similarity of the peaks from the different residues makes unambiguous assignment difficult. The availability of one-dimensional versions of the traditional 2-D experiments through the use of semiselective gaussian pulses[103] allows the peaks from each residue to be separated and trivially assigned.

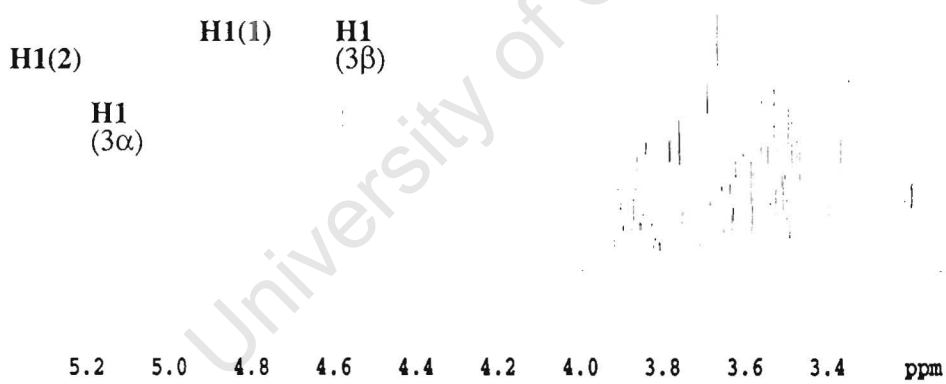


Figure 6.8: Panose ^1H Spectrum

One Dimensional TOCSY

The H-1 protons of Panose are easily assigned by peak height and coupling. The reducing residue α peak is very small while the β peak shows a stronger coupling to H-2. The remaining two peaks are assigned on the knowledge that the $\alpha(1\rightarrow4)$ linkage peaks are always upfield of those from $\alpha(1\rightarrow6)$ linkages. Having assigned the H-1's, a series of one-dimensional TOCSY experiments[104] was run to separate the ^1H peaks by residue. For

each glucosyl unit a single 1-D TOCSY was recorded with irradiation at the frequency of the H-1 proton (mixing time=500ms). Because the proton signals of the reducing terminus were all split by the anomeric effect, there were in fact four sets of data (an example is shown with assignments in figure 6.9).

Figure 6.9: Selective 1-D TOCSY for the central Panose ring

A second series of 15 tocsy experiments was then run, with the mixing time arrayed from 10ms to 150ms in increments of 10ms. All spectra were acquired on a 500MHz Varian Spectrometer at 303.1K using 32K data points and processed using the Varian VNMR software. 16 Transients were collected for each arrayed spectrum, and 82 for the single mixing time experiments. An example of such an array is given in figure 6.10.

Figure 6.10: 1-D TOCSY for central Panose ring arrayed in mixing time

From the non-arrayed spectra, all the peaks belonging to a certain residue could be identified, in most cases clearly resolved. In the arrayed spectra, the peaks appeared

in the order H-2, H-3, H-4, H-5, H-6. It was possible to unambiguously assign H-2, H-3 and H-4 directly from this information, while distinguishing H-5 from the two H-6's relied on the two dimensional heteronuclear correlation experiment (below). The ^1H assignments are given in table 6.3.

	H1	**H2**	**H3**	**H4**	**H5**	**H6**	**H6**
α-reducing terminus	5.15	3.495	3.895	3.565	3.876	3.798	3.738
β-reducing terminus	4.574	3.200	3.693	3.565	3.528	3.701	3.855
central ring	5.318	3.520	3.608	3.421	3.844	3.900	3.658
non-reducing terminus	4.878	3.485	3.667	3.353	3.636	3.775	3.690

Table 6.3: ^1H assignments for Panose (values in *ppm*)

Two Dimensional Heteronuclear Correlation Experiment

Following the assignment of the proton spectrum, the carbon 13 assignment using a heteronuclear correlation experiment was straightforward. A gradient HSQC allowed all the ^{13}C peaks to be assigned, including some which were overlapped. The full set of carbon 13 assignments are given in table 6.4.

	C1	**C2**	**C3**	**C4**	**C5**	**C6**
α-reducing terminus	92.125	71.538	73.436	77.656	70.230	60.963
β-reducing terminus	95.991	74.236	76.421	77.425	74.827	61.094
central ring	99.909	71.88	73.337	69.676	71.580	66.194
central ring (α) [a]	100.00	71.95	73.337	69.696		
non-reducing terminus	98.286	71.724	73.337	69.808	72.072	60.772

[a]Some of the central ring carbons exhibited anomeric splitting; these are the alpha peaks

Table 6.4: ^{13}C assignments for Panose (values in *ppm*)

T_1 Relaxation Measurements

A standard inversion-recovery sequence was used to obtain T_1 first order relaxation constants for all the carbons in Panose. T_1 values at 200, 300 and 400 MHz were recorded and are shown in table 6.5.

Ring	Carbon	δ (ppm)	T_1 (200 MHz)	T_1 (300 MHz)	T_1 (400 MHz)
1	C1	95.99	0.40(2)	0.52(3)	0.61(6)
	C2	74.24	0.37(2)	0.54(3)	0.57(6)
	C3	76.42	0.33(1)	0.49(3)	0.55(4)
	C4	74.83	0.32(2)	0.50(2)	0.49(4)
	C5	74.98	0.39(3)	0.45(2)	0.38(3)
	C6	61.09	0.15(1)	0.21(2)	0.21(7)
	C1(α)	92.13	0.35(3)	0.37(1)	0.46(1)
	C2(α)	71.54		0.45(2)	0.56(4)
	C3(α)	73.44	0.31(1)	0.49(3)	0.55(2)
	C4(α)	77.66	0.36(4)	0.52(2)	0.40(1)
	C5(α)	70.23	0.31(3)	0.48(4)	0.51(2)
	C6(α)	60.96	0.15(1)	0.20(2)	0.29(2)
2	C1	99.91	0.28(2)	0.37(2)	0.41(3)
	C2	71.88	0.25(1)	0.38(3)	0.38(3)
	C3	73.34			
	C4	69.68	0.24(1)	0.33(2)	0.38(4)
	C5	71.58		0.38(2)	0.36(3)
	C6	66.19	0.15(1)	0.20(1)	0.21(1)
	C1(α)	100.00	0.24(1)	0.34(1)	0.37(2)
	C2(α)	71.95	0.28(2)	0.36(2)	0.34(2)
	C3(α)	73.34			
	C4(α)	69.70			
3	C1	98.29	0.36(2)	0.48(2)	0.50(4)
	C2	71.72	0.27(1)	0.41(1)	0.45(2)
	C3	73.34			
	C4	69.81	0.30(2)	0.41(2)	0.46(3)
	C5	72.07	0.28(1)	0.40(1)	0.46(6)
	C6	60.77	0.20(1)	0.29(2)	0.21(1)

Table 6.5: ^{13}C Experimental data for panose: chemical shifts and T_1 measurements (in seconds) at 200, 300 and 400 MHz. Missing values are due to overlapped or missing peaks at the frequency in question.

Comparison of Relaxation times from Simulation and Experiment

The ^{13}C longitudinal relaxation times, T_1, were calculated from the molecular dynamics simulations by the same method as for maltose and isomaltose. The rotational constants τ_M were estimated from the decay of the P_2 autocorrelation function of the fixed vector between C1 and C4 of the central residue as before, giving values of 134 ps and 148 ps from simulations I and II respectively. Relaxation times calculated by this method are listed together with those from experiment in table 6.6. The experimental and simulated times are in excellent agreement for all the carbons, indicating a very good reproduction of the overall motional behaviour. The trends in the values within each ring are the same as observed for the disaccharide models. In addition, the terminal residues are observed to have a longer relaxation time than the central one for both the simulated and experimental data. This indicates that the central and terminal residues have different motional properties, as would be expected. Benesi and Brant have reported longer relaxation times in general for the terminal residues of oligosaccharides,[91] in agreement with our findings.

Order parameters for panose calculated from the relaxation data and from the simulation (assuming $t_M = 0.2ns$ and $t_e = 0.0ns$) are given in table 6.7. The caution regarding these assumptions expressed for maltose and isomaltose applies equally here. The order parameters were reduced in the case of the terminal carbons relative to those in the central residue for both the simulated and experimental data, supporting the observation from the T_1 relaxation times that the terminal residues have a different motion to the central ones. Moreover, this shows that the motion of the terminal residues is less spatially restricted than that of the central one.

Ring	Carbon	T_1 (I)	T_1 (II)	T_1 (exp)
1	C1	0.53	0.48	0.61(6)
	C2	0.52	0.47	0.57(6)
	C3	0.57	0.51	0.55(4)
	C4	0.52	0.47	0.49(4)
	C5	0.55	0.48	0.38(3)
	C6	0.27	0.22	0.21(7)
2	C1	0.43	0.39	0.41(3)
	C2	0.43	0.39	0.38(3)
	C3	0.46	0.42	
	C4	0.45	0.41	0.38(4)
	C5	0.45	0.41	0.36(3)
	C6	0.22	0.20	0.21(1)
3	C1	0.53	0.45	0.50(4)
	C2	0.51	0.43	0.45(2)
	C3	0.55	0.48	
	C4	0.55	0.46	0.46(3)
	C5	0.54	0.46	0.46(6)
	C6	0.26	0.23	0.21(1)

Table 6.6: Longitudinal relaxation times for panose (in seconds) as calculated from the two simulations (I (starting from crystal structure) and II) as measured from experiment at 400 MHz.

Ring	Carbon	$\mathcal{S}^2$ (I)	$\mathcal{S}^2$ (II)	$\mathcal{S}^2$ (exp)
3	C1	0.56	0.70	0.59(2)
	C2	0.58	0.74	0.72(1)
	C3	0.57	0.72	
	C4	0.59	0.74	0.69(2)
	C5	0.59	0.75	0.76(2)
	C6	0.64	0.73	0.54(1)
2	C1	0.87	0.92	0.74(3)
	C2	0.89	0.91	0.85(3)
	C3	0.88	0.91	
	C4	0.88	0.91	0.93(2)
	C5	0.90	0.93	0.80(0)
	C6	0.85	0.90	0.69(2)
1	C1	0.69	0.70	0.49(2)
	C2	0.69	0.69	0.57(2)
	C3	0.68	0.69	0.61(2)
	C4	0.72	0.72	0.57(2)
	C5	0.70	0.72	0.62(2)
	C6	0.70	0.81	0.71(2)

Table 6.7: Generalized order parameters ($\mathcal{S}^2$) calculated from each of the panose simulations and from experimental T_1 values.

6.2 Tetrasaccharide Model

Although panose includes both $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkages, it does not represent a true branch point, being essentially a linear fragment, as might be found in pullulan. For the purpose of more completely mimicking the branch in amylopectin, one more glucopyranosyl monomer was added to panose, to make a T-shaped tetramer, consisting of a central (branch point) monomer $\alpha(1\rightarrow4)$ linked at positions 1 and 4 and $\alpha(1\rightarrow6)$ linked at position 6 to the remaining three glucosyl units. Figure 6.11 illustrates the molecule (α-D-glucopyranosyl-(1→4)- [α-D-glucopyranosyl-(1→6)]-α-D-glucopyranosyl-(1→4)-D-glucopyranose), and the naming conventions applied to it, while table 6.8 gives the atoms defining the glycosidic and primary alcohol dihedrals.

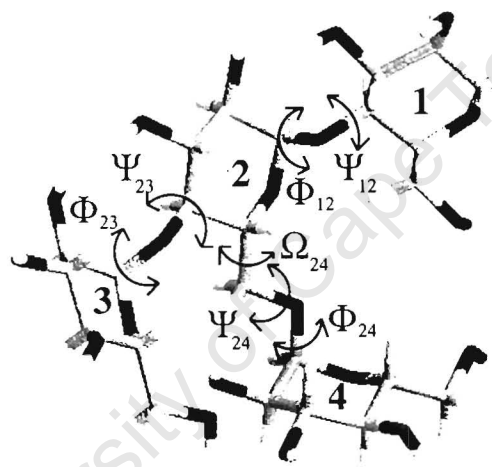
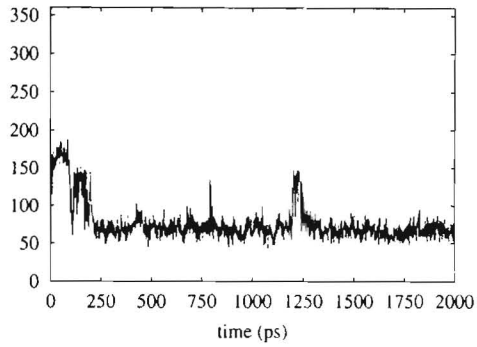


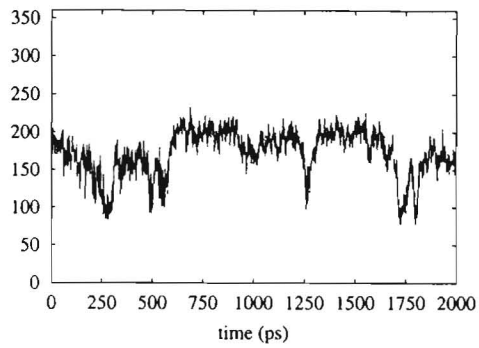
Figure 6.11: Residue numbering and torsion angle name conventions for the four residue branch model

Φ_{12} :2H1-2C1-2O1-1C4	Ψ_{24} :4C1-4O1-2C6-2C5
Ψ_{12} :2C1-2O1-1C4-1H4	Ω_{24} :4O1-2C6-2C5-2C4
Φ_{23} :3H1-3C1-3O1-2C4	Θ_1 : 1C4-1C5-1C6-1O6
Ψ_{23} :3C1-3O1-2C4-2H4	Θ_3 : 3C4-3C5-3C6-3O6
Φ_{24} :4O5-4C1-4O1-2C6	Θ_4 : 4C4-4C5-4C6-4O6

Table 6.8: Dihedral angle naming scheme for four residue model.



Φ_{24}, I



Ψ_{24}, II

Figure 6.12: Φ_{24} dihedral time series from first simulation and Ψ_{24} time series from the second simulation of the four monomer model.

6.2.1 Molecular Dynamics Simulations

Two simulations in solution were set up for the four residue model. The $\alpha(1\rightarrow4)$ linkages were initialised from the $(-50, -40)$ (A) well from the maltose map for both simulations. The two starting conformations were distinguished by the $\alpha(1\rightarrow6)$ linkage: the first was started from the A well (809 waters) and the second from the C well in the isomaltose map (811 waters). The primary alcohols were set to the *gt* rotamer in the first run, and in *gg* in the second run, since some experimental studies have shown *gt* to be almost as favoured as *gg*. Each simulation was equilibrated and run for 2 ns in the microcanonical ensemble under the same conditions as the panose simulation.

6.2.2 Dynamic Events

The four residue model showed a more interesting evolution of dihedral angles. The $\alpha(1\rightarrow4)$ linkages behaved in much the same way as for panose, however the $\alpha(1\rightarrow6)$ linkages and primary alcohols behaved differently. The first simulation, in which the $\alpha(1\rightarrow6)$ linkage of the tetrasaccharide started in the A well, underwent two transitions in Φ from 80^o to 160^o (figure 6.12. This does not correspond to any of the minima identified on the isomaltose adiabatic map, however it does lie within a low energy region corresponding to a single transition in Φ from well A, so it is not plausible. The second simulation makes a transition in Ψ similar to that made in the panose simulation from the crystal structure (figure 6.12).

Two of the primary alcohols in the first simulation of the four-residue model make transitions from the *gt* rotamer where they were initialised to the *gg* rotamer which is most favoured by the force field.[79] This is quite unusual, as previous studies[53] have not found transitions of this sort using this force-field. Graphs of the evolution of these

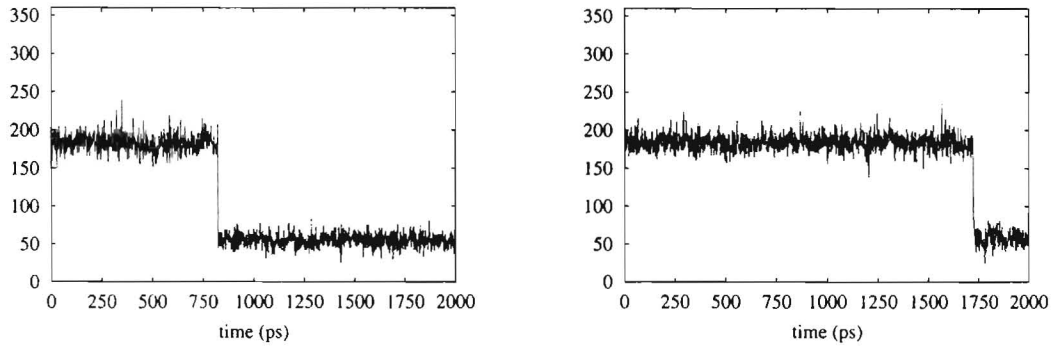


Figure 6.13: Θ_1 and Θ_4 time series from the first simulation of the four monomer model.

dihedrals are shown in figure 6.13.

6.2.3 Cluster Analysis

For the four residue model, cluster analysis data are given in table 6.9. The data fall into a large number of clusters, indicating that the branch point cannot be simply described as a small discrete set of conformers, due to its flexibility. This overlap was also evident in the cumulative number of members in each cluster (not shown), which in each case was substantially larger than the actual number of structures assigned to that cluster. Each cluster given in table 6.9 differs from the others, there being no sets of similar clusters as was the case for panose. The $\alpha(1\rightarrow6)$ linkage of all clusters besides T6 and T11 has its Φ_{24}-Ψ_{24}-Ω_{24} dihedrals in the isomaltose global minimum conformation (well A). The $\alpha(1\rightarrow4)$ linkages of all the clusters fall within either the A or B populations in the maltose map, which constitute the global minimum region for maltose. The only exception to these observations is cluster T6 from the first simulation. This cluster, with $\Phi_{24} = 147.9^o$, $\Psi_{24} = 199.2^o$ and $\Omega_{24} = 195.7^o$, represents a brief transition in the $\alpha(1\rightarrow6)$-linkage to the low energy region in the isomaltose map which was not among the six chosen minima discussed in chapter 5. The second simulation has a very similar behaviour to the first, clusters T1 to T4 being similar to those for the first simulation. Cluster T11 does not match any of the minima from the first simulation. With dihedrals $\Phi_{24} = 65.7^o$, $\Psi_{24} = 108.1^o$ and $\Omega_{24} = 200.9^o$, the $\alpha(1\rightarrow6)$ linkage of this cluster fits within the E well of the isomaltose adiabatic map. To simplify further analysis, only clusters 1 and 6 will be considered.

Illustrations of structures close to clusters T1 and T6 and clusters T7 and T11 are depicted in figure 6.14, on the basis of being the most populated (T1 and T7) and the most distinct (T6 and T11). Contrasting with panose, the cluster centres

sim.	clus.	members	std. dev.	Φ_{12}	Ψ_{12}	Φ_{23}	Ψ_{23}	Φ_{24}	Ψ_{24}	Ω_{24}
I	T1	12870	0.140e+02	-28.5	-25.6	-26.6	-24.1	70.9	193.0	193.3
	T2	10391	0.112e+02	-31.2	-27.3	-14.6	-20.3	68.3	159.6	193.2
	T3	5252	0.989e+01	-31.0	-27.3	-47.1	-36.6	69.8	167.6	193.1
	T4	4005	0.145e+02	-26.8	-23.4	-24.2	-21.9	75.1	194.5	174.8
	T5	3416	0.120e+02	-26.0	-24.1	-25.7	-25.5	70.1	165.9	174.8
	T6	2931	0.133e+02	-11.9	-17.5	-15.7	-20.1	147.9	199.2	195.7
II	T7	10619	0.110e+02	-19.7	-22.9	-19.5	-23.4	73.5	197.8	195.2
	T8	5656	0.108e+02	-16.2	-17.3	-53.5	-42.7	75.6	195.5	193.9
	T9	5087	0.966e+01	-29.1	-26.9	-18.7	-22.1	64.2	154.5	190.5
	T10	4285	0.956e+02	-26.7	-25.4	-49.2	-38.4	63.5	162.9	171.9
	T11	3943	0.120e+02	-23.3	-27.5	-26.8	-28.5	65.7	108.1	200.9

Table 6.9: Results of clustering for the two simulations of the four residue model. For each cluster center, only the linkage dihedrals are shown.

for this molecule all form several interresidue hydrogen bonds. Firstly, there are several instances of the *maltose-like* O2 - O3 hydrogen bonds, and the clusters shown have their $\alpha(1\rightarrow4)$ dihedrals in the maltose B well, so are capable of forming this type of hydrogen bond. The second type of hydrogen bond is that *between O6 of residue 4 and the O2 on either of residues 1 or 3* (residue 3 for cluster T6, and residue 1 for the remainder). The prototype for this is the hydrogen bond between O6 of residue 1 and O2 of residue 3 in the crystal structure of panose. The anomalous formation of this hydrogen bond to residue 3 instead of 1 in cluster T6 is made possible by the transition of the Φ_{24} dihedral of the $\alpha(1\rightarrow6)$ linkage from the usual value of around $70-80^o$ to 148^o. The absence of a well in this region of conformational space from the isomaltose map is now easy to explain, since no such interaction could occur for the simple disaccharide.

The branched glucose unit (residue 4) has thus been seen to adopt conformations in which it is closer to either residue 1 or residue 3. The origin of this preference is the hydrogen bond formed between O2 of the branched glucose to the primary alcohol of either residue 1 or 3. It must be emphasized that the cluster centre structures in figure 6.14 only represent a small sample of the conformational variation of this molecule in solution, as evident from the large number of distinct clusters. This behaviour is evidence of a flexibility about the $\alpha(1\rightarrow6)$ linkage which has been previously observed through indirect experimental measurements.

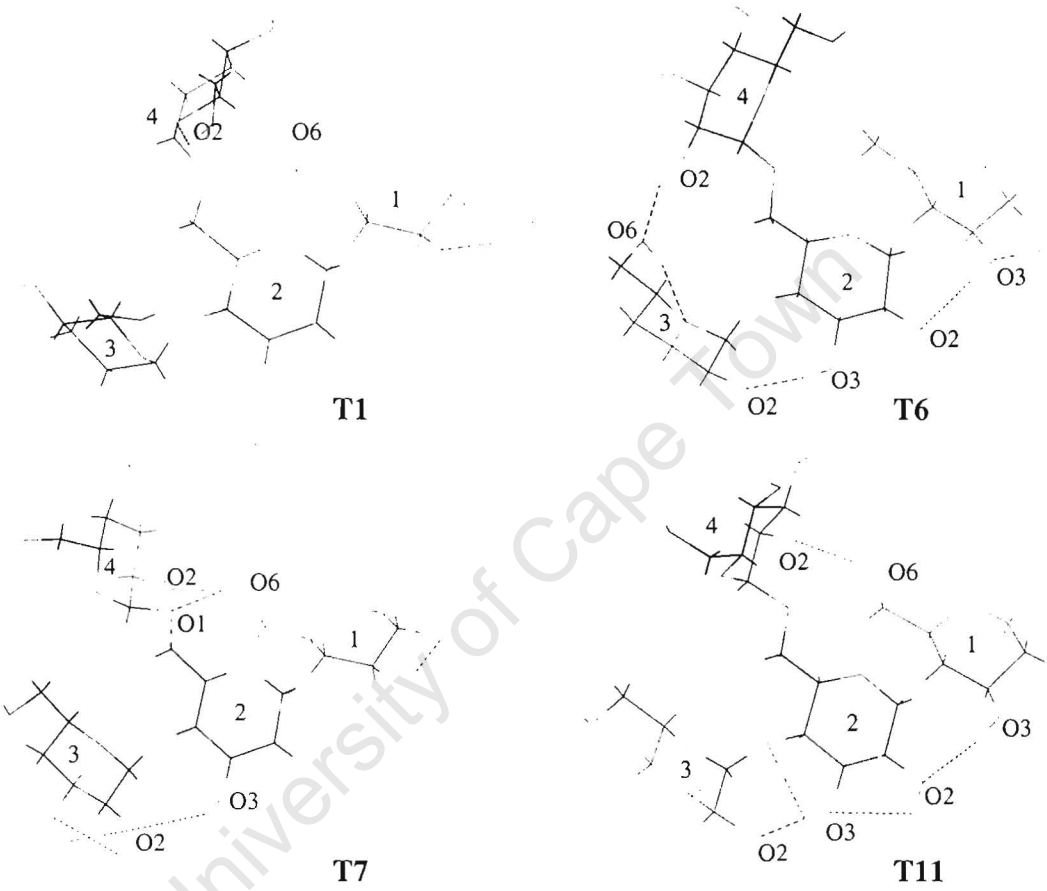


Figure 6.14: Structures closest to the cluster centres obtained from the simulation of the tetramer model.

6.2.4 Hydrogen Bond Analysis

The hydrogen bonding analysis for panose showed the direct measurement of potential oxygen-oxygen distances to be a valuable method for a flexible molecule. As the cluster analysis showed the tetrasaccharide model to form even more diverse clusters than panose, the validity of a discrete structure analysis is highly questionable in this case. The direct measurement of hydrogen bond distances is therefore likely to be the only useful method of analyzing the hydrogen bonding behaviour.

The time series for the oxygen-oxygen distances of several potential hydrogen bond donors in the four residue model are shown in figure 6.15. The first two distances tracked are the maltose-like O3-O2 hydrogen bonds which could be formed between residues 1 and 2, and 2 and 3. The latter two distances are non-sequential interactions between O2 of residue 4 and O6 on either residue 1 or 3. The hydrogen bond between residues 4 and 1 is that observed for panose in the crystal structure and in solution, as well as in most of the cluster centres identified for this molecule in figure 6.14. That between residues 4 and 3 has only been seen in cluster T6.

Little correlated formation of hydrogen bonds analogous to that observed for the second panose simulation is observed for either of the simulations of the four residue model. The clusters described for this model are distributed throughout the trajectory, and the cluster to which a frame of the trajectory belongs changes frequently during the run. This is consistent with the flexibility of the molecule. Some features of the hydrogen bonding are noteworthy, nonetheless. In simulation 1, in which the primary alcohols are both started in the *gt* rotamer, the unusual hydrogen bond between O2 of residue 4 and O6 of residue 3 is observed initially. Although this pairing is soon disrupted (after 200 ps), the substitute pairing between O2 of residue 4 and O6 of residue 1 is not formed for any substantial length of time until the primary alcohol Θ_1 makes the transition from *gt* to *gg* at 800 ps. Short lived pairings between these atoms are indeed observed between 200 and 800 ps; however, the analogous interactions are also observed in the first panose simulation in which the corresponding primary alcohol is also in the *gt* form, and no long-lasting O2 - O6 interactions are observed there either. A prerequiste for hydrogen bonding between residues 1 and 4 thus seems to be the occupation of the *gg* rotamer by Θ_1.

6.2.5 Solvent Structure Analysis

As for panose, the heterogeneity of conformations adopted requires some kind of restriction to be imposed on the structures analysed. This is especially true in this case, as the large number of dihedral angles used in the clustering process produces many

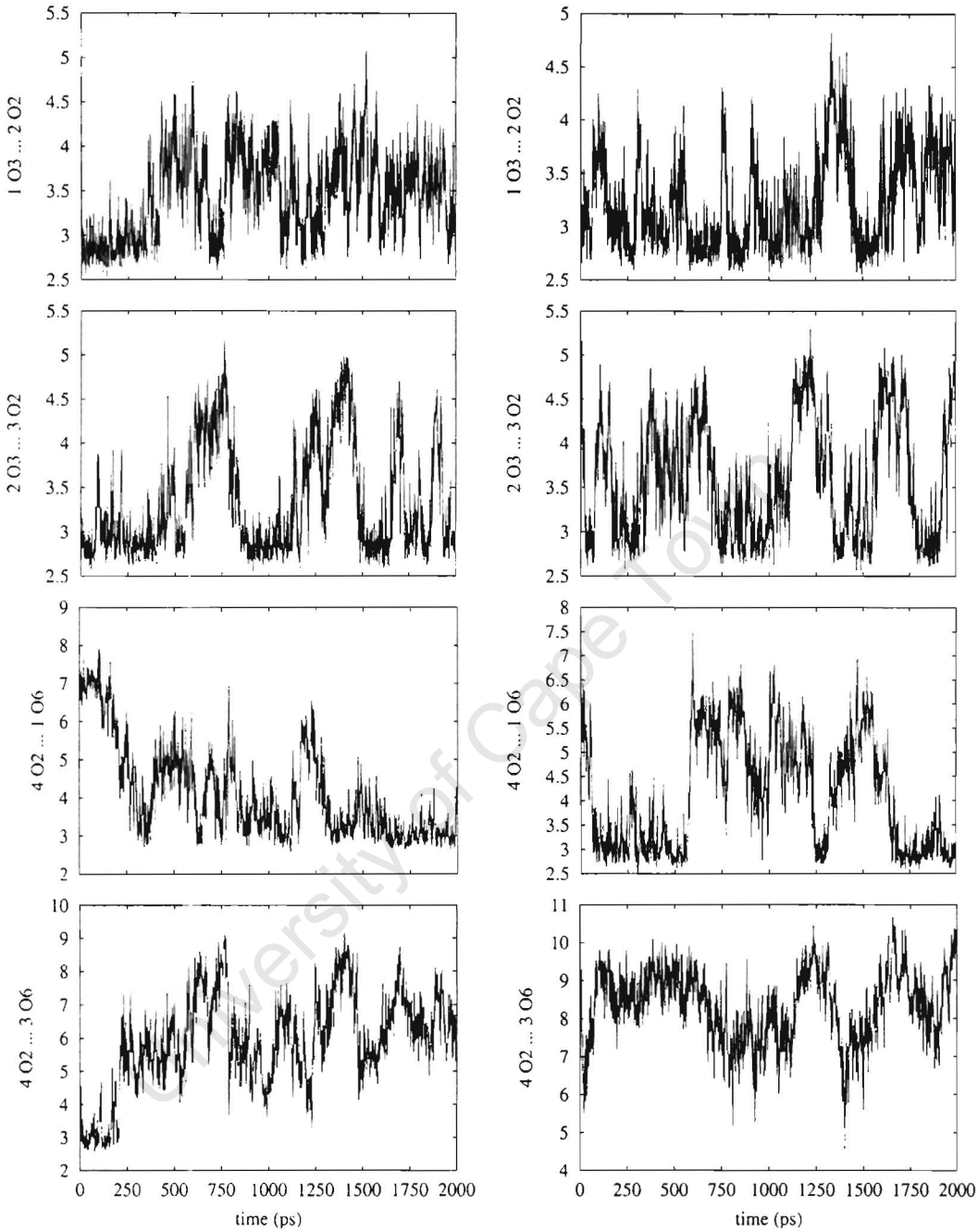


Figure 6.15: Time series of distances between potential hydrogen bonded partners for four residue model, simulation 1 (a) and simulation 2 (b).

clusters, many of which are not significantly different. Only clusters which are distinct, and therefore structure the solvent differently, will be used in further analysis.

Clusters T1 and T6 were used to analyse the solvent distribution for the four residue model. T1 represents the main part of the trajectory in which O2 of residue 4 forms a hydrogen bond to O6 of residue 1, causing residue 4 to fold back towards residue 1. Cluster T6 exemplifies the alternative in which the O2 on residue four instead hydrogen bonds to O6 on residue 3, causing residue 4 to lie closer to residue 3 than to 1. The two cases thus represent the alternatives of a backfolded branch (T1) or one at right angles (T6).

The spatial distribution functions for cluster T1 are given in figure 6.16 and 6.17. The peaks of water density around the cluster T1 do follow the general trend of being located between the hydroxyl groups. However, there is considerably more disorder than would be the case for a simple monosaccharide. The water density surrounding residue 4 is especially scattered, suggesting a high degree of mobility for this residue, which would support the hypothesis that the $\alpha(1\rightarrow6)$ linkage is more flexible. However, in addition to the usual regions of high density, new peaks have appeared. Between residues 1 and 2 are two new water density peaks; being within hydrogen bonding distance of both O3 on residue 1 and O2 on residue 2, these are very likely involved in a hydrogen bonded water bridge replacing the standard direct hydrogen bond of maltose. Other water density peaks appear within the two clefts formed by the "chain" stubs abutting the central residue (2). Although these are generally not involved in any kind of special interaction with the solute, their high density may be attributable to a lower rate of diffusional exchange with the bulk solvent than may be the case for water molecules peripherally hydrogen bonded to the molecule.

The shape of cluster T6 is determined by the strong maltose-like O2-O3 hydrogen bonds between residues 1, 2 and 3, and the unusual hydrogen bond between residues 3 and 4. The "main chain" comprising monomers 1 to 3 forms a helical segment, which together with residue 4 form a bowl-like structure (in figure 6.17, the bowl faces away from the viewer). Most of the water density falls within this bowl, the majority of the hydroxyl groups pointing in toward the bowl region. This in contrast to the much flatter structure T1, which has approximately the same water density on both sides.

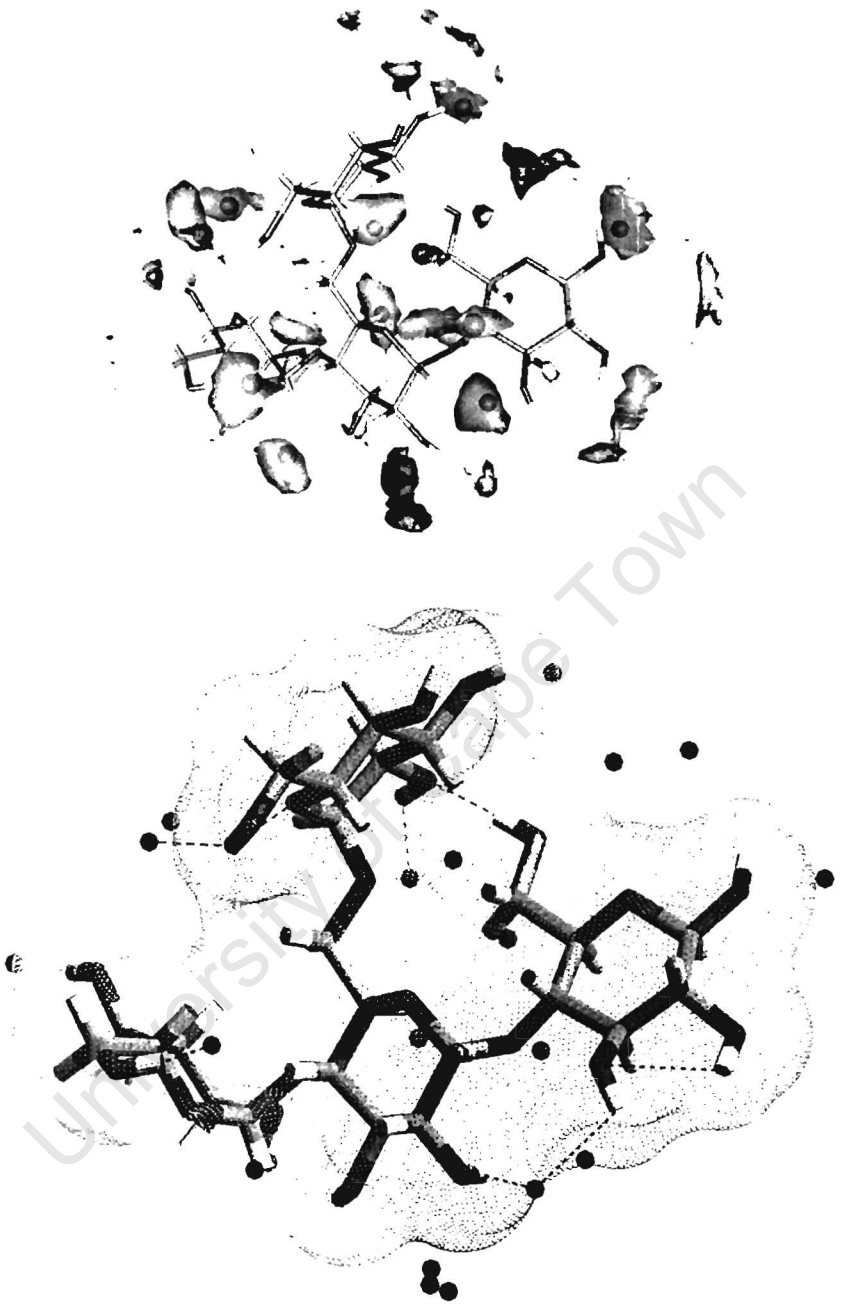


Figure 6.16: Three dimensional contour plot of water density around the branch model, for structures from cluster T1 (above); Points of maximum water density around the cluster centre structure with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

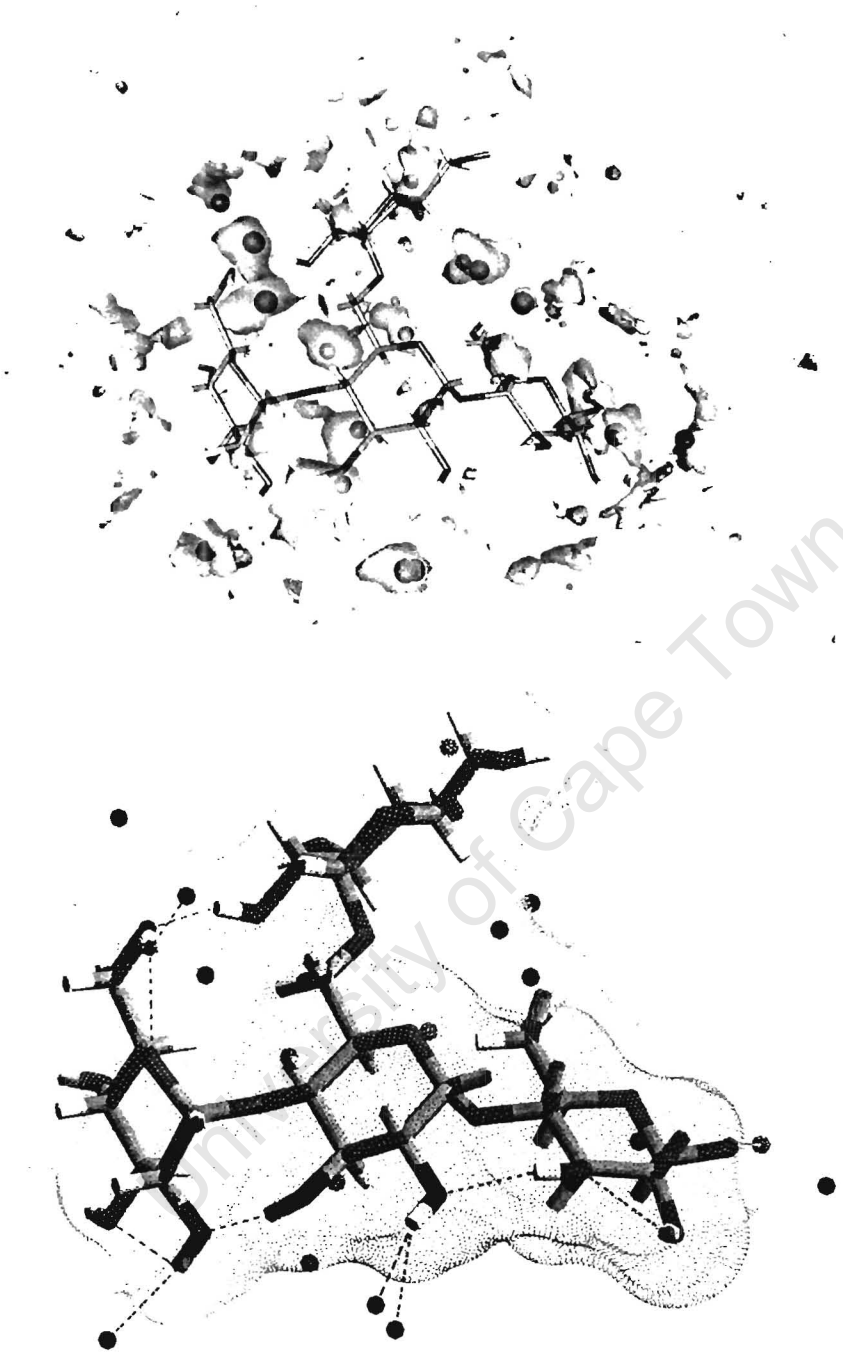


Figure 6.17: Three dimensional contour plot of water density around the branch model, for structures from cluster T6; Points of maximum water density around the cluster centre structure with putative hydrogen bonds shown (dotted skin indicates Connolly solvent accessible surface) (below).

6.3 Branch models in the context of larger polysaccharides

The three and four residue branch models are able to tell us something about the structure and dynamics of the polysaccharide branch point which they are intended to mimic. The existence of multiple clusters for both molecules, and the frequent exchange of conformation between clusters attests to the fact that they are much more flexible than simple $\alpha(1\rightarrow4)$ linked chains. This is especially the case for the tetrasaccharide, for which multiple clusters were observed, defying a simplistic description in terms of static structures. Both the $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ linkages sampled essentially the same areas of the maltose and isomaltose adiabatic maps as were sampled by the disaccharides themselves. In addition, the dynamic behaviour of each type of linkage in the three and four residue models, as measured by the range of dihedral space sampled and rate of transitions made, was similar to that observed for the isolated disaccharides: the $\alpha(1\rightarrow4)$-linkage made frequent transitions (corresponding to the breaking and forming of the *maltose-like* O2-O3 type hydrogen bond) between maltose minima A and B. Although the $\alpha(1\rightarrow6)$-linkage did not undergo as many actual transitions to different wells, its flexibility, especially in the Ψ dihedral was evident in all runs. This confirms the hypothesis that Ψ, sampling a range of 150°, is the principal source of conformational variation for the $\alpha(1\rightarrow6)$-linkage. The similarity in behaviour of the isolated linkages in isomaltose and maltose to the linkages in the larger molecules validates the use of disaccharide models to map the potential surface of glycosidic linkages in oligo- and polysaccharides.

In addition to the interactions between sequential residues, which are contained in the disaccharide adiabatic maps, the use of more than two residues allows non-sequential interactions to be modelled. Both models provide us with more information about the joint than would be available from studying the dimeric residues alone. There are both direct and water-bridged hydrogen bonded pairings between non-sequential residues which act to stabilize the joint to some extent. The most important of these is the direct hydrogen bond observed in the panose crystal structure between O6 of residue 1 and O2 of residue 3. This hydrogen bond is seen in all the simulations in which it can potentially form (see note about primary alcohols below).

As a model, panose provides the most important features and groups which were found to induce structure in the four residue model, notably the hydrogen bonding between residues 1 and 3. What it does not provide is the alternative hydrogen bonding partner seen in cluster T6, which causes residue 4 of the four residue model to move closer to residue 3 than 1. This has the effect of opening up the panose fragment in the

tetrasaccharide, something which was not observed for panose. However, the frequency of occurrence of this structure appears to be very low, and therefore not very important.

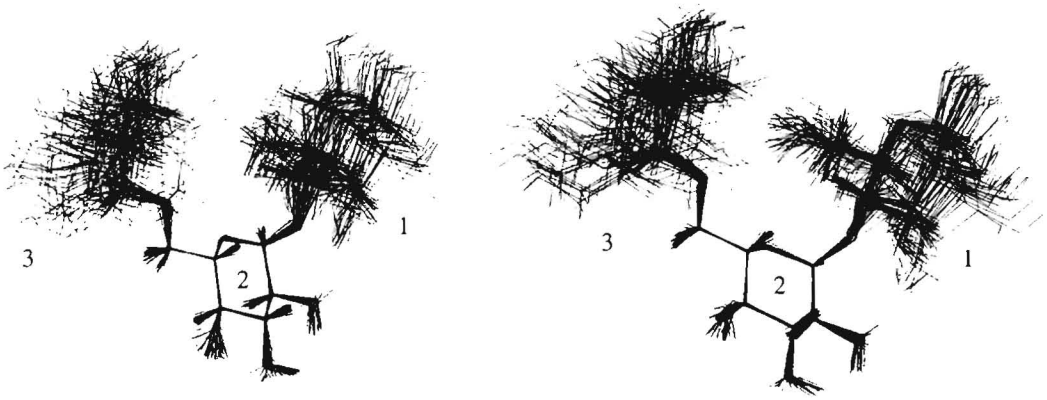
The influence of the primary alcohols on the available hydrogen bonding partners and hence on three-dimensional structure has already been mentioned: for both panose and the tetrasaccharide, it appears as though the *gg* rotamer is a requirement for the formation of a durable non-sequential O2-O6 hydrogen bond between the terminal residues of a panose fragment, although bridging hydrogen bonds in solution have been observed for the *gt* rotamer. The difficulty of rotation about the C5-C6 bond and the overwhelming preference for the *gg* rotamer limit the available conformations. This aspect of the PHLB force-field is one which is currently being addressed[79] in order to get a good sampling of conformational space in future simulations, and to avoid the separation of conformations dependent on the starting primary alcohol configuration, as was the case for panose.

The good agreement between the experimentally measured T_1 relaxation times and those calculated from the simulations gives confidence that the force-field and simulation conditions are able to produce an accurate representation of the dynamics of the molecule.

If the model compounds adopted are viewed as being fragments of a larger polysaccharide, some important observations can be made. The largest clusters of structures for both panose and the four residue model encourage back-folding of the chain attached to residue 4, in such a fashion that it is approximately parallel with that attached to residue 3. This accords with solid state studies in which the $\alpha(1\rightarrow4)$ linked chains of amylopectin are found to lie in a parallel orientation in starch. Only a few of the clusters provided alternative relative chain orientations, and these were small in size.

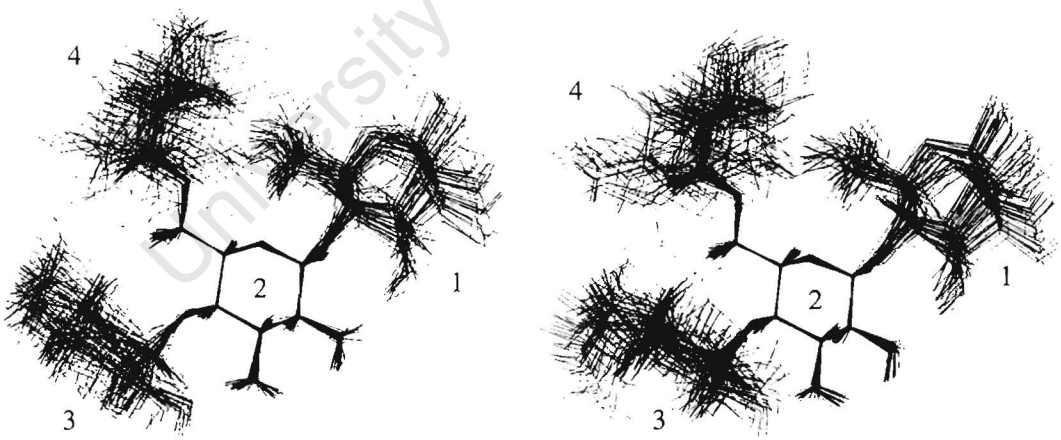
Overlays of the trajectories for panose and the tetramer have been constructed in figures 6.18 and 6.19 respectively, using frames sampled at even intervals throughout the simulation. These give an overall visual impression of the flexibility and range of the linkage. The close association of rings 1 and 3 in panose and 1 and 4 for the tetramer indicates the preference for a "back-folded" structure in all cases.

Clearly, such inferences about the structure from small models are premature at this stage: the direction of the strands about each branch are after all being deduced from just a single residue in each direction. This is not necessarily unreasonable however, since the behaviour of the tri- and tetrasaccharide models would be expected to be mirrored by larger molecules, just as these models themselves reflected the underlying behaviour of disaccharides. Nonetheless, an important component of any future work in this area would have to be the treatment of models with several residues on each side of the junction.



Simulation 1 Simulation 2

Figure 6.18: Regularly sampled structures from each panose simulation overlayed using the central ring (2) carbons for fitting.



Simulation 1 Simulation 2

Figure 6.19: Regularly sampled structures from each simulation of the tetrasaccharide model overlayed using the central ring (2) carbons for fitting.

References

1. Zugenmaier, P.; Sarko, A. *Biopolymers* **1976,** *15,* 2121-2136.
2. Wu, H. C.; Sarko, A. *Carb. Res.* **1978,** *61,* 7-40.
3. Horii, F.; Yamamoto, H.; Hirai, A.; Kitamaru, R. *Carb. Res.* **1987,** *160,* 29-40.
4. Waigh, T. A.; Hopkinson, I.; Donald, A. M.; Butler, M. F.; Heidelbach, F.; Riekel, C. *Macromol.* **1997,** *30,* 3813-3820.
5. Galinsky, G.; Burchard, W. *Macromol.* **1997,** *30,* 4445-4453.
6. Waigh, T. A.; Jenkins, P. J.; Donald, A. M. *Faraday Discussions* **1996,** *103,* 325-337.
7. Fukuda, M.; Hindsgaul, O., Eds.; *Molecular Glycobiology;* Oxford University Press: Oxford, United Kingdom, 1994.
8. Doane, W. M. *Starch/Stäerke* **1992,** *44,* 293-295.
9. Shogren, R. L.; F.Fanta, G.; Doane, W. M. *Starch/Stäerke* **1993,** *45,* 276-280.
10. Benczédi, D.; Tomka, I.; Escher, F. *Macromol.* **1998,** *31,* 3055-3061.
11. Benczedi, D.; Tomka, I.; Escher, F. *Macromol.* **1998,** *31,* 3062-3074.
12. Imberty, A.; Perez, S. *Carb. Res.* **1988,** *181,* 41-55.
13. Jeffrey, G. A. *Carb. Res.* **1991,** *222,* 47-55.
14. Quigley, G. J.; Sarko, A.; Marchessault, R. H. *J. Am. Chem. Soc.* **1970,** *92,* 5834-5839.
15. Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980,** *36,* 2783-2792.
16. Bose, B.; Zhao, S.; Stenutz, R.; Cloran, F.; Bondo, P. B.; Bondo, G.; Hertz, B.; Carmichael, I.; Serianni, A. S. *J. Am. Chem. Soc.* **1998,** *120,* 11158-11173.
17. Poppe, L. *J. Am. Chem. Soc.* **1993,** *115,* 8421-8426.
18. Bock, K.; Duus, J. O. *J. Carb. Chem.* **1994,** *13,* 513-543.
19. Rockwell, G. D.; Grindley, T. B. *J. Am. Chem. Soc.* **1998,** *120,* 10953-10963.
20. Dais, P. *NATO ASI Ser.* **1994,** *87,* 263-278.

21. Dais, P. *NATO ASI Ser.* **1994,** *87,* 279-292.

22. Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982,** *104,* 4546-4559.

23. Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W. *Carb. Res.* **1988,** *180,* 207-221.

24. Ha, S. N.; Madsen, L. J.; Brady, J. W. *Biopolymers* **1988,** *27,* 1927-1952.

25. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Comp. Chem.* **1997,** *18,* 1955-1970.

26. Palma, R.; Himmel, M. E.; Liang, G.; Brady, J. W. *In Glycosyl Hydrolases in Biomass Conversion: Vol. ACS Symposium Series,* in press.

27. Hardy, B. J.; Egan, W.; Widmalm, G. *Intl. J. Biol. Macromol.* **1995,** *17,* 149-159.

28. Engelsen, S. B.; du Penhoat, C. H.; Perez, S. *J. Phys. Chem.* **1995,** *99,* 13334-13351.

29. Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry;* Dover Publications: London, First ed.: 1996.

30. Jensen, F. *Introduction to Computational Chemistry;* John Wiley and Sons: New York, 1999.

31. Riedi, P. C. *Thermal Physics: an introduction to thermodynamics, statistical mechanics, and kinetic theory;* Oxford University Press: Oxford, Second ed.; 1988.

32. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983,** *4,* 187-217.

33. Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984,** *5,* 129-145.

34. Metropolis, N.; Metropolis, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953,** *21,* 1087.

35. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids;* Oxford Science Publications: Oxford, First ed.; 1989.

36. Karplus, M.; Petsko, G. A. *Nature* **1990,** *347,* 631-639.

37. van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem. Intl. Ed.* **1990,** *29,* 992-1023.

38. Andersen, H. C. *J. Chem. Phys.* **1980,** *72,* 2384-2393.

39. Carr, R.; Parrinello, M. *Phys. Rev. Lett.* **1985,** *55,* 2471-2474.

40. Molteni, C.; Parrinello, M. *J. Am. Chem. Soc.* **1998,** *120,* 2168-2171.

41. Verlet, L. *Phys. Rev.* **1967,** *159,* 98-103.

42. Tasaki, K.; McDonald, S.; Brady, J. W. *J. Comp. Chem.* **1993,** *14,* 278-284.

43. Steinbach, P. J.; Brooks, B. R. *J. Comp. Chem.* **1994,** *15,* 667-683.

44. Kitchen, D. B.; Hirata, F.; Westbrook, J. D.; Levy, R.; Kofke, D.; Yarmush, M. *J. Comp. Chem.* **1990,** *11,* 1169-1180.

45. Adams, D. J. *Chem. Phys. Lett.* **1979,** *62,* 329-332.

46. Fincham, D.; Heyes, D. M. *Adv. Chem. Phys.* **1985,** *63,* 493-572.

47. Bekker, H. *J. Comp. Chem.* **1997,** *18,* 1930-1942.

48. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995,** *103,* 4613-4621.

49. Soper, A. K.; Phillips, M. G. *Chem. Phys.* **1986,** *107,* 47-60.

50. Svishchev, I. M.; Kusalik, P. G. *J. Chem. Phys.* **1993,** *99,* 3049-3058.

51. Immel, S.; Lichtenthaler, F. W. *Liebigs Annalen* **1995,** 1925-1937.

52. Liu, Q.; Brady, J. W. *J. Am. Chem. Soc.* **1996,** *118,* 12276-12286.

53. Liu, Q.; Brady, J. W. *J. Phys. Chem. B* **1997,** *101,* 1317-1321.

54. Kuttel, M. "Developing Analytical Tools for Saccharides in Condensed Phases", Master's thesis, University of Cape Town, 1999.

55. Vishnyakov, A.; Widmalm, G.; Kowalewski, J.; Laaksonen, A. *J. Am. Chem. Soc.* **1999,** *121,* 5403-5412.

56. Astley, T.; Birch, G. G.; Drew, M. G. B.; Rodger, P. M. *J. Phys. Chem. A* **1999,** *103,* 5080-5090.

57. Brady, J. W.; Schmidt, R. K. *J. Phys. Chem.* **1993,** *97,* 958-966.

58. Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975,** *97,* 1354-1358.

59. Carpenter, G. A.; Grossberg, S. *Appl. Opt.* **1987,** *26,* 4919-4930.

60. Karpen, M. E.; Tobias, D. J.; Brooks, C. L. *Bioch.* **1993,** *32,* 412-420.

61. Zhu, S.-B.; Singh, S.; Robinson, G. W. *Adv. Chem. Phys.* **1994,** *85,* 627-731.

62. Narton, A. H.; Levy, H. A. *J. Chem. Phys.* **1971,** *55,* 2263-2269.

63. Thiessen, W. H.; Narten, A. H. *J. Chem. Phys.* **1982,** *77,* 2656-2662.

64. Thiessen, W. H.; Narten, A. H. *Science* **1982,** *217,* 1033-1034.

65. Soper, A. K. *Chem. Phys.* **1986,** *107,* 61-74.

66. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983,** *79,* 926-935.

67. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987,** *91,* 6269-6271.

68. Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996,** *105,* 1902-1921.

69. Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. (USA)* **1993,** *90,* 9135-9139.

70. Ryckaert, J. P.; Cicotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977,** *23,* 327-341.

71. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Intermolecular Forces;* Dordrecht: Dordrecht, First ed.; 1981.

72. Bernal, J. D.; Fowler, R. H. *J. Chem. Phys.* **1933,** *77,* 4156.

73. Stillinger, F. H.; Rahman, A. *J. Chem. Phys.* **1974,** *60,* 1545-1557.

74. Ferguson, D. M. *J. Comp. Chem.* **1995,** *16,* 501-11.

75. Bernardo, D. N.; Ding, Y.; Krogh-Jespersen, K.; Levy, R. M. *J. Phys. Chem.* **1994,** *98,* 4180-4187.

76. Jorgensen, W. L.; Jenson, C. *J. Comp. Chem.* **1998,** *19,* 1179-1186.

77. Brady, J. W. *J. Am. Chem. Soc.* **1989,** *111,* 5155-5165.

78. Marchessault, R.; Perez, S. *Biopolymers* **1979,** *18,* 2369.

79. Homans, S. W. *Bioch.* **1990,** *29,* 9110-9118.

80. Naidoo, K. J ; Kuttel, M. M. Unpublished data.

81. Momany, F. A.; Willet, J. L. Unpublished data.

82. Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995,** *99,* 3832-3846.

83. Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982,** *104,* 4559-4570.

84. Vold, R. L.; Waugh, J. S.; Klein, M. P.; Phelps, D. E. *J. Phys. Chem.* **1968,** *48,* 3831-3832.

85. Derome, A. E. *Modern NMR Techniques for Chemistry Research;* Pergamon Press: Oxford, 1987.

86. Ott, K.-H.; Meyer, B. *Carb. Res.* **1996,** *281,* 11-34.

87. Shashkov, A. S.; Lipkind, G. M.; Kochetkov, N. K. *Carb. Res.* **1986,** *147,* 175-182.

88. Gress, M. E.; Jeffrey, G. A. *Acta Cryst.* **1977,** *B34,* 213-218.

89. Stevens, E. S.; Sathyanarayana, B. K. *J. Am. Chem. Soc.* **1989,** *111,* 4149-4154.

90. Schmidt, R. K.; Teo, B.; Brady, J. W. *J. Phys. Chem.* **1995,** *99,* 11339-11343.

91. Brady, J. W.; Naidoo, K. J. Unpublished data.

92. Benesi, A. J.; Brant, D. A. *Macromol.* **1985,** *18,* 1109-1116.

93. McCain, D. C.; Markley, J. L. *J. Mag. Res.* **1987,** *73,* 244-251.

94. Kowalewski, J.; Widmalm, G. *J. Phys. Chem.* **1994,** *98,* 1033-1034.

95. Maler, L.; Widmalm, G.; Kowalewski, J. *J. Phys. Chem.* **1996,** *100,* 17103-17110.

96. Mandel, A. M.; Akke, M.; Palmer, A. G. *J. Mol. Bio.* **1995,** *246,* 144-163.

97. III, A. G. P.; Rance, M.; Wright, P. E. *J. Am. Chem. Soc.* **1991,** *113,* 4371-4380.

98. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983,** *220,* 671-680.

99. Naidoo, K. J.; Brady, J. W. *Chem. Phys.* **1997,** *224,* 263-273.

100. Fischer, S.; Karplus, M. *Chem. Phys. Lett.* **1992,** *194,* 252-261.

101. Chiu, S.-W.; Clark, M.; Subramaniam, S.; Jakobsson, E. *J. Comp. Chem.* **2000,** *21,* 121-131.

102. Arnosti, C.; Repeta, D. J. *Starch/Stäerke* **1995,** *47,* 73-75.

103. Kessler, H.; Oschkinat, H.; Griesinger, C. *J. Mag. Res.* **1986,** *70,* 513-543.

104. Kessler, H.; Anders, U.; Gemmecker, G.; Steuernagel, S. *J. Mag. Res.* **1989,** *85,* 1-14.

Appendix A

Isomaltose dynamics plots

A.1 Time Series plots

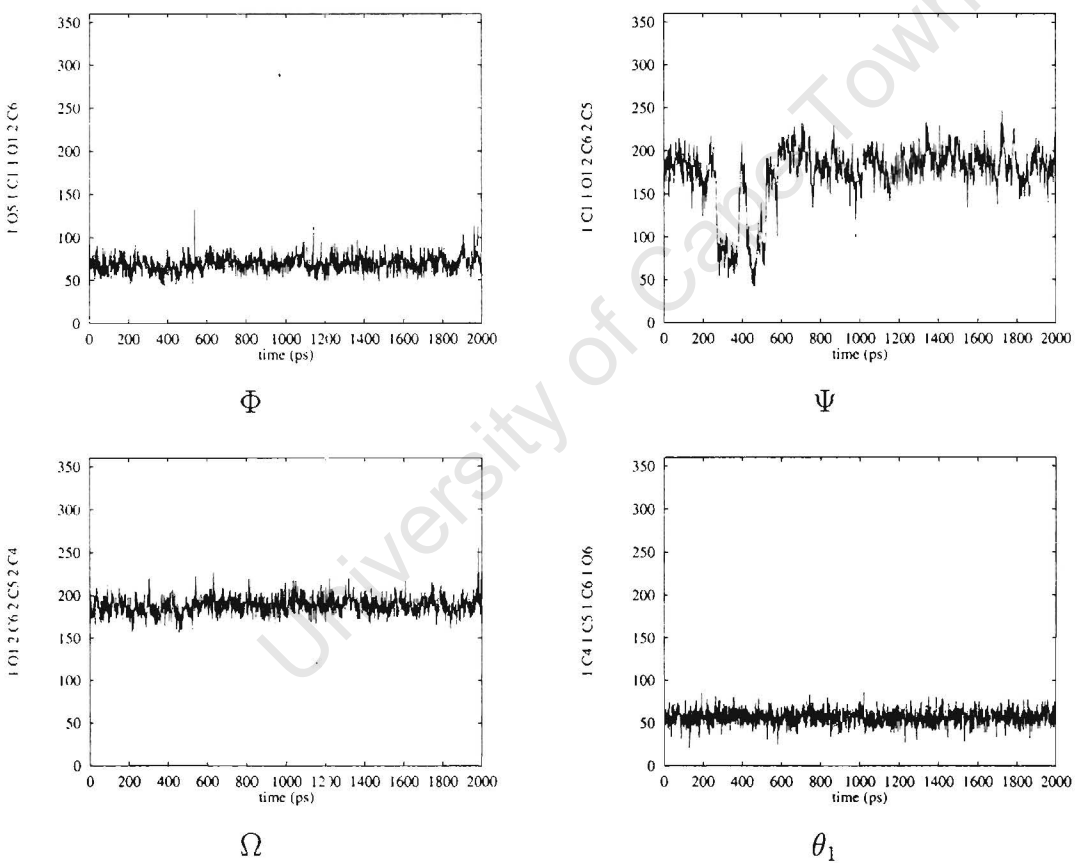


Figure A.1: Timeseries for isomaltose in solution from minimum A

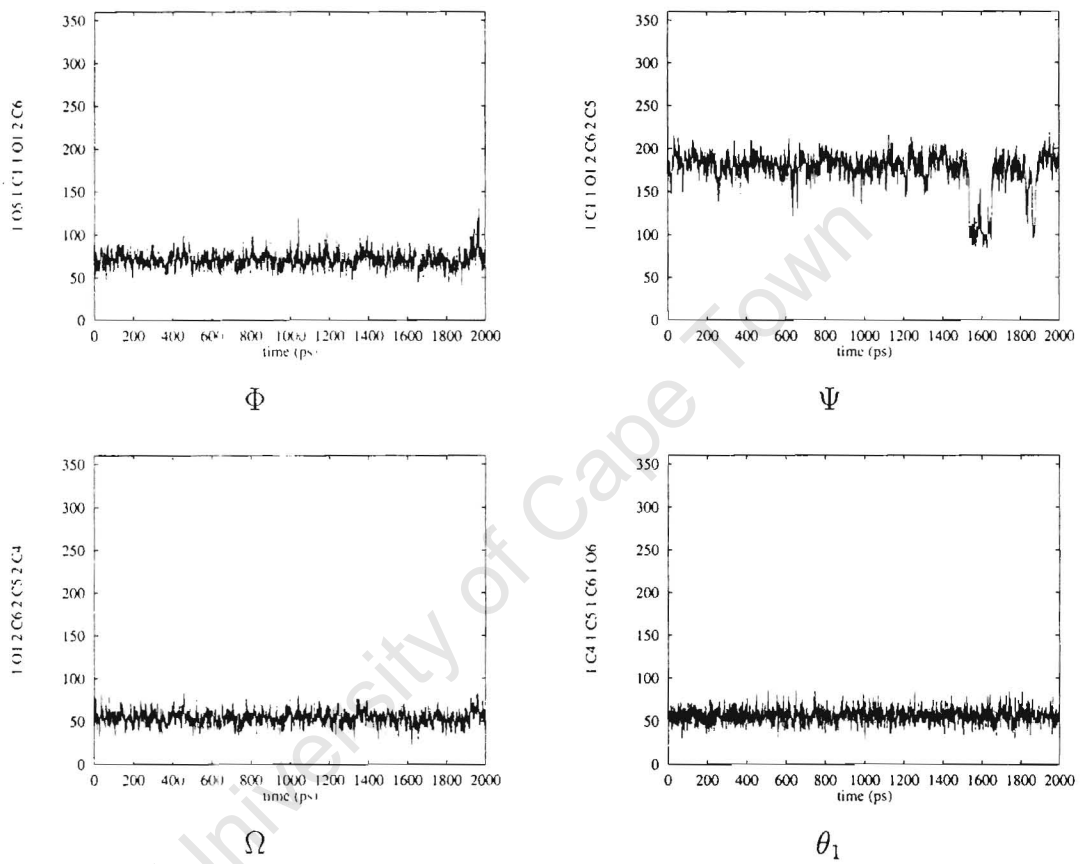


Figure A.2: Timeseries for isomaltose in solution from minimum B

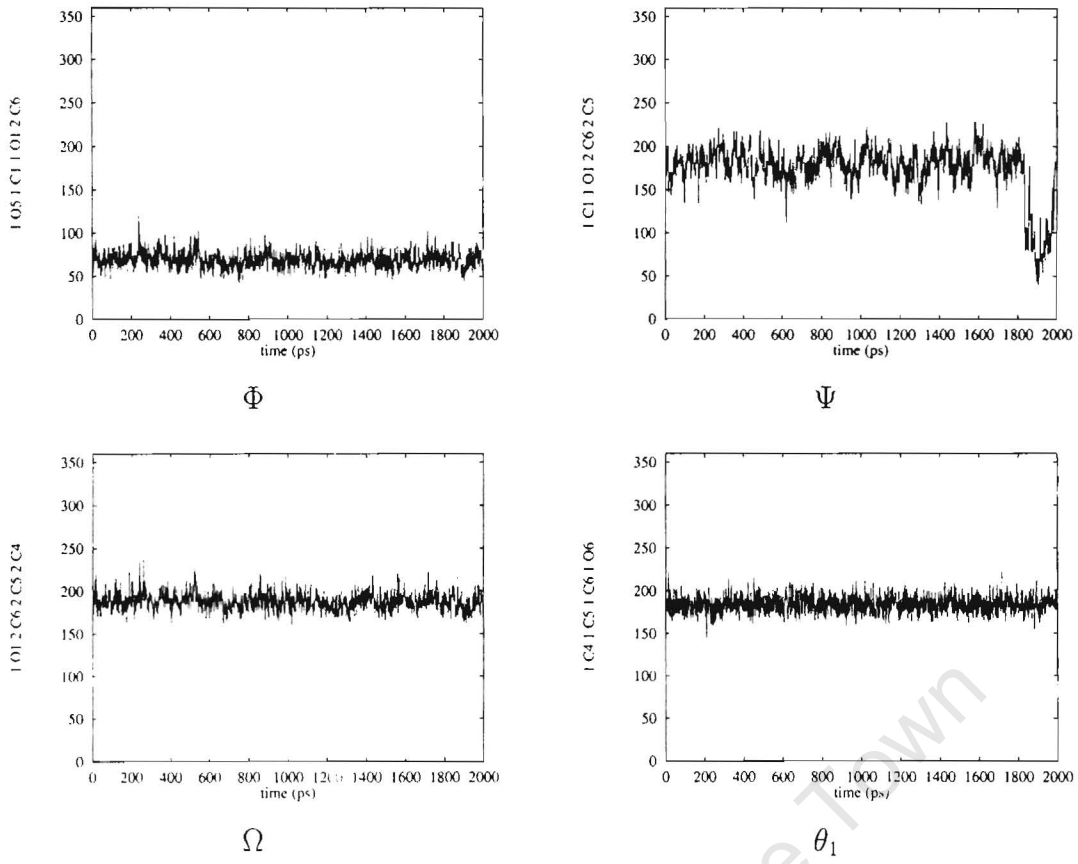


Figure A.3: Timeseries for isomaltose in solution from minimum C

A.2 Correlation plots

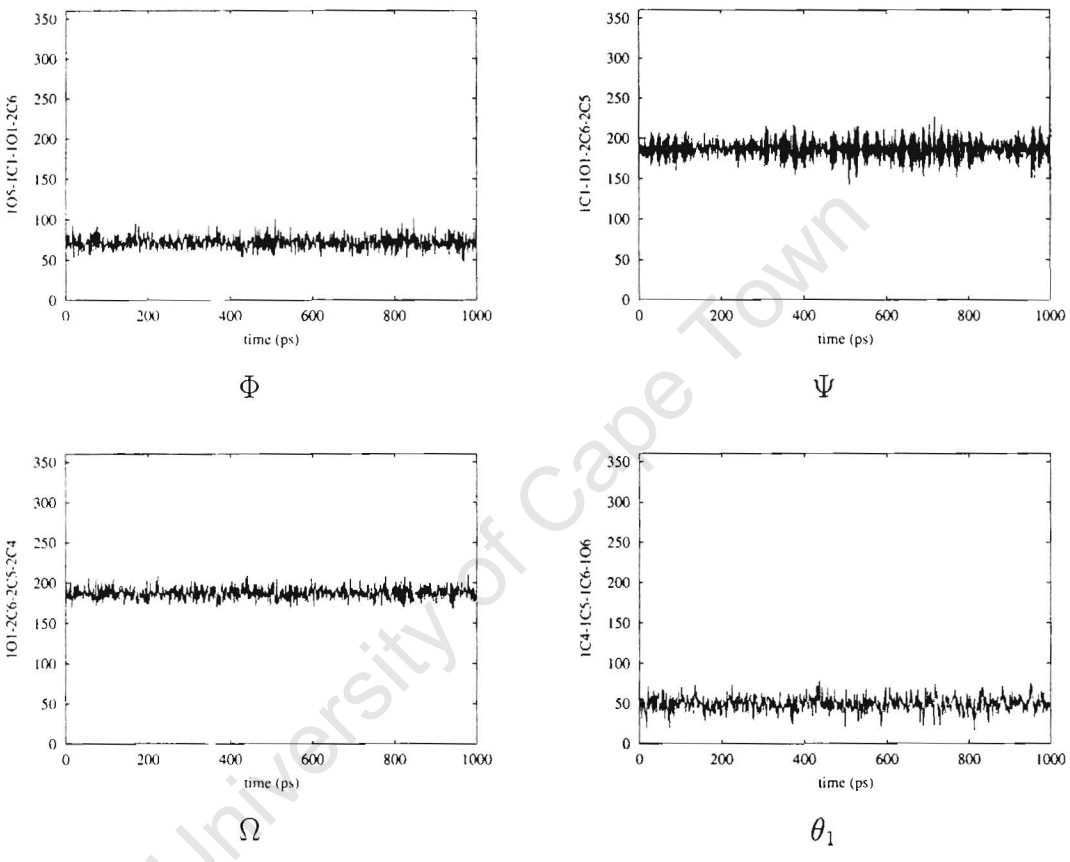


Figure A.4: Timeseries for isomaltose in vacuum from minimum A

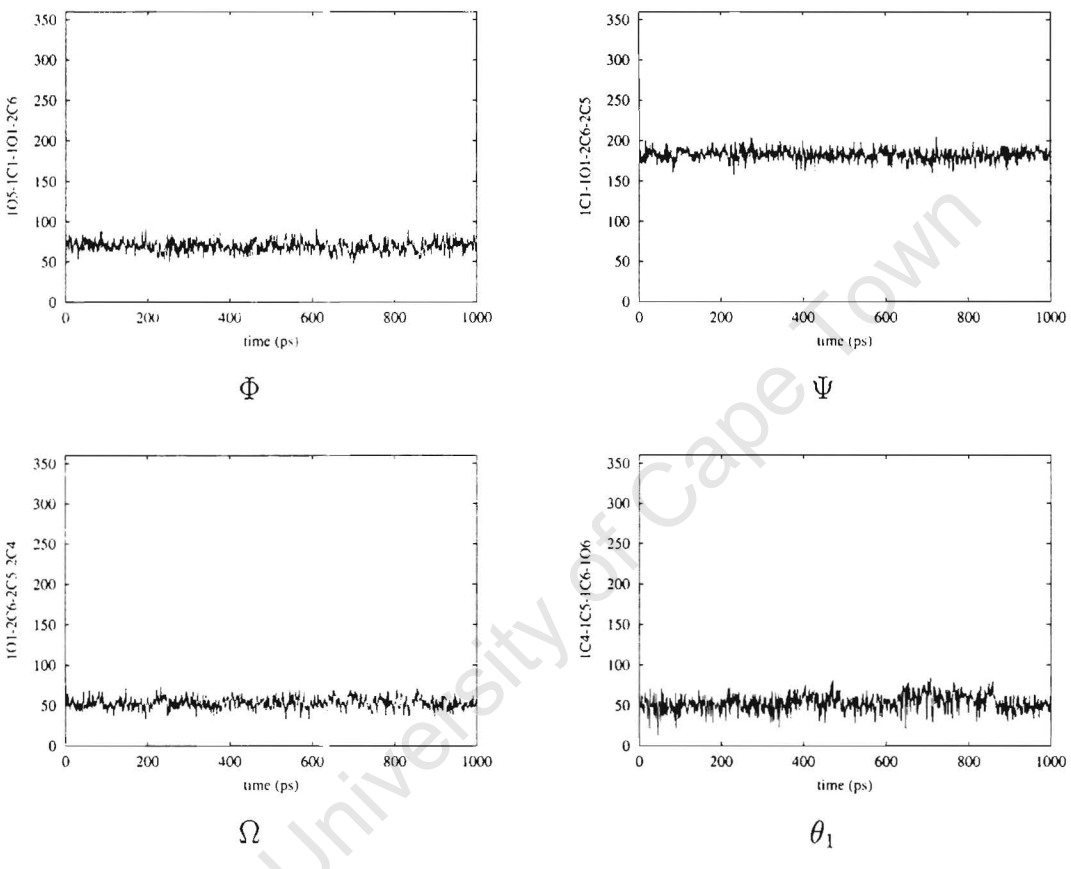


Figure A.5: Timeseries for isomaltose in vacuum from minimum B

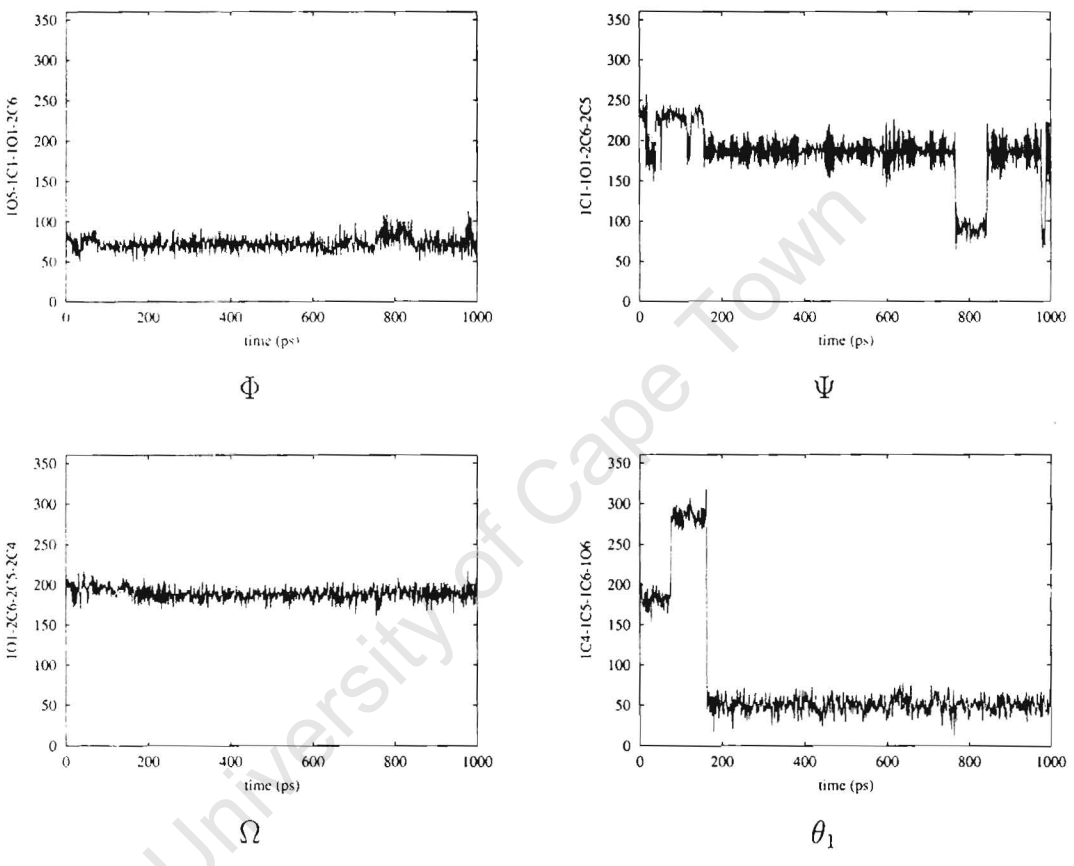


Figure A.6: Timeseries for isomaltose in vacuum from minimum C

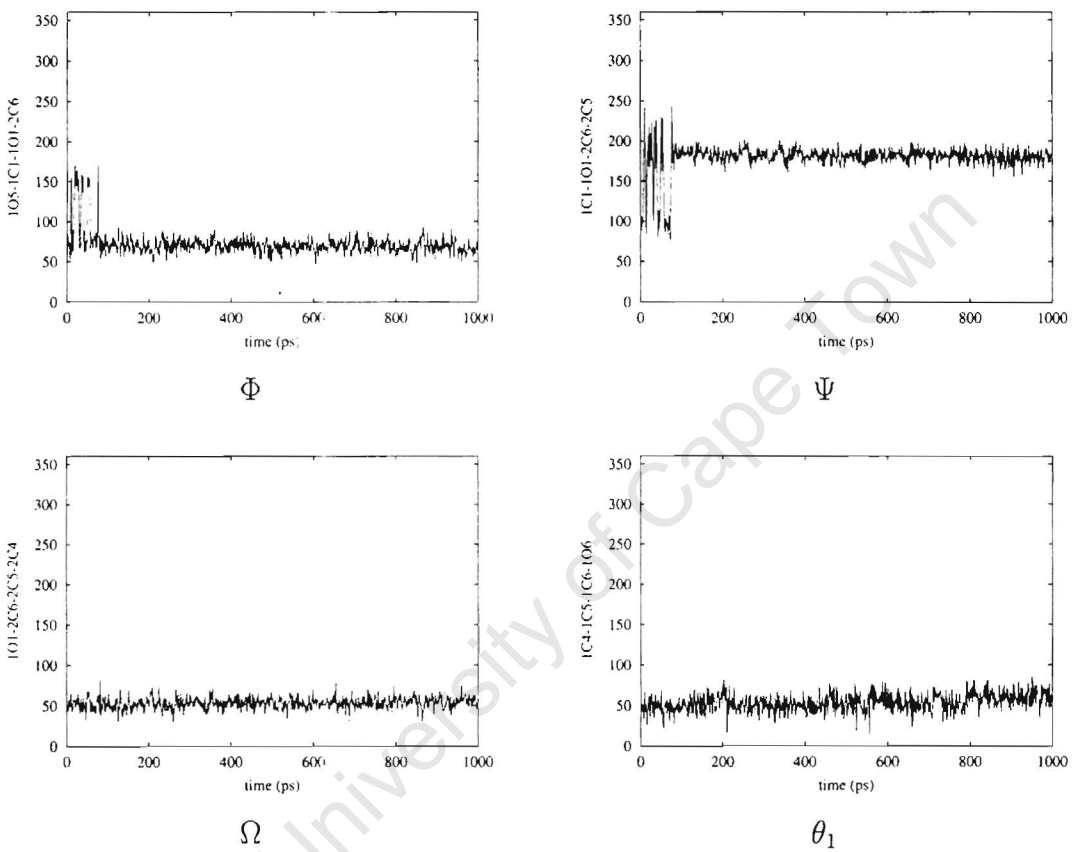


Figure A.7: Timeseries for isomaltose in vacuum from minimum D

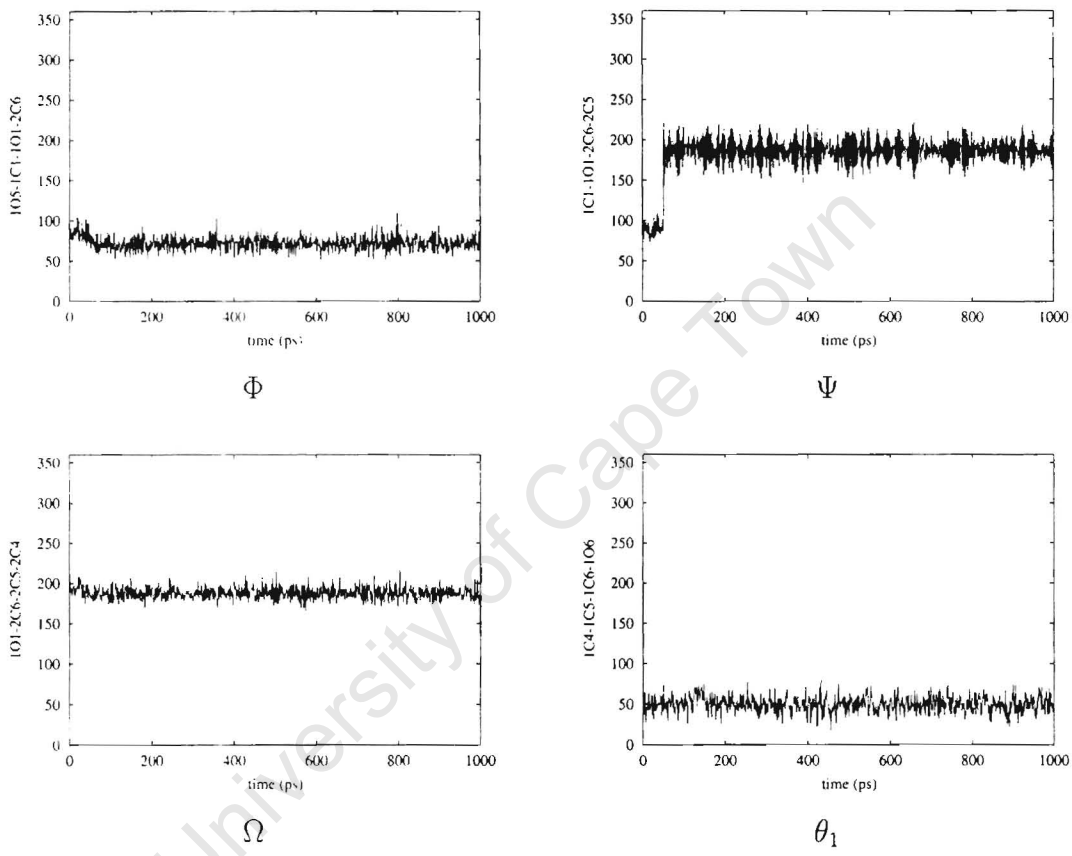


Figure A.8: Timeseries for isomaltose in vacuum from minimum E

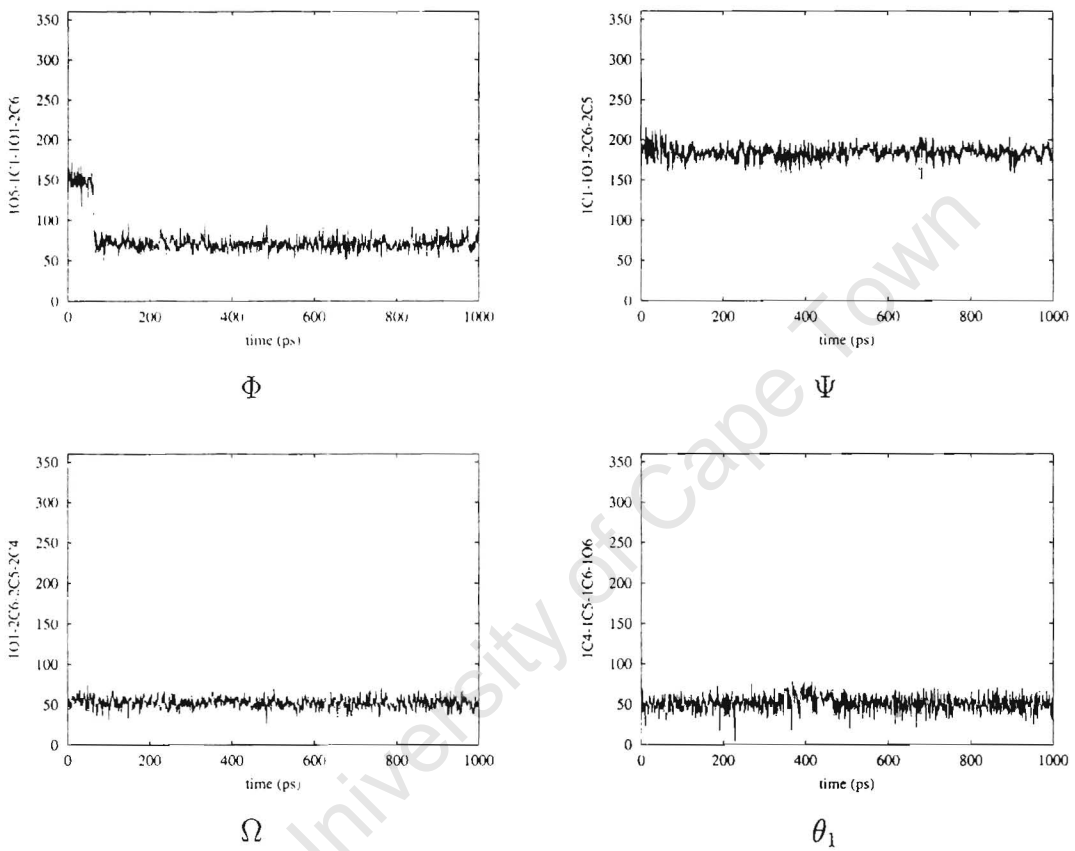


Figure A.9: Timeseries for isomaltose in vacuum from minimum F

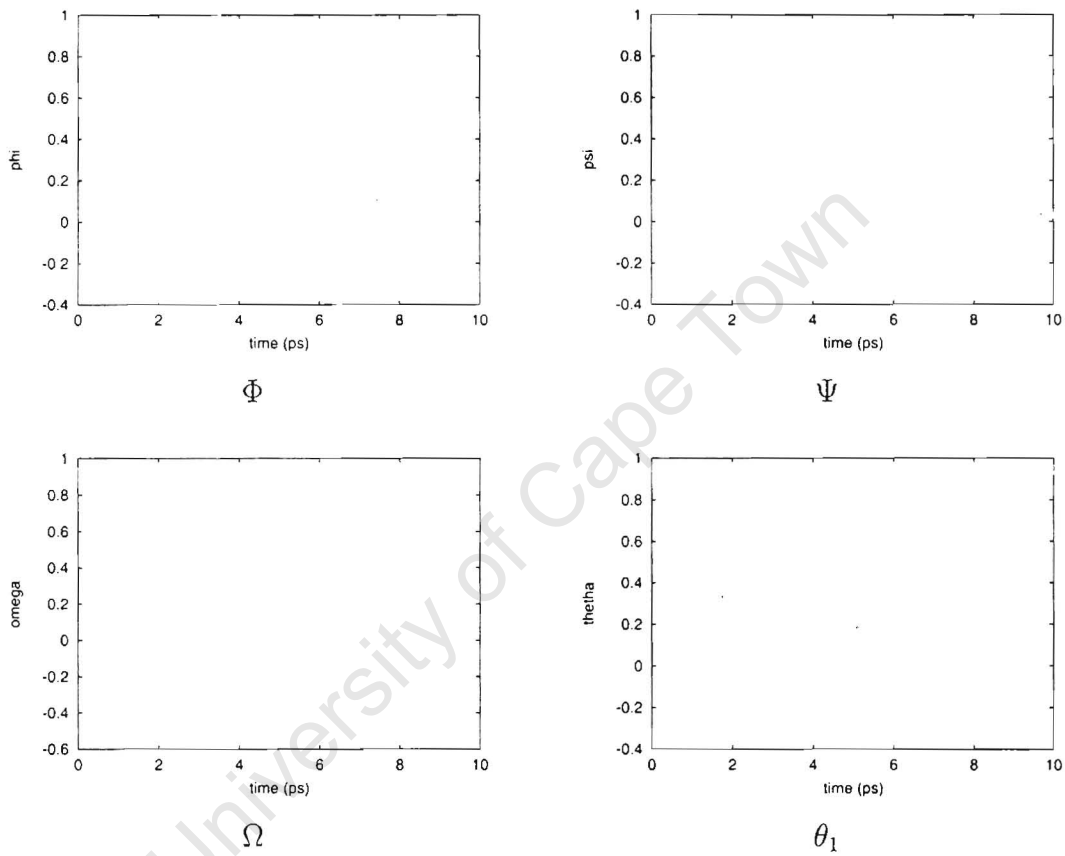
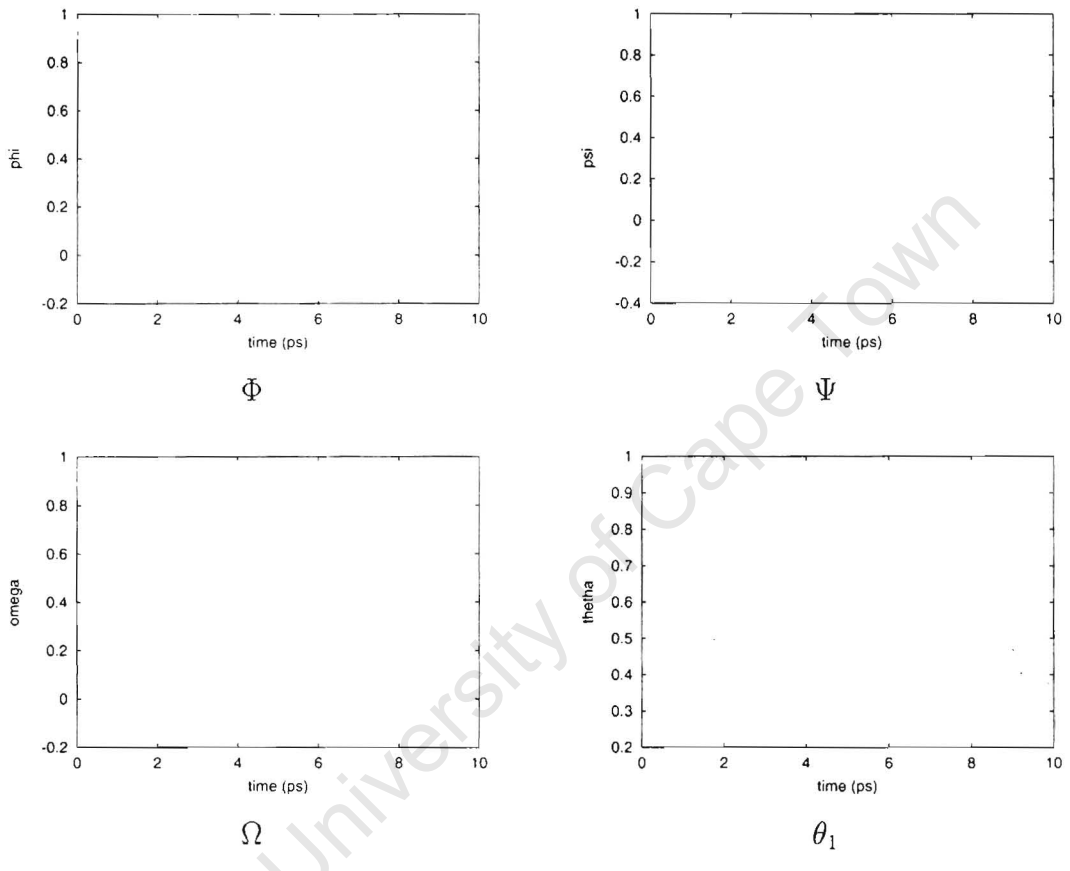


Figure A.10: Time correlation functions for isomaltose in vacuum from minimum B



Φ Ψ Ω θ_1

Figure A.11: Time correlation functions for isomaltose in vacuum from minimum C

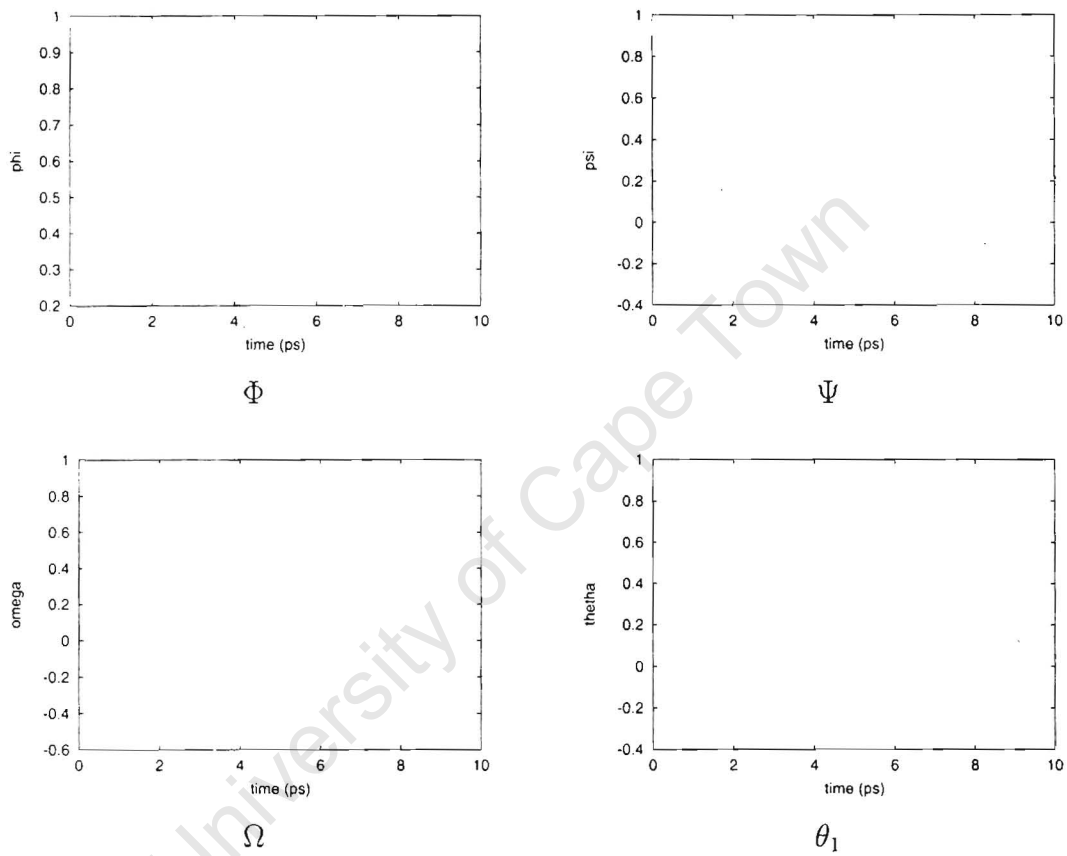
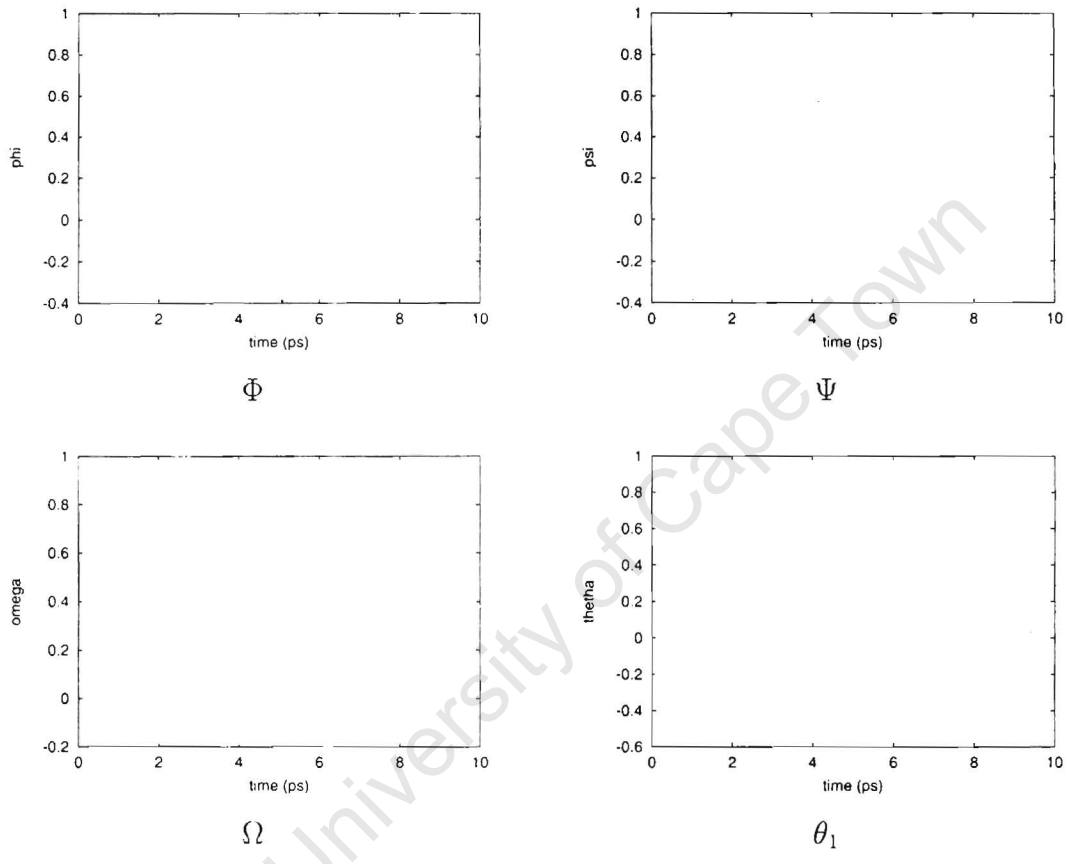


Figure A.12: Time correlation functions for isomaltose in vacuum from minimum D



Φ Ψ Ω θ_1

Figure A.13: Time correlation functions for isomaltose in vacuum from minimum E

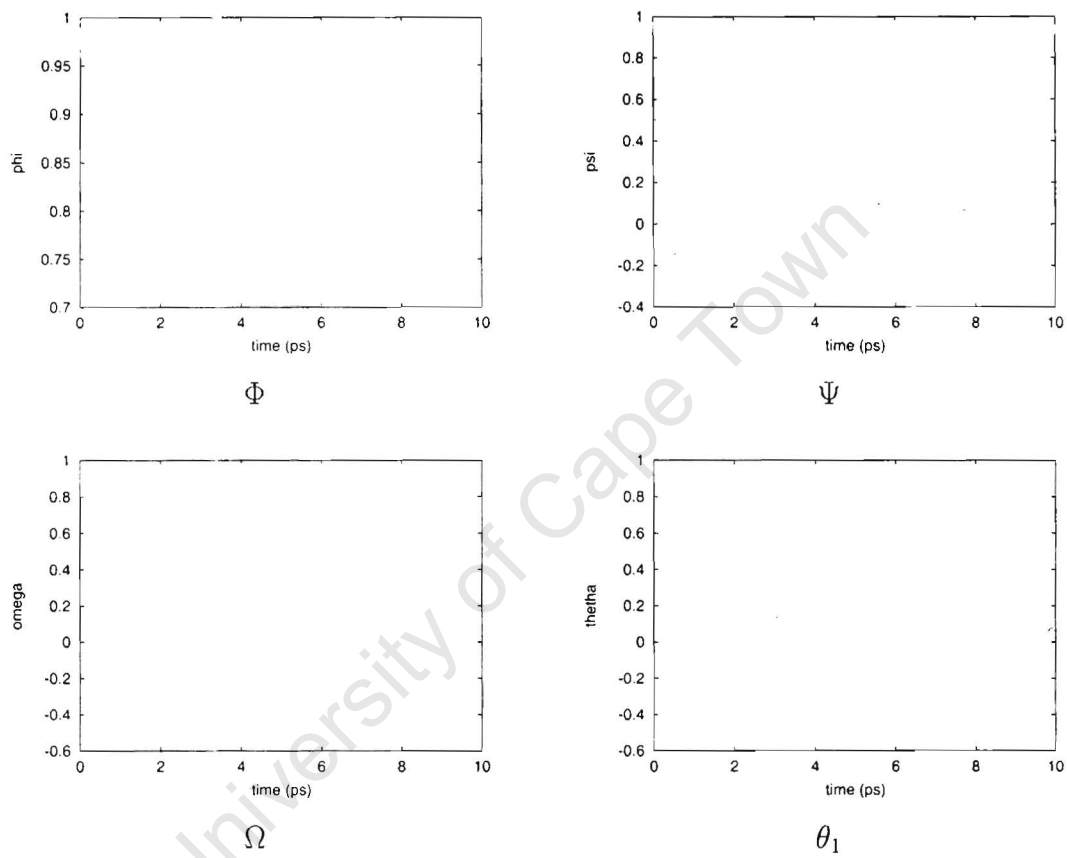


Figure A.14: Time correlation functions for isomaltose in vacuum from minimum F

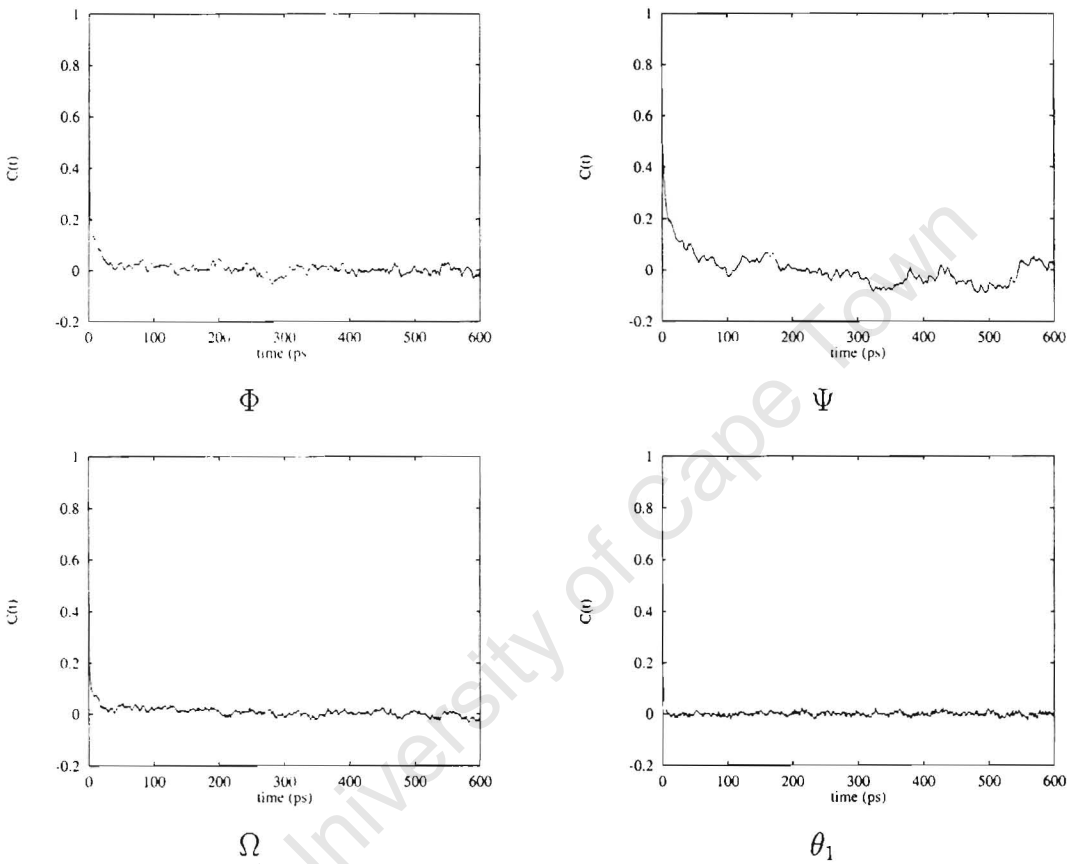


Figure A.15: Time correlation functions for isomaltose in solution from minimum A

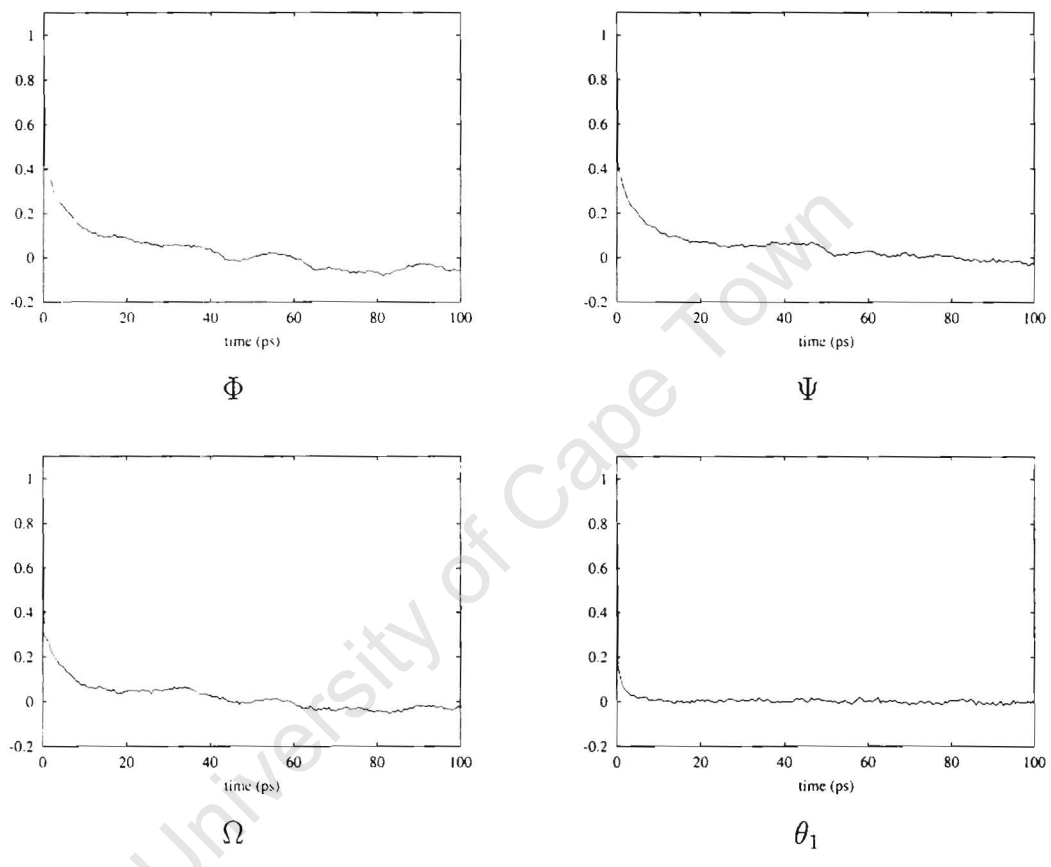


Figure A.16: Time correlation functions for isomaltose in solution from minimum B

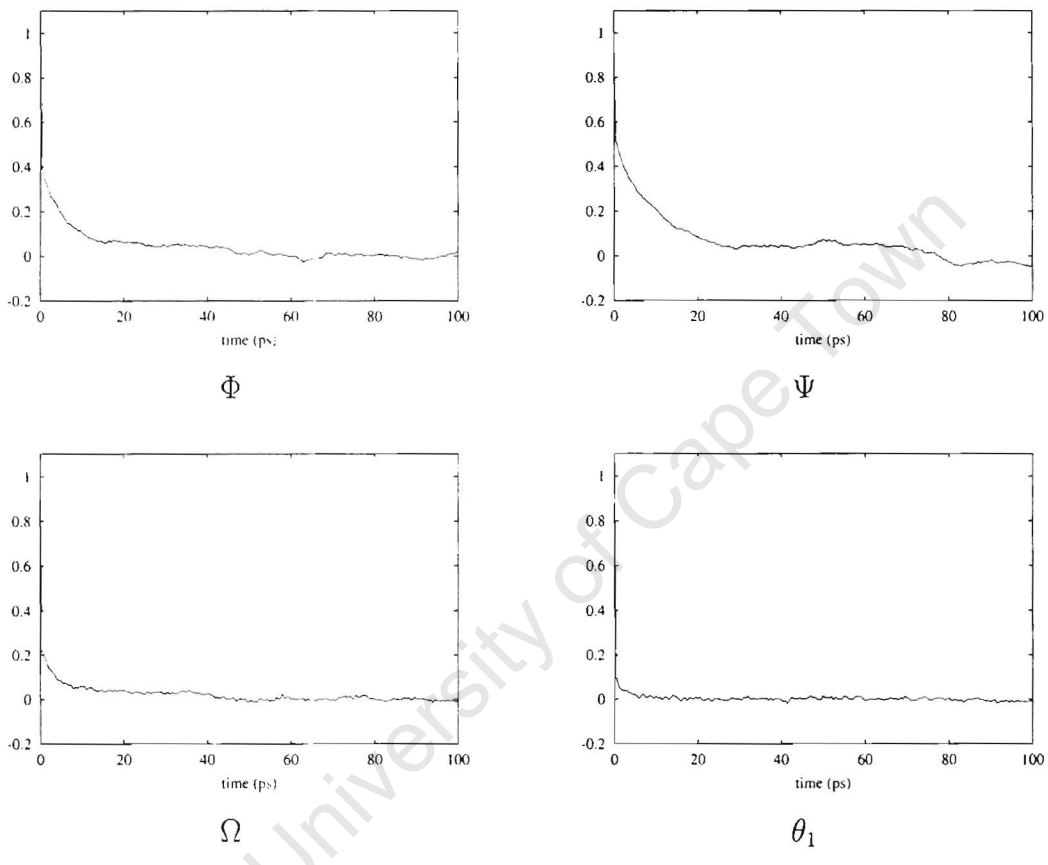


Figure A.17: Time correlation functions for isomaltose in solution from minimum C

Appendix B

Panose Dynamics Plots

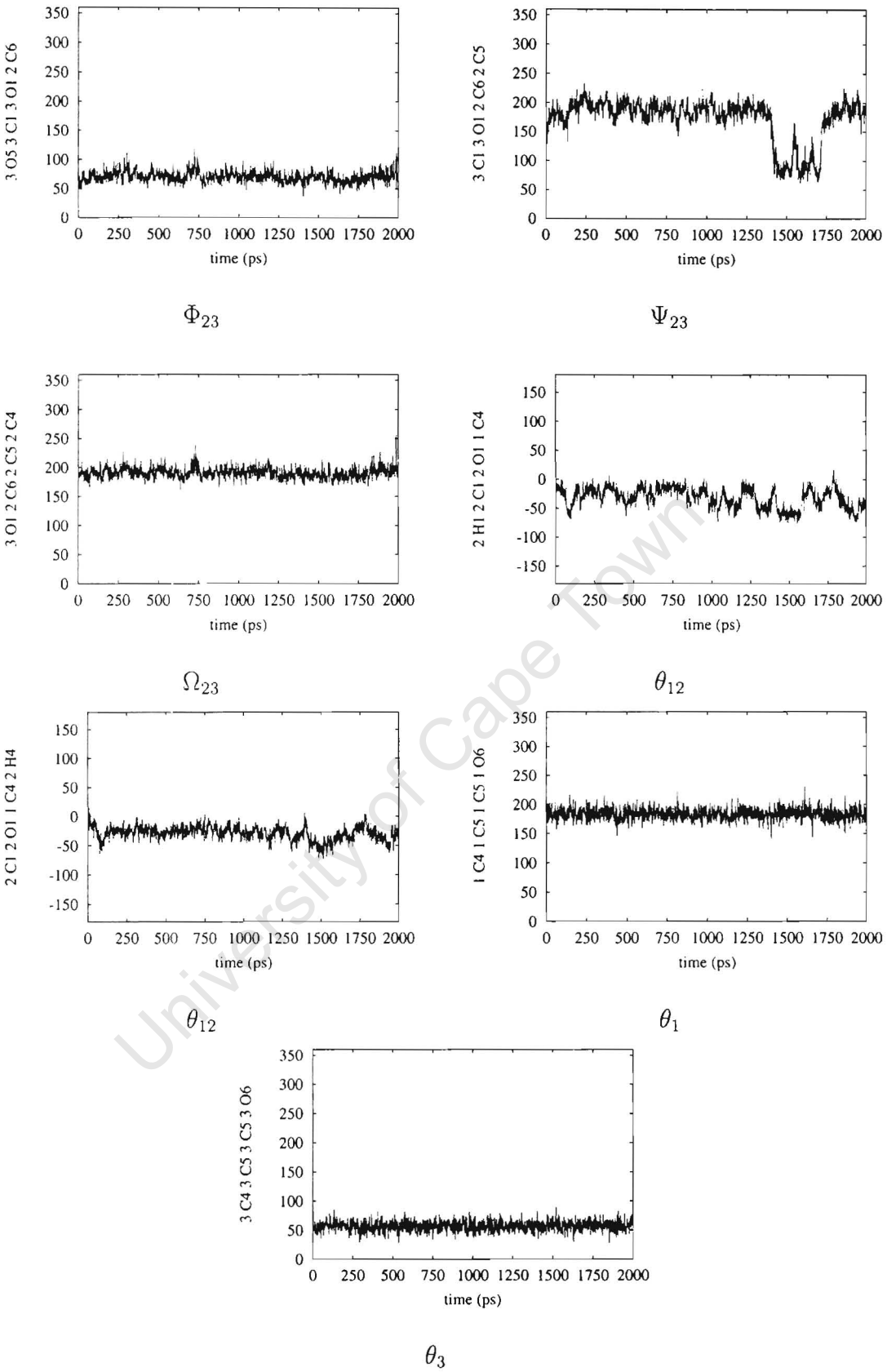
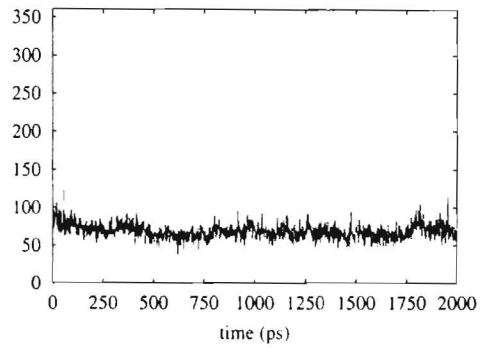
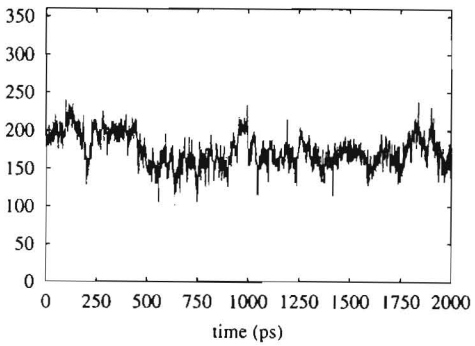


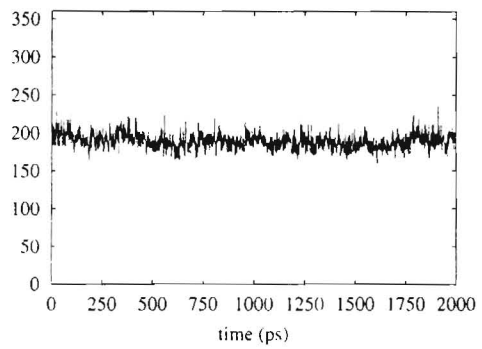
Figure B.1: Timeseries for Panose in solution starting from the crystal structure.



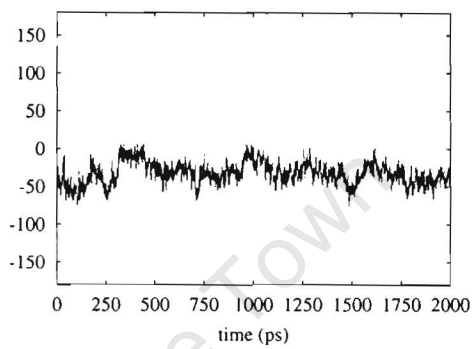
Φ_{23}



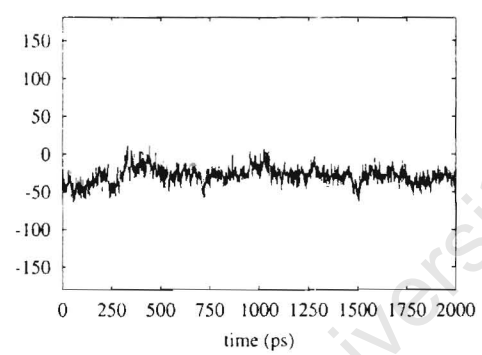
Ψ_{23}



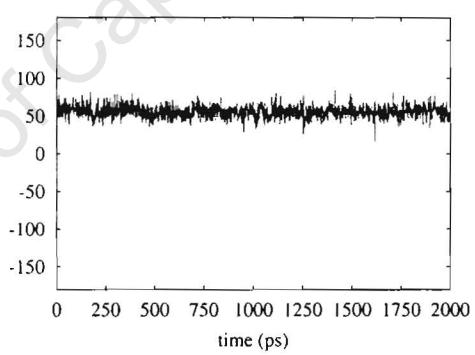
Ω_{23}



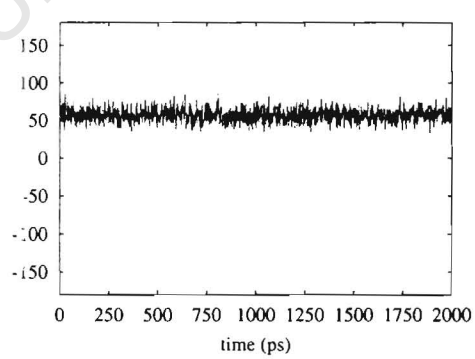
θ_{12}



θ_{12}



θ_1



θ_3

Figure B.2: Timeseries for Panose for second solution run.

Appendix C

Detailed hydrogen bond analyses for branch models

Hydroxyl	To Water	Bridging	Internal
1 OH1	2.25 (0.72)	0.34 (0.10)	0.09 (0.07)
1 OH2	1.94 (0.47)	0.12 (0.13)	0.18 (0.06)
1 OH3	1.30 (0.59)	0.15 (0.17)	0.74 (0.27)
1 O5	0.56 (0.17)	0.31 (0.10)	0.01 (0.06)
1 OH6	1.92 (0.54)	0.54 (0.21)	0.10 (0.13)
2 O1	0.04 (0.08)	0.03 (0.06)	0.06 (0.06)
2 OH2	1.23 (0.48)	0.20 (0.14)	0.79 (0.30)
2 OH3	1.05 (0.46)	0.21 (0.20)	0.34 (0.07)
2 OH4	0.89 (0.63)	0.08 (0.22)	0.17 (0.06)
2 O5	0.08 (0.22)	0.00 (0.00)	0.00 (0.00)
3 O1	0.13 (0.12)	0.08 (0.07)	0.00 (0.00)
3 OH2	1.14 (0.48)	0.50 (0.20)	0.29 (0.10)
3 OH3	1.00 (0.53)	0.21 (0.18)	0.34 (0.06)
3 OH4	1.26 (0.72)	0.14 (0.26)	0.15 (0.06)
3 O5	0.06 (0.17)	0.02 (0.08)	0.05 (0.08)
3 OH6	0.18 (0.29)	0.02 (0.07)	0.05 (0.08)

Table C.1: Panose Hydrogen Bonding Statistics for 500-1000ps of the first simulation, corresponding to clusters P1-P4. Main figures are the average number of hydrogen bonds, both as donor and acceptor, made by that sugar oxygen to the partner specified. Figures in brackets indicate average lifetimes in picoseconds.

Hydroxyl	To Water	Bridging	Internal
1 OH1	0.49 (0.39)	0.10 (0.10)	0.11 (0.06)
1 OH2	0.25 (0.27)	0.07 (0.18)	0.33 (0.06)
1 OH3	0.35 (0.36)	0.29 (0.19)	0.29 (0.12)
1 O5	0.08 (0.14)	0.07 (0.10)	0.02 (0.05)
1 OH6	0.69 (0.39)	0.03 (0.16)	0.03 (0.05)
2 O1	0.01 (0.06)	0.00 (0.00)	0.00 (0.00)
2 OH2	0.12 (0.22)	0.01 (0.09)	0.35 (0.09)
2 OH3	0.23 (0.33)	0.41 (0.13)	0.04 (0.13)
2 OH4	0.97 (0.49)	0.05 (0.12)	0.15 (0.06)
2 O5	0.18 (0.14)	0.13 (0.09)	0.00 (0.00)
3 O1	0.33 (0.16)	0.26 (0.10)	0.10 (0.07)
3 OH2	0.70 (0.31)	0.34 (0.17)	0.25 (0.07)
3 OH3	1.04 (0.45)	0.17 (0.18)	0.29 (0.06)
3 OH4	1.33 (0.49)	0.16 (0.15)	0.16 (0.06)
3 O5	0.29 (0.17)	0.10 (0.09)	0.05 (0.07)
3 OH6	1.59 (0.53)	0.10 (0.10)	0.05 (0.06)

Table C.2: Panose Hydrogen Bonding Statistics for 1450-1650ps of the first simulation, corresponding to cluster P5. Main figures are the average number of hydrogen bonds, both as donor and acceptor, made by that sugar oxygen to the partner specified. Figures in brackets indicate average lifetimes in picoseconds.

Appendix D

NMR Spectral Plots for Panose

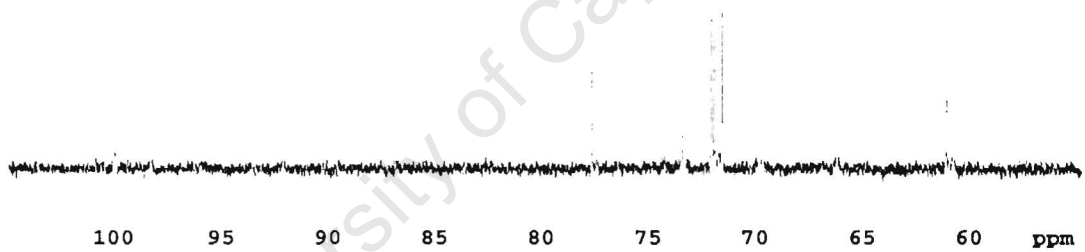


Figure D.1: ^{13}C Spectrum

Figure D.2: One-dimensional TOCSY with variable mixing time

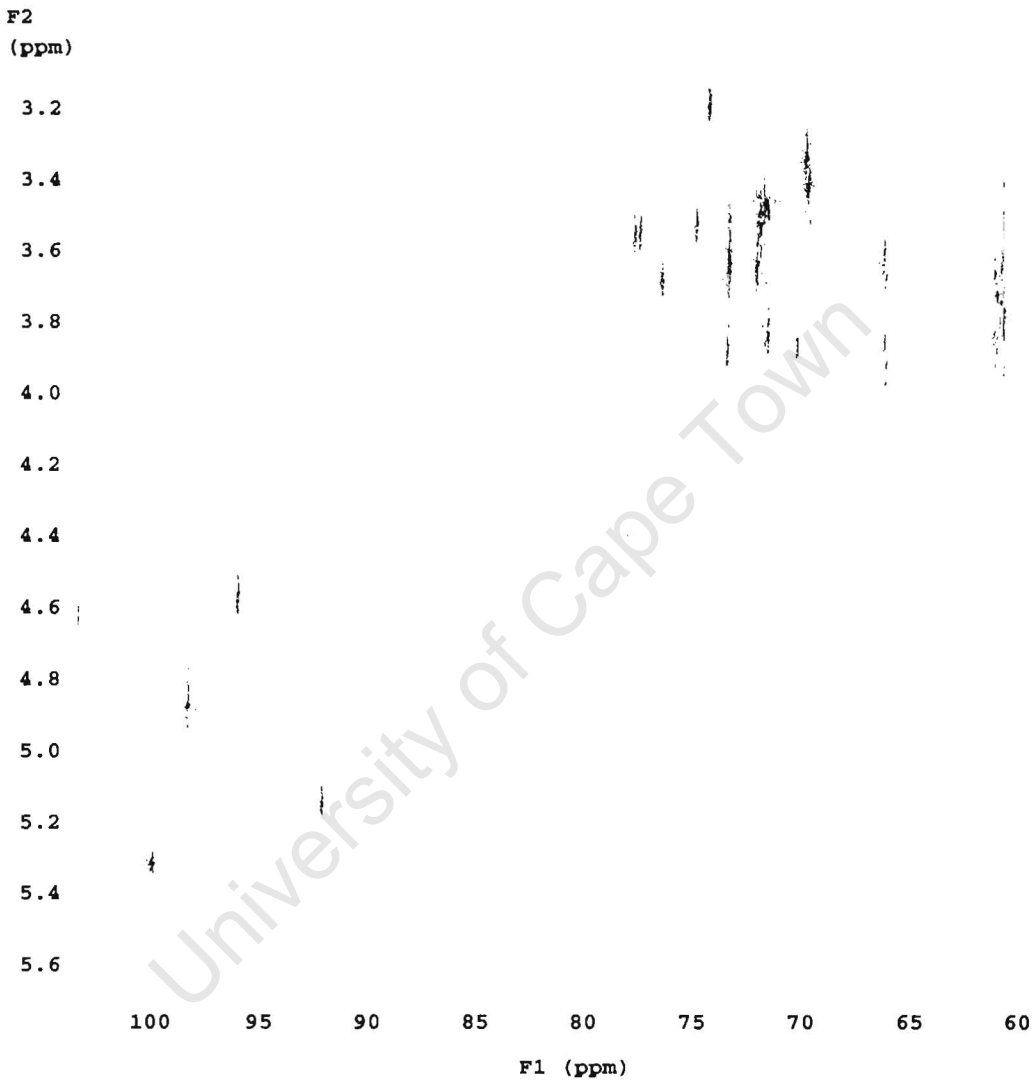


Figure D.3: Gradient HSQC

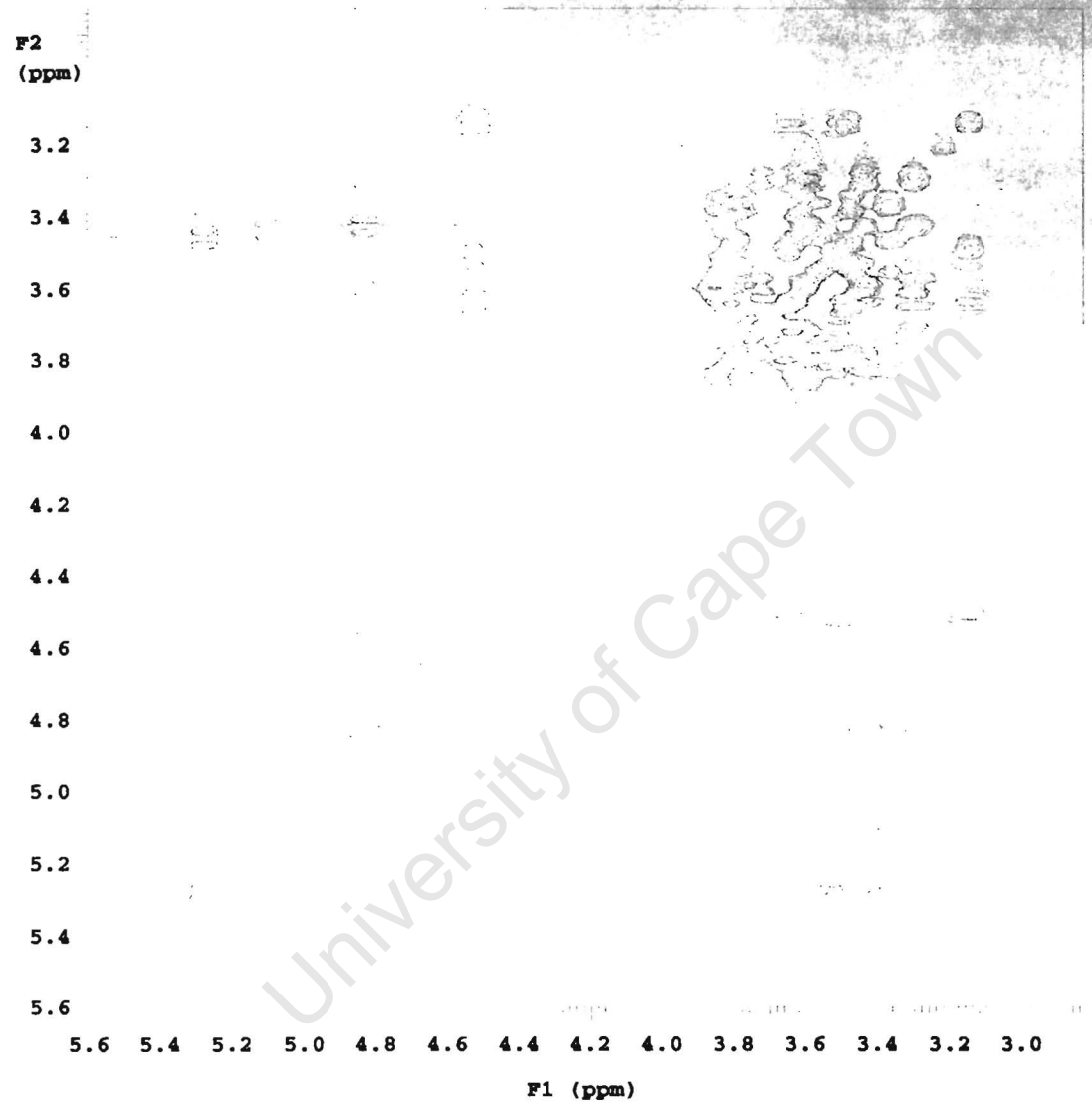


Figure D.4: Phase-sensitive TOCSY (mixing time = 50ms)