The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Simulations of Carbohydrate Conformational Dynamics and Thermodynamics

by
Michelle Mary Kuttel

A thesis submitted to the
University of Cape Town
in fulfilment of the requirements for the degree of
Doctor of Philosophy

University of Cape Town

June 12, 2003

Simulations of Carbohydrate Conformational Dynamics and Thermodynamics

Michelle Kuttel April 2003

Abstract

In this thesis, free energy calculations are employed to establish the conformational freedom of selected carbohydrates. The focus is on the biologically important (1→4)–linked glucans, although the $\alpha(1\rightarrow1)\alpha$ and $\alpha(1\rightarrow6)$ linkages are also investigated. Two principal types of potential of mean force (PMF) calculation are used: free energy calculations for rotations about dihedral angles and end–to–end distance free energy calculations for the extension and compression of oligosaccharide strands. In addition, a new method for simulating the atomic force microscopy stretching of single molecules is presented.

PMFs for rotation about the hydroxymethyl group in β–glucose and β–galactose are used to validate the reparameterisation of the hydroxymethyl group. Two–dimensional PMF surfaces for rotation about the glycosidic linkage dihedral angles are computed for three (1→4)–linked model disaccharides (maltose, dixylose and cellobiose) and two tetrasaccharides (maltotetraose and cellotetraose) as well as the $\alpha(1\rightarrow1)\alpha$–linked trehalose and the $\alpha(1\rightarrow6)$–linked isomaltose disaccharides. These PMF calculations are the first *complete* PMF calculations to be performed on glycosidic linkages in either solution or vacuum. Comparison of the vacuum and solution free energy surfaces shows that all the (1→4) linkages are affected by an aqueous solvent, whereas the $\alpha(1\rightarrow1)\alpha$ linkage remains unchanged. However, the PMF surfaces for the tetrasaccharides indicate that, while the maltose disaccharide is a good model for the $\alpha(1\rightarrow4)$–linkage, the cellobiose disaccharide is not for the $\beta(1\rightarrow4)$–linkage.

Oligosaccharide dynamics are explored using both end–to–end PMF calculations and stretching simulations. A key finding of this study is that ψ dihedral angle transitions in the $\alpha(1\rightarrow4)$ glycosidic linkage are more important to both the conformation and the dynamics of the amylose polymer than is currently thought. In particular, a complex helix–ribbon–helix conformational transition is identified as being primarily responsible for the elastic response of the amylose polysaccharide to a stretching force. Moreover, stretching simulations suggest that glycosidic linkage rotations also dominate the dynamic response of the cellulose ($\beta(1\rightarrow4)$–linked) and dextran ($\alpha(1\rightarrow6)$–linked) polysaccharides to a stretching force.

University of Cape Town

University of Cape Town

Acknowledgements

I would particularly like to thank my supervisor, Dr Kevin J. Naidoo, for his advice and enthusiasm throughout the course of my post graduate studies. I am very grateful for the interest he has shown not only in the work presented here, but also in my future career. I would also like to thank Prof. John W. Brady of Cornell University, for his academic guidance and financial support during my three month visit to Cornell and for subsequent helpful discussions.

Dr Neil Ravenscroft of the U.C.T Chemistry Department greatly assisted me by generously sharing his carbohydrate expertise. He was an extremely thorough proof-reader and made innumerable suggestions on style. I would like to thank him for this and also his well-stocked snack drawer.

My family provided invaluable support. I thank my husband, Peter Ouwehand, for proof-reading and discussions on mathematical issues, as well as his patience and humour with the length of the process. My parents, Sue and Ted Kuttel, never fail to provide enthusiatic encouragement and support. Thank you.

Finally, I am very grateful to the University of Cape Town for providing additional financial support, particularly in my final year. I would also like to extend my thanks to the National Research Foundation of South Africa for their financial support.

University of Cape Town

University of Cape Town

Publications and Presentations

Sections of this work have been published in the following article:

1. *Carbohydrate Solution Simulations: Producing a Force Field with Experimentally-Consistent Hydroxyl Rotational Frequencies*, M. Kuttel, J. W. Brady and K. J. Naidoo, *J. Comp. Chem.*, 23(13),1236-1243

and presented at the following conferences:

1. 6 March, 2003: Materials Modelling Meeting, University of the North, Sovenga, South Africa
   Simulations of polysaccharide dynamics: the role of conformational transitions in determining physical properties, M. Kuttel and K. J. Naidoo

2. 2 -14 Oct, 2000: Workshop on Methods for Macromolecular Modelling, New York University, New York City, U.S.A.
   Towards realistic primary alcohol rotational behaviour in carbohydrate models, M. Kuttel and K. J. Naidoo

3. 27 Oct -1 Sept, 2000: 20th International Carbohydrate Symposium, University of Hamburg, Hamburg, Germany.
   Towards realistic primary alcohol rotational behaviour in carbohydrate models, M. Kuttel and K. J. Naidoo

University of Cape Town

University of Cape Town

List of Abbreviations

AMBER Assisted Model Building and Energy Refinement

AFM Atomic Force Microscopy

CHARMM Chemistry at Harvard Macromolecular Mechanics

CSFF Carbohydrate Solution Force Field

DMSO Dimethylsulphoxide

fs femotosecond

FEP Free Energy Pertubation

GROMOS Gronigen Molecular Simulation program

HGFB Ha, Giammona, Field and Brady carbohydrate force field for CHARMM

HSEA Hard Sphere Exo-Anomeric approach

kcal kilocalories

MD Molecular Dynamics

MM Molecular Mechanics

NMR nuclear magnetic resonance

nOe nuclear Overhauser effect

ns nanosecond

PHLB Palma, Himmel, Liang and Brady carbohydrate force field for CHARMM

PMF Potential of Mean Force

pN picoNewton

ps picosecond

QM Quantum Mechanics

SAXS Small Angle X-ray Scattering

T_g Glass transition temperature

TI Thermodynamic Integration

WHAM Weighted Histogram Analysis Method

University of Cape Town

University of Cape Town

Contents

University of Cape Town

University of Cape Town

University of Cape Town

University of Cape Town

List of Figures

University of Cape Town

University of Cape Town

University of Cape Town

University of Cape Town

List of Tables

University of Cape Town

University of Cape Town

Chapter 1

Introduction

The carbohydrate class of molecules acquired their name when early elemental analysis of starch and sugars showed them to be "hydrates of carbon": they contain hydrogen and oxygen in the same proportions as in water ($C_xH_{2y}O_y$). The most abundant biopolymers on the planet, carbohydrates are uniquely diverse and perform a plethora of biological roles, ranging from food storage through structural support to molecular recognition. The simplest of the carbohydrates are the monosaccharides. These are vital basic cellular food sources — glucose is the primary fuel for biological cells — but are also involved in water control in many drought– and cold–resistant organisms. Monosaccharides link together in a number of ways to form oligo– and polysaccharides. The most biologically abundant are the polyglucoses, of which amylose and amylopectin (the food polymers comprising starch) and the structural polysaccharide cellulose, are the most familiar examples. A wealth of different of oligo– and polysaccharides are found physiologically, where they are often in the form of complexes (glycoconjugates) with other classes of molecules, such as proteins (glycoproteins) or lipids (glycolipids). Glycoconjugates play physically supportive, protective and lubricatory roles in tissues and fluids. In addition, the outer edge of cell membranes exhibits a variety of polysaccharides, glycolipids and glycoproteins that are involved in complex cellular functions, such as cell–cell adhesion, recognition processes and immunological protection.

A desirable goal in the characterisation of all biomolecules is elucidation of the relationship between molecular structure and biological function. This is particularly important for the carbohydrates, which display exceptional conformational variety and flexibility. Only by acquiring knowledge of the mechanics of carbohydrates on a microscopic level can we hope to understand, and perhaps alter, their properties. A key aspect of carbohydrate chemistry is the specific interactions they have with water through their hydroxyl groups. Many carbohydrates function in an aqueous environment and solvation effects

University of Cape Town

are known to play an important role in determining their conformation and properties. Carbohydrates in the solid state also interact strongly with water; water is present in many carbohydrate crystals and water content is known to affect the physical properties of carbohydrate glasses.[1–4]

Experiment has thus far provided limited information on carbohydrate structure and dynamics, particularly for oligo– and polysaccharides in solution. Studies have lagged well behind those of proteins and nucleic acids and thus the molecular basis for many of the fascinating variety of tasks performed by the carbohydrates is still imperfectly understood. A typical example is amylose, one of the first described and most investigated of the carbohydrate polymers. Amylose is a structurally homogeneous linear polysaccharide composed of glucose monomers joined by $\alpha(1\rightarrow4)$ glycosidic linkages. Together with its $\alpha(1\rightarrow6)$ branched partner, amylopectin, amylose comprises the ubiquitous carbohydrate mixture known as starch. The varying properties of starch under different conditions provides the basis for its variety of culinary uses, particularly as a thickening and gelling component.[5] In addition, high–amylose cornstarch has potential for industrial use as a biodegradable plastic.[6–9] However, the range of applications in which it can be used are limited by the brittleness of aged starch glasses[10] and the dramatic effect of water on this material.[8,11–14] Although amylose is known to occur in various helical crystalline and semi–crystalline forms, only one crystalline form, V–amylose, has been fully characterised.[15–17] Moreover, the structure of amylose strands in solution, though thought to be helical, remains to be established.[18]

The solid state structures of carbohydrates are typically investigated by X–ray diffraction or neutron scattering techniques, with fibre diffraction being the method of choice for crystal structure determination of the less ordered polysaccharides. The crystal structures for many mono– and disaccharides[19–24] and some tri– and tetrasaccharides[25,26] have been determined. However, with a few conformationally–restricted cyclic exceptions, such as the cyclomaltoses[17,27–29] and the less flexible cellulose,[30] oligo– and polysaccharides have proved notoriously difficult to crystallise. In addition, results from fibre diffraction studies of polysaccharides have often proved ambiguous.[16] Even when available, crystal structures are of limited value in that they at best provide a precise structural description of a carbohydrate's conformation in the crystal environment, and are of little use as a probe of a flexible carbohydrate's dynamic behaviour or solution structure.

In solution, the conformation of carbohydrates has been principally investigated by the use of ^{1}H– and ^{13}C–NMR spectroscopy,[31–41] though a number of other methods, such as complex formation and optical rotation, have also been employed.[42–44] NMR coupling constants, both homo– ($^3J_{H,H}$) and heteronuclear ($^3J_{H,C}$), and nuclear Overhauser effects

University of Cape Town

(NOE) are typically used for conformational analysis. Scalar ^3J coupling constants involve three bonds and can be used to infer bond torsion angles. However, scalar coupling methods have limitations for conformational analysis, especially for large molecules such as oligo– and polysaccharides, as there are no proton–proton couplings across the glycosidic linkages. Dipolar coupling (from which the NOE parameters, T_1 and T_2 relaxation times are determined) gives information about intermolecular distances and torsion angles. In the case of oligosaccharides, conformational analysis is based largely on intra– and interresidue NOEs. However, the number of measurable NOEs which can be observed relative to the degrees of freedom in the molecule is, in most cases, insufficient to establish a unique solution conformation.[45] In addition, the conformational flexibility of glycosidic linkages results in the difficulties experienced with NMR and many of the other experimental probes of their solution structure. Carbohydrates can undergo numerous transitions in solution and interconversion between major conformational states of oligo– and polysaccharides is estimated to take place on the nanosecond time scale.[46] As few experimental techniques probe this time scale, this means that the observed signal of many spectroscopic or electrochemical techniques often represents a "virtual" structure, comprising the dynamic average of several structures.[31,47–49]

Two new techniques for the analysis of polysaccharide structures – scanning probe microscopy (SMP) and small–angle X–ray scattering (SAXS)– are becoming increasingly important.[50] Atomic force microscopy has been recently used to image a variety of polysaccharides[18] as well as to investigate the mechanical response of single polysaccharide molecules to tensile stress.[51–53] Small–angle X–ray scattering has provided a probe of polysaccharide structure in a range (2 to 25 nm) intermediate between NMR and light scattering. In addition, a recent advance in NMR techniques is the application of dilute crystalline phases (bicelles) as solvents. This enables determination of through–space magnetic dipole–dipole (DD) interactions which provides a powerful tool for analysis of saccharide molecular structure in ordered phases.[54,55]

Carbohydrate dynamics in solution includes both the interconversion between conformers and the association of carbohydrate solute molecules with each other. Relatively slow dynamic processes (seconds or longer) provide visible separate resonances in standard NMR spectra. Fast processes, with correlation times of approximately 10^{-11} to 10^{-10} s, have been investigated using NMR relaxation observables: T_1, T_2, NOE.[34,56,57] However, spectral complexity limits the usefulness of low frequency NMR for oligo– and polysaccharides. Other experimental techniques, e.g. infrared–,[58] fluorescence–, mechanical–[59] and dielectric[3,60] spectroscopy, may be used to examine motions in intermediate time regimes between 10^{-10} and 10^{-3} s. These methods are often complementary to NMR. For

University of Cape Town

example, dielectric techniques have a larger dynamic range than NMR and can be used on glassy samples, whereas NMR permits the assignment of particular motions. Broadband ultrasonic absorption measurements have recently provided insight into the conformational flexibility, pseudorotation and *exo*–cyclic group isomerisation of monosaccharides in water.[61,62] In addition to solution studies, mechanical[59] and dielectric spectroscopy[3] as well as solid state NMR[63] have been used to investigate the conformational dynamics and molecular mobility in polysaccharide glasses.

As structural information on carbohydrates remains scarce, computational studies are being increasingly employed to investigate the conformation and mobility of oligo– and polysaccharides. These can serve both to supplement experimental data, where an accurate model can be vitality important for interpreting results, and as an independent probe of microscopic interactions.[35,64] Simple Molecular Mechanics (MM), or the more sophisticated *ab initio* minimisations can be used to predict the preferred structure of a carbohydrate. In the case of disaccharides, spectroscopic data is usually supplemented by MM calculations of the potential energy surfaces[31,42–44,47,48,64–66] in order to give an estimate of different population distributions. Increasingly, Molecular Dynamics (MD) simulations are being used to interpret information from NMR experiments.[46,57,67–73] MD simulations are particularly useful for investigating the hydration of a carbohydrate at the molecular level[67,70,74–81] and can provide mechanistic information unobtainable by current experimental methods. Numerous simulation studies have investigated various properties of saccharides, including conformational flexibility,[57,74,82–87] conformational free energy differences,[88–91] the anomeric effect,[92] hydrogen bonding in saccharide solutions[75,79,81] and the dielectric effect.[93]

University of Cape Town

This thesis is concerned with a computational exploration of the physical conformations and properties of carbohydrates, particularly glucose–based carbohydrates, or glucans. The next section provides some background information on carbohydrates: their composition, structure and conformations. This is followed by an overview of the current state of research into carbohydrate dynamics and interactions with water. Section 1.4 gives a brief introduction to carbohydrate simulations. The final two sections contain a description of the objectives of the study and an overview of the contents of the rest of the thesis.

1.1 Carbohydrate Conformation

1.1.1 Monosaccharide Structure and Conformation

The simplest carbohydrates, the monosaccharides, are classified according to the number of carbon atoms they possess. For example, the *trioses*, *tetroses*, *pentoses* and *hexoses* have 3, 4, 5 and 6 carbon atoms respectively. Naturally occurring monosaccharides with more than six carbons are comparatively rare, the most common monosaccharides being the pentoses and hexoses. Monosaccharides possess at least one asymmetric carbon and their various stereoisomers have distinct physical and biological properties. In aqueous solution, the flexible open–chain forms of the hexoses typically undergo a cyclisation reaction between the carbonyl group and one of the hydroxyl groups to form a hemiacetal. This reaction creates a five– or six–membered ring with an additional chiral atom at C_1, termed the anomeric carbon. The two possible anomers at C_1 are designated α (O_1 in an axial conformation) and β (O_1 in an equatorial conformation). Cyclisation of a monosaccharide is easily reversible and monosaccharides in aqueous solution exist in a dynamic equilibrium between a variety of different chain and ring forms, with the lower–energy cyclic forms predominating. Aldohexoses, such as glucose, in water are typically in equilibrium between the α– and β–anomers of the pyranoid ring.[61]

The Anomeric Effect

In pyranoses, the anomeric effect[94,95] refers to the tendency of an electronegative group at C_1 to have an increased population of the axial (α–D) orientation at the expense of the sterically–favoured equatorial (β–D) orientation. The anomeric effect can be explained in terms of electrostatic interactions (the equatorial β–anomer is destabilised by unfavourable dipole–dipole interactions)[94–96] and negative hyperconjugation (the axial α–anomer is stabilised by delocalisation of the free electron pair on the ring oxygen into the antiperiplanar $\sigma^*_{C_1-O}$ orbital).[97–99] For glucose, the anomeric equilibrium in water is 64% β and 36% α,[94–96] although the precise ratio is solvent and temperature dependent.[100] This corresponds to a ($\alpha \rightarrow \beta$) free energy difference in water of -0.3 kcal/mol. Solvation is thought to play a role in determining this anomeric ratio by preferentially stabilising the equatorial position,[99,101] though whether this involves specific hydrogen bonding interactions or is dominated by electrostatic effects remains under debate.[102] Computational calculations have attempted to resolve which effects decide the anomeric ratio. To date, several quantum mechanical (QM) studies have focused on the ($\alpha \rightarrow \beta$) free energy difference in glucose.[102–107] The most recent of these studies, by Barrows et al., calculated

University of Cape Town

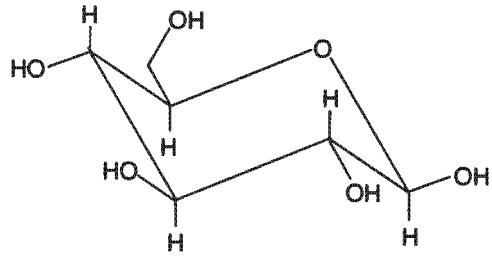


Figure 1.1: The $^4C_1$ conformation of β-D-glucopyranose.

a free energy difference ($\alpha \rightarrow \beta$) in water of -0.2 kcal/mol, with solvation favouring the β anomer by 0.6 kcal/mol.[106] An earlier study by Cramer and Truhlar found no solvation effect for the anomeric free energy difference.[102] However, though they can be very accurate, computational constraints restrict QM calculations to a limited set of conformations without the inclusion of explicit solvent molecules.

A variety of Molecular Mechanics studies have attempted to replicate the experimental ($\alpha \rightarrow \beta$) free energy difference in glucose,[108–110] and other monomers such as D–mannose[92] and D–xylose.[111] Some studies investigated simpler molecules with fewer rotatable bonds, such as 2–methoxytetrahydropyran,[112,113] as models of the anomeric effect in carbohydrates. Results vary depending on the simulation conditions and the force fields employed. Ha et al. found the α–anomer to be more stable than the β–anomer by 0.31 kcal/mol, with aqueous solvation favouring β by 3.6 kcal/mol.[108] In contrast, calculations by van Eijck et al. found the ($\alpha \rightarrow \beta$) free energy difference to be -0.86 kcal/mol, with solvation favouring the α–anomer by 0.5 kcal/mol.[109] Convergence of these free energy calculations is a problem, as there is a small free energy difference between the anomers and a large number of potentially accessible conformations that must be explored.

University of Cape Town

Pyranoid Ring Conformation

Each of the various carbohydrate ring forms has a variety of possible conformations. For example, glucopyranose has 6 distinct boat forms, 6 twist–boat forms and 2 chair forms. A common classification system represents the ring conformation with a capital letter, using numerals to distinguish between the different forms of each shape: e.g. $^4C_1$ denotes a chair conformation in which the C_4 lies above and the C_1 below the plane of the pyranose ring (see Figure 1.1). A more rigorous definition of ring puckering coordinates was provided by Cremer and Pople in 1975.[114] For six–membered rings, such as glucose, a coordinate system is defined which maps all types of puckering onto the surface of a sphere. The polar positions correspond to chair conformations, and the equatorial positions to various twist–boat forms.

The preferred conformation of a pyranose ring is determined by non–bonded interactions (Van Der Waals and electrostatic interactions) as well as solvent effects, with intramolecular hydrogen bonds also playing a role. In general, the chair conformations of the pyranose ring are favoured, because the boat and skew forms cause crowding of the atoms in the ring and are thus significantly higher in energy. Pyranose rings have two distinct chair conformations, $^4C_1$ and $^1C_4$. These flexible chair conformations are entropically stabilised because a range of chair distortions can occur without a significant change in energy.[115,116] The $^4C_1$ conformation of α–D–glucose, with all ring substituents in the equatorial position, is preferred over the $^1C_4$ conformation in which all substituents are in the axial orientation. The $^4C_1$ conformation is found in crystal structures of α–D–glucose,[117] β–D–glucose,[23] various other pyranose mono– and disaccharides and the cyclodextrins. In addition, optical rotation and ultrasonic relaxation experiments have shown that in aqueous solution the pyranose ring persists in the $^4C_1$ conformation, both in simple sugars and in larger polysaccharides.[42,62] Intermediate between the two chair conformations are a variety of stable boat conformers. These have energies 5 to 8 kcal/mol above the energy of $^4C_1$ and are separated from the chair conformations by energy barriers of approximately 12 kcal/mol.[118,119]

Interestingly, in certain polysaccharides the pyranose ring has recently been proposed to undergo a transition from a chair to a twist–boat conformation when subjected to stretching forces.[52,53,119–121] This suggested transition is dependent on the presence of axial glycosidic bonds. *Ab initio* calculations on three pyranose monomers, α– and β–glucose[119] and α–galactose,[122] have shown that the O_1O_4 vector increases in length for the chair–to–boat conversions in α–glucose and the chair inversion in α–galactose,[122] while there is no increase in length for any ring inversion in β–glucose. Therefore, stretching the pyranose ring with an external force along the O_1O_4 vector is expected to favour boat conformations of the glucose ring, as these provide greater separation of the ring oxygens and will lead to greater extension of the molecule. Moreover, calculations have shown the resultant twist–boat conformations to be energetically stable minima.[118] Polysaccharide stretching experiments are discussed in more detail in Section 1.3.1.

Hydroxymethyl Orientations

The orientation of the *exo*–cyclic 6–hydroxymethyl group, also termed the primary alcohol (Figure 1.2), affects the relative energies of the different ring conformations in hexoses. In addition, dynamic rotations of this group are thought to play an important role in relaxation processes that occur in a variety of hexopyranose carbohydrate glasses, from

University of Cape Town

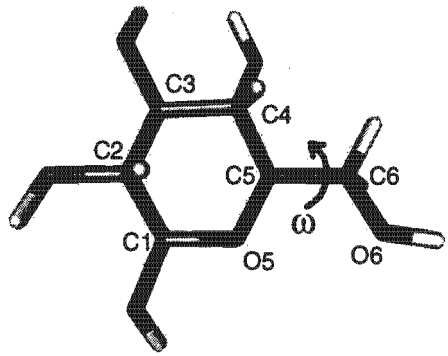


Figure 1.2: Three–dimensional structure of a β-D-glucose molecule in the $^4C_1$ conformation, showing the primary alcohol dihedral angle, ω.

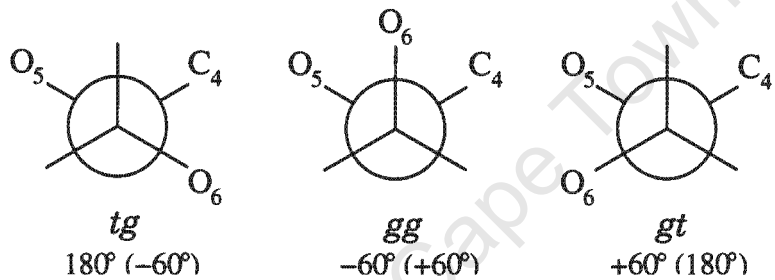


Figure 1.3: Newman projections showing the definitions of the three minimum–energy hydroxymethyl positions for a pyranose ring. Values for the ω dihedral angle (O_5–C_5–C_6–O_6) are shown below each conformation, with the corresponding ω^* values in brackets.

glucose to cellulose.[3,4,59,60]

The orientation of the hydroxymethyl group in relation to the pyranoid ring is usually described by either of two dihedral angles:

$$\omega = O_5 - C_5 - C_6 - O_6 \tag{1.1}$$

or;

$$\omega^* = C_4 - C_5 - C_6 - O_6 \tag{1.2}$$

Here the first definition (ω) is used. Generally, the hydroxymethyl group adopts one of three staggered orientations. These orientations are designated *tg*, *gg* and *gt* respectively, where the first letter refers to the *trans* or *gauche* orientation of the ω dihedral and the second the orientation of the ω^* dihedral. The three hydroxymethyl orientations and their corresponding ω values are illustrated in Figure 1.3.

The rotameric distribution of the hydroxymethyl group for different monosaccharides has proved difficult to establish experimentally. Conformational preferences have been determined chiefly from X–ray crystallography and NMR studies, and to some extent from

University of Cape Town

optical rotation experiments.[123] While there is variation in the reported experimental rotameric distributions, the *gauche* conformations (*gg* and *gt*) are known to be favoured over the *trans* (*tg*) conformation. The overall *gg*:*gt*:*tg* population rankings in aqueous solution are estimated to be approximately 6:4:0 for glucopyranosides and 2:6:2 for galactopyranosides.[123–125] Specifically, Bock et al. report α–D–glucose to have a distribution of *gg*:*gt*:*tg* = 52:41:76 and α–D–galactose a distribution of *gg*:*gt*:*tg* = 12:56:32.[123] Earlier studies by Nishida et al. give a ratio for α–D–glucose of *gg*:*gt*:*tg* = 53:45:2[124] and α–D–galactose of *gg*:*gt*:*tg* = 18:61:21.[125]

The conformational preferences of the pyranose hydroxymethyl group in solution are assumed to be determined by solvation effects, stereoelectronic effects (the 1,3–diaxial interaction and *gauche* effect*) and hydrogen bonding.[123] However, there is no consensus on the relative importance of these effects and opinion remains divided as to which influence predominates. Experimental studies tend to discount the effects of hydrogen bonding. Rockwell and Grindley investigated the solvent dependence of hydroxymethyl rotamer populations by analysing NMR $^3J_{H5,H6R}$ and $^3J_{H5,H6S}$ coupling constants of fully methylated glucose and glucose methylated at all positions except for the primary alcohol in solvents of varying polarity.[126] As the aqueous rotameric populations of the two molecules were identical, they concluded that the hydroxymethyl conformational preferences in water are due to a polarisation effect on the molecule, rather than an explicit hydrogen bond effect. A similar study of galactose derivatives by De Vries and Buck concluded that the differences in rotamer populations were due to stereoelectronic effects.[127]

In contrast, some theoretical studies have emphasised the importance of hydrogen bonding. Cramer and Truhlar concluded from semi–empirical calculations of D–glucopyranose conformers in a continuum solvent that the hydroxymethyl conformational equilibrium is dominated by solute–solvent hydrogen bonding interactions, with unfavourable solvation of the *tg* conformer destabilising it relative to the *gg* and *gt* conformations.[102] Molecular orbital calculations of carbohydrate model compounds by Tvaroška et al. indicate that *gauche* preference of the hydroxymethyl group is due to the presence of hydrogen bonding and not the *gauche* effect.[112,128] However, a recent theoretical study by Kirschner and Woods[129] suggests that water plays a central role by disrupting the intramolecular hydrogen bonding in both glucose and galactose, allowing the rotamer populations to be determined by internal electronic interactions and steric repulsions. In conclusion, the relative degree of importance of the various influences on the hydroxymethyl population

*The *gauche* effect is defined as the tendency for a molecule to adopt that structure which has the maximum number of synclinal (*gauche*, ±60) interactions between adjacent electron pairs and/or polar bonds in a molecular fragment X–C–C–Y, where X and Y are two electronegative substituents.

University of Cape Town

in solution remains undecided.

1.1.2 Disaccharide Conformation

Disaccharides are formed by a condensation reaction between two monosaccharides. A glycosidic bond joins the anomeric carbon on one monosaccharide with any one of the hydroxyls of the second monosaccharide. Thus, glucopyranose disaccharides may have linkages that connect C_1 to C_1', C_2', C_3', C_4' or C_6' in either the α or β form, resulting in 11 distinct linkages. If the disaccharide maintains a free lactol (i.e. hemiacetal hydroxyl) group, it is a reducing disaccharide. For example, the glucopyranose disaccharide cellobiose is connected from the β C_1 carbon in one monosaccharide to the C_4 carbon of the second monosaccharide to constitute a $\beta(1 \rightarrow 4)$ linkage and is a reducing disaccharide. If there is no free lactol group, it is non–reducing (e.g. the disaccharides sucrose or trehalose). Primed labels (e.g C_5', O_6') are used to distinguish the atoms belonging to the reducing residue in a disaccharide from those in the nonreducing residue.

Disaccharide conformation is dominated by the glycosidic linkage conformation. The position and stereochemistry of the linkage governs the extent of interaction between successive sugar residues, which in turn determines the preferred conformations and relative freedom of a given linkage. The relative orientations of the two residues involved in a two–bond glycosidic linkage can be defined completely by two torsion angles, ϕ and ψ (Figure 1.4):

$$\phi = H_1 - C_1 - O_1 - C_x' \tag{1.3}$$

$$\psi = C_1 - O_1 - C_X' - H_x' \tag{1.4}$$

where C_X' and H_X' are the aglyconic atoms. The definitions for ϕ and ψ are analogous to ϕ_H and ψ_H in IUPAC convention. In this thesis, a further dihedral, τ, is defined as:

$$\tau = H_1 - C_1 - C_X' - H_x' \tag{1.5}$$

This non–IUPAC dihedral angle describes rotation about a virtual C_1–C_X bond. The value of τ is simply the sum of ϕ and ψ and it is thus a one–dimensional measure of the relative orientation of the glucose subunits involved in the glycosidic linkage.

In the case of C_6–linked disaccharides, the three bond torsion angles (ϕ_6 and ψ_6 and ω_6) are here defined as:

$$\phi_6 = O_5 - C_1 - O_1 - C_6' \tag{1.6}$$

$$\psi_6 = C_1 - O_1 - C_6' - C\prime_5 \tag{1.7}$$

$$\omega_6 = O_1 - C_6' - C_5' - O_5' \tag{1.8}$$

University of Cape Town

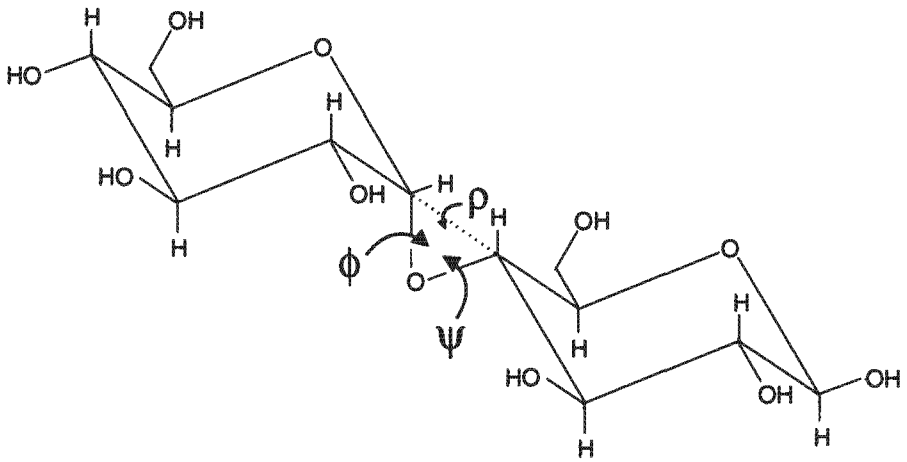


Figure 1.4: A line diagram of the β–maltose disaccharide (4-*O*-α–D–glucopyranosyl–β–D–glucopyranose), showing the ϕ, ψ and τ torsion angles.

in keeping with existing definitions.[57]

The preferred orientations of the glycosidic linkage torsion angles for a particular linkage determine the relative orientations of the monomers and thus the overall conformation of the disaccharide. The ϕ–dihedral angle conformation is governed by the anomeric and *exo*–anomeric effect, which results in a principal conformation with ϕ close to 0° or 180°, depending on the anomeric conformation and the absolute configuration of the sugar residue. The *exo*–anomeric effect has the same stereoelectronic origin as the anomeric effect (page 5), but it influences the orientation of the aglycone; it describes the preference for a *gauche* conformation of the aglycone O–R bond with respect to the *endo*–cyclic *C*–*O* bond, regardless of whether the aglycone is axial or equatorial. These stereoelectronic effects do not influence the ψ angle, which therefore has freer rotation than the anomeric *C*–*O* bond. In general, the ψ angles have values around 0°, resulting in a *syn* relationship between sugar residues. However, there is evidence that the *anti*–conformers (ψ=180°) may be slightly populated in some disaccharides.[46,66,73,130]

Steric clashes between the *exo*–cyclic groups, particularly the more bulky hydroxymethyl groups, can limit the conformational space and thus affect the preferred glycosidic linkage conformations in disaccharides. Theoretical calculations in the form of Ramachandran, or adiabatic, contour maps of the conformational energy as a function of the glycosidic torsion angles, ϕ and ψ, are often used to assess the conformational space of disaccharides.† These energy surfaces have shown that the majority of linkage orientations for disaccharides, and particularly polysaccharides, are forbidden, mainly by non–bonded electrostatic and Van Der Waals interactions.[32] Conversely, certain conformations of a

†adiabatic maps are discussed in detail in Section 2.2.1

University of Cape Town

glycosidic linkage can be favoured because they allow for strong interresidue hydrogen bonds between hydroxyls on different rings. Interactions of this type have been observed in carbohydrate crystal structures. However, the extent to which these hydrogen bonds affect the orientation of glycosidic linkages in solution has not been resolved and depends on the degree to which they are disrupted by water.

Disaccharides form the simplest conceptual models of the glycosidic linkages, and are thus an obvious starting point for analysis of polysaccharide conformation. Often the first approach to investigating disaccharides is to determine their crystal structures.[19,21,23,131] However, as the crystal structure conformation may not correspond to the solution conformation, other experimental methods such as optical rotation and NMR have been used to investigate the solution conformation of disaccharides.[31,44,48,132,133] For example, the maltose disaccharide (Glc–α(1,4)–Glc) has been shown to adopt a conformation in solution very different from that in its crystal structure.[31]

Few experimental methods currently provide information on the range of conformational space available to a disaccharide glycosidic linkage and therefore computational studies are being increasingly employed to this end. Several key disaccharides have been studied extensively, either in their own right or because of their value as models for a glycosidic linkage in a polysaccharide. Examples are the maltose disaccharide,[32,35,57,68,69,72,75,82,83,88,134–136] cellobiose (Glc–β(1,4)–Glc)[84,85,137,138] and the nonreducing disaccharides sucrose (Glc–α(1,2)β–Fru)[90,139–142] and α,α–trehalose (Glc–α(1,1)α–Glc).[78,81,143–146] Maltose and cellobiose are typically seen as models of the $\alpha(1\rightarrow4)$– and $\beta(1\rightarrow4)$–linkages respectively, while sucrose is of interest because of its wide usage in the food industry. Trehalose has received a lot of attention because of its known cryoprotective properties. Adiabatic map analyses have shown that the $\alpha(1\rightarrow4)$ linkage in the maltose disaccharide has a relatively small domain of ϕ, ψ conformational space available to it.[32,35,57,72,82,83,134,135] In contrast, adiabatic maps for cellobiose show that the $\beta(1\rightarrow4)$–linkage is intrinsically much more flexible than the α–linkage.[83,84,137,147,148] The $\alpha(1\rightarrow6)$–linked isomaltose (Glc–α(1,6)–Glc) is the most flexible of all the glucan disaccharides, because the extra bond reduces steric clashes and thus increases the range of available conformational space.[57,149] The least flexible is the non–reducing α,α–trehalose.[78,81] However, despite numerous studies, none of the glycosidic linkages have as yet been fully characterised. This is required not only for understanding disaccharide structure and dynamics, but also for polysaccharides, as will now be discussed.

University of Cape Town

1.1.3 Polysaccharide Conformation

A carbohydrate chain typically contains a large number of monosaccharide units linked together by glycosidic bonds. Oligosaccharides are usually defined as comprising up to 20 monosaccharide units, while the terms *polysaccharide* and *glycan* refer to longer chains. *Glucans* are polysaccharides comprised entirely of glucose residues.

For reference purposes, the glucose residues in an oligo– or polysaccharide chain are typically numbered from the reducing glucose residue to the non–reducing residue. In this work, the glycosidic linkage dihedral angles are similarly numbered. Thus, ϕ_4 refers to the fourth ϕ–dihedral angle from the reducing end of the chain.

Oligo– and polysaccharides are similar to peptides in that they possess a secondary structure, such as a helix or random coil, organised from stiff structural elements. This secondary structure is dictated principally by the conformation of the backbone glycosidic linkages between sugar residues, though in pyranose residues the hydroxymethyl group may play a significant role in relatively stabilising or destabilising chain conformations. The glycosidic linkages are also the source of most of a chain's flexibility, as the $^4C_1$ chair form of the pyranose ring is comparatively rigid.[42,141]

The nature of the glycosidic linkages thus governs the general characteristics of the polysaccharide chain conformation and flexibility[43] and ultimately affects the properties and function of a polymeric material. The more flexible the linkages, the less likely it is that an oligosaccharide chain will adopt a unique, regular conformation. Although cooperative intramolecular interactions within a polysaccharide chain can potentially combine to limit the range of a flexible linkage and thus fix the molecule into an ordered shape, this is less likely to occur in solution than in the solid state. In fact, many polysaccharides are highly flexible molecules (particularly in solution): both experiment and theoretical calculations have shown oligosaccharides to exhibit large conformational flexibility.[43,46]

Glucose polymers are a case in point: the D–glucose residue is the basic component of a large group of carbohydrate polymers with widely differing conformations and functions, ranging from ordered conformations such as a ribbon or a helix, to random coils.[43] A frequently cited comparison is that of the amylose and cellulose homoglucans, which differ only in the configuration of their glycosidic linkages ($\alpha(1\rightarrow4)$–linked in amylose and $\beta(1\rightarrow4)$–linked in cellulose) but have vastly different physical properties and aqueous solubilities. Cellulose (β-1,4 D-glucan) is a rigid support molecule and forms structural fibres in plants. Amylose (α–1,4 D–glucan), which was discussed earlier, is the linear component of starch. Structurally cellulose and amylose differ remarkably. Cellulose has regular ribbon–like chain conformations in the solid state[30,150] and cellulose strands

University of Cape Town

are stiff and highly extended in solution.[36,119] In contrast, amylose forms various flexible helical structures in the solid state[17,151,152] and is thought to dissolve as a random coil with a disordered, sixfold helical trajectory on the basis of optical rotation, light scattering, limiting viscosity, sedimentation and NMR measurements.[153–159] The contrasting physical and chemical properties of amylose and cellulose are attributable entirely to their differing glycosidic linkages.

Another example is dextran, the corresponding $\alpha(1\rightarrow6)$–linked glucan. This molecule is very flexible as a result of the extra bond in its backbone linkages and it forms highly disordered, flexible and compact random coils in solution.[160] Furthermore, the regularly–repeating linear "copolymer" pullulan [(→6)–α–D–Glc–(1→4)–α–D–Glc(1→4)–α–D–Glc–(1→] contains both the $\alpha(1\rightarrow4)$- and $\alpha(1\rightarrow6)$–linkages and can be thought of as a mixture between dextran and amylose. It forms a random coil with a weak tendency to follow a pseudo–helical trajectory.[160]

The solution structure of oligo– and polysaccharides is often difficult to characterise experimentally because of their flexibility and disorder. Therefore, small oligosaccharide fragments are often studied as more tractable models of polysaccharides. In many cases, short oligosaccharide strands of polysaccharides have been shown to mimic the dynamics of the polymer. For example, in the case of pullulan, the effective unit for demonstrating segmental motions of the higher polymer contains 12 or fewer glucose residues[161] while strand lengths of 6 units are sufficient in the case of amylose and dextran.[56] However, experimental methods often cannot investigate individual polysaccharide strands in solution and typically provide results gained from an average of all the strands in the polymer. A single chain will adopt a wide variety of conformations over time. According to the ergodic hypothesis, a large group of chains adopts the same distribution of conformations at any one instant.[35] Polysaccharides are therefore often described using the statistical–mechanical theory of polymer chain conformation developed by Flory.[162] Statistical measures are used to describe the average conformation of a polymer strand, such as the root–mean–square distance between the polymer ends or the *radius of gyration* — the root–mean–square distance of the elements of the chain from its center of gravity. In addition, simple models of polymer structure have been developed, such as the *freely jointed chain* model. The freely jointed chain consists of a series of x links of length l joined in a linear sequence with no restrictions on the angles between successive bonds.

Simulations have attempted to reproduce the average statistical parameters measured by experiment. For example, predictions of the scattering function and radius of gyration for a pullulan hexasaccharide based on oligosaccharide linkage properties were found to be in excellent agreement with SAXS data.[163] However, simulation techniques are also

University of Cape Town

limited by polysaccharide flexibility and structural analyses are often reduced to consideration of the constituent glycosidic linkages, with the polysaccharide chain approximated by stiff elements joined by flexible linkages.[32,34]

1.2 Carbohydrates in the Solid State

The less mobile carbohydrates crystallise readily. Examples are the simple carbohydrates, the less flexible cyclic carbohydrates and rigid polysaccharides such as cellulose. However, crystals of the more mobile oligo– and polysaccharides are difficult to obtain. In contrast, most carbohydrates can form amorphous or partially–crystalline glasses in the solid state. The physical properties of these vitreous polysaccharide materials (such as volume, enthalpy, heat capacity, mechanical and dielectric relaxation behaviour) are dependent not only on the molecular weight of the carbohydrate, but also on the type of glycosidic linkage, the molecular configuration and the water content of the glass.[4,10,38,60,164]

Glasses are characterised by a glass transition, which occurs around a particular glass transition temperature (T_g). At T_g there is a change in the material characteristics of the glass, from a brittle glassy form to a less rigid rubbery form, with a concomitant sharp change in heat capacity. The glass transition is a physico–chemical event — a change of state but not of phase — and the temperature at which it occurs is of relevance for the use of carbohydrate polymers as structural materials.[165] The macroscopic changes occurring in carbohydrate glasses at the glass transition have not yet been explained at the molecular level, but they are probably related to a change in molecular mobility. Dielectric relaxation studies of pentopyranoses above their T_g show a primary α–relaxation which is thought to be associated with molecular reorientation.[60,166] (Relaxation processes detected in carbohydrates are categorised according to their strength, with the strongest relaxations referred to as α–processes, the next strongest as β–processes and so on.) Carbohydrate glasses also exhibit sub–T_g molecular mobilities — so–called secondary relaxations — which result in their macroscopic mechanical or dielectric behaviour.[3] These processes are generally attributed to either localised, thermally–activated rotations of side groups or motions of the main chain units, which can be precursors to the main α–relaxation process involved in the glass transition. In particular, the hexopyranoses have a β–relaxation below the calorimetric T_g, which is thought to be associated with rotations of the hydroxymethyl group.[3,59,60,63,166–168]

Water has a potent plasticising effect on carbohydrate glasses, which depresses the glass transition temperature.[1,2,4,37,38] Relatively small amounts of water can reduce the glass transition temperature dramatically. For example, in maltose the T_g drops from 353

University of Cape Town

K to 241 K upon addition of 20 wt% water.[60] This effect is thought to occur because water molecules disrupt the hydrogen bonding network between carbohydrate molecules and thus facilitate molecular motion.[37] Hindered mobility of water molecules in carbohydrate glasses has been reported by modelling studies[169,170] and experiment,[171] with the mobility of water in carbohydrate glasses increasing gradually with water content.[38] Starch and maltose glasses exhibit brittleness on ageing.[10] This retrogradation process in starch is strongly affected by moisture content.[13]

Water also appears to affect the secondary relaxation processes in some vitreous carbohydrates. For example, dielectric relaxation experiments have shown the strength of the β–relaxation process in both glucose and maltose glasses to increase with increasing water content.[60] In these glasses, sub–T_g relaxations are thought to arise from a combination of the motion of the hydroxymethyl groups and reorientation of the water molecules. Cellulose is another example of the effect of water on sub–T_g relaxations. Mechanical spectroscopy experiments have shown cellulose to have two sub–T_g relaxation processes: β and γ. The strength of the β–relaxation is strongly dependent on water content.[59] In fact, dielectric spectroscopy measurements on dry cellulose showed only a γ–relaxation.[3] The molecular origin of these relaxations remains to be established, though the γ–relaxation process has been attributed to a combination of hydroxymethyl and hydroxyl side group motions.

Glasses of simple saccharides have antidesiccant and cryoprotectant properties; they are known to preserve proteins and phospholipid bilayers from denaturation and disruption during extreme dehydration.[172–174] They play important roles in seeds, pollen and in the dormant states of drought–resistant organisms and have practical applications in the preservation and storage of labile foods, therapeutic proteins and pharmaceutical products. The antidesiccant properties of disaccharides tend to follow their glass transition temperatures,[164,175] with the cryoprotective ability of the simple sugars decreasing in the order:[175] trehalose > maltose > fructose > sucrose > glucose > cellobiose > glycerol
Trehalose has a significantly higher T_g than other disaccharide–water systems at all water contents, a fact that has been termed the trehalose anomaly.[175] Its outstanding cryoprotective properties have been the focus of numerous experimental and simulation studies.[78,81,132,176–178] However, the energetic and structural factors responsible for the saccharides' cryoprotective properties remain to be elucidated. One mechanism suggested is the "water replacement hypothesis",[179] whereby saccharides replace lost water around biomacromolecules by forming hydrogen bonds between the saccharide hydroxyl groups and the biomacromolecules' polar residues.

University of Cape Town

1.3 Carbohydrates in Aqueous Solution

Carbohydrates interact strongly with water and the effect of a water solvent on the preferred molecular structure, conformation and dynamics is of prime importance to the biological functions of carbohydrates.

The simple carbohydrates are very soluble in water, though not all to the same degree. For example, ranked in order of increasing solubility:

$$\text{lactose} < \text{glucose} < \text{sucrose} < \text{fructose.}^{5}$$

Although, as a general rule, the solubility of a carbohydrate is proportional to its molecular weight, oligo– and polysaccharides also vary in their interactions with water according to their structures. For example, the solution properties of cyclic and linear oligosaccharides differ dramatically. While the α–, β– and γ–cyclodextrins are of limited solubility in water, with β–cyclodextrin being the least soluble,[180] their linear analogs are almost indefinitely soluble.[152]

Solvents make two basic contributions to the environment experienced by a solute: the first is an electrostatic polarisation effect of the bulk solvent on the electronic structure of the solute molecule, and the second contribution is direct solvent–solute interactions that encompass hydrogen bonding.[181,182] Direct hydrogen bonds between water molecules and the numerous hydroxyl groups possessed by carbohydrates have considerable potential for altering the preferred structures and conformations of these molecules. The "top" and "bottom" surfaces of a glucopyranose ring in the $^4C_1$ conformation present aliphatic C–H groups and are hence hydrophobic. The edges, which are fringed with hydroxyl groups, are hydrophilic. Force field calculations have shown that in vacuum, or a non–aqueous environment, these hydroxyl groups typically form a clockwise or anti–clockwise arrangement of intraresidue hydrogen bonds to minimise the energy.[84] Hydrogen bonds, generally accepted to be short–range and angle–dependent interactions, have a strength usually in the range of 3-6 kcal/mol. This is sufficiently larger than kT at room temperature for them to be quite resistant to thermal disruption. Consequently, specific hydrogen–bonds have been shown to be important in determining the conformation of many biopolymers. However, in aqueous solution, water molecules compete as alternative hydrogen–bonding partners for the hydroxyl groups, with potential conformational consequences if the relative energies of the different conformations are altered. Interactions with water can potentially perturb intramolecular hydrogen bonding within a carbohydrate and alter the hydroxymethyl conformation. Di– and higher saccharides have the added possibility of an altered ϕ,ψ conformation, with water perhaps acting as a bridge between monosaccharide residues in a disaccharide.[89] Polysaccharide helical chain folding and chain association

University of Cape Town

are apparently significantly mediated by hydrogen–bonded interactions of the sugar hydroxyls, though the aqueous environment is not usually thought to be conducive to strong intramolecular hydrogen bond interactions.[35]

The extent to which internal hydrogen bonding persists in solution and the extent to which it continues to influence the conformational properties of carbohydrates has not yet been resolved. Although the mobility of water around carbohydrates has been studied by measuring the relaxation behaviour of the mobile protons using proton nuclear magnetic resonance (^{1}H NMR) experiments,[38] experimental techniques are seldom able to act as a probe of individual hydrogen bonds in solution. However, indirect evidence for some hydrogen bond perturbation can be inferred from altered molecular conformations in aqueous solution. For instance, solvation is known to have an effect on the preferred geometry of the glycosidic linkages in some disaccharides.[31,42–44,47] In oligo– and polysaccharides this could lead to a change in secondary structure.

In the absence of experimental data, simulation methods are a valuable tool for directly investigating specific hydrogen bond interactions. However, results from simulations are not always conclusive as they are dependent on the force fields used to model both the water solution and the carbohydrate. For example, Brady and Schmidt analysed the hydrogen–bonding in a simulation of the disaccharide maltose in solution and concluded that internal hydrogen bonds were not present to any great extent when in solution.[75] However, a subsequent study by Ott and Meyer found a high prevalence of internal hydrogen bonds.[68]

The structural details of specific carbohydrates, particularly their flexibility and the location of the hydroxyl groups, determine exactly how they interact with the surrounding water. Many thermodynamic parameters of aqueous solutions of carbohydrates are sensitive to differences in the stereochemistry of the solute,[183] but the manner in which the hydration of a particular carbohydrate depends on its hydroxyl topology is still a matter of debate.[76] It has been proposed that the anomeric effect governs carbohydrate hydration.[184,185] Franks postulated the stereospecific hydration model, which assumes that an equatorial hydroxyl group is more fully hydrated than an axial hydroxyl group; hence the hydration of a carbohydrate would depend on the ratio of equatorial to axial hydroxyl groups.[186,187] Although this theory has found considerable support, doubts have arisen concerning its general applicability. Recently, a modified stereospecific hydration model has been proposed,[76] which emphasises that the hydration of a carbohydrate depends on the relative distance of the nearest neighbour oxygens of a carbohydrate compared to the oxygen–oxygen distances of water. In this view, the solvation of carbohydrates in water depends essentially on how well the carbohydrate molecule fits into the three–dimensional

University of Cape Town

hydrogen–bonded network of water. This means that particular carbohydrates that have pairs of oxygen atoms the same distance apart as the oxygen atoms of water are more completely solvated.[76]

In turn, it has been proposed that carbohydrates affect water by imposing an "ice–like" structure on the solution molecules around them. This allows for a distinction between water of hydration and unbound "bulk" water.[93] However, analysis of the first solvation shell around simulated maltose and maltohexaose in water has shown no evidence for this.[72]

1.3.1 Carbohydrate Dynamics

As a result of their flexibility, there is often no single structure of a carbohydrate in solution and the solution conformation of a carbohydrate can often only be described in terms of population distributions in conformational space.[44,47–49] Both the identity of the solvent and temperature can affect the relative abundance of the populations.

A variety of relaxation processes can occur in aqueous solutions of monosaccharides. These include pseudorotation interconversions between different linear and ring forms (which require bond breaking), interconversion between the various chair and boat forms of the hexose ring, rotations of the primary and secondary hydroxyls and, where applicable, rotations about the glycosidic bonds. Acoustic absorbance spectra for various monosaccharides in water show five relaxation regimes, with relaxation times on the order of 1 μs, 100 ns, 10 ns, 1 ns and 100 ps each.[61,62,188] These have been assigned, respectively, to $^4C_1 \rightleftharpoons {}^1C_4$ chair–chair conformational inversion dynamics, two modes of pseudorotation of ring structures, *exo*-cyclic group isomerisation and a possible molecular association mechanism. Interconversion between different anomers and between the furanose and pyranose ring forms require bonds to be broken and are therefore comparatively slow processes, on the order of seconds. Chair–chair conformational relaxation is the slowest process in this range — it appears to occur in a time scale of the order of 200 nanoseconds.[61] Hydroxymethyl rotational motion was found to occur in the nanosecond time scale,[188] which is in agreement with previous studies which estimated exchange between the different conformations of the hydroxymethyl groups to occur on a time scale of between 10^{-10} and 10^{-3} seconds.[123,141]

Oligosaccharide and polysaccharide molecules can be very flexible and the dynamics of their interconversion between low–energy structures and conformations is central to their physical and chemical properties in aqueous solution. Because of the complexity of the motion, little can be determined about the amplitude and time scale of carbohydrate mo-

University of Cape Town

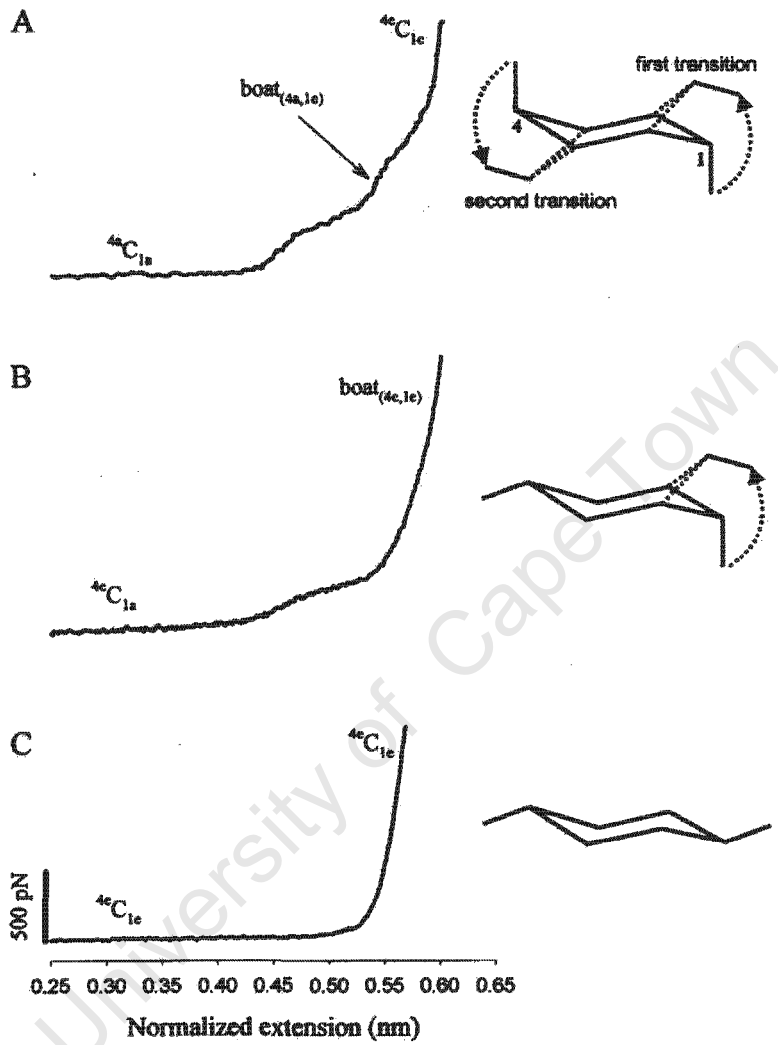
tion experimentally. NMR conformational analysis, based on NOEs and scalar couplings, of a single oligosaccharide will give a minimum range of conformations necessary to fit the experimental data, but will not give the time scales of any conformational changes.[64] A number of NMR relaxation studies have probed the solution relaxation of amylose,[56,189] dextran[56] and other oligosaccharide strands.[190,191]

Motions within oligo– and polysaccharide molecules are likely primarily to involve rotations about the glycosidic linkages. However, although it is expected that the nature and time dependence of conformational changes of oligo– and polysaccharides in solution are largely dependent on the flexibility and physics of the glycosidic linkages,[192] the influence of the glycosidic linkage type on the dynamics of the conformational changes in polysaccharides remains to be be clarified.[35] In addition, the nature of these rotations is poorly understood. At low temperatures, rotations about the glycosidic linkages are often considered to be localised, with the glycosidic linkages rotating independently of their neighbours.[3,192,193] However, if hydrogen–bonding spans the glycosidic linkage of a polysaccharide chain, another possibility is the existence of coupled segmental motions of the polymer backbone. These involve cooperative motions of two, three or more sugar residues bonded in sequence and are often referred to as crankshaft motions.[192] The concept of cooperativity is often associated with a non–zero entropy in the total energy barrier involved in the conformational change[194,195]— if the entropy is large, the motions are cooperative and involve segments of the main chain.

Molecular dynamics simulations can provide a detailed description of the dynamic behaviour as well as the average conformer(s) of a glycan. A combined approach is to use interproton distances determined by simulation and experimental NOE intensities to calculate the dynamic behaviour of specific linkages in an oligosaccharide.[64,196] Brant and co–workers have done a large number of theoretical and experimental studies on the physics, mobility and dynamics of oligo– and polysaccharide strands in solution, investigations which span over three decades.[34,35,56,161,163,192,193,197,198] Their recently–developed theoretical approach to polymer dynamics — optimised Rouse Zim local dynamics (ORZLD) theory — showed that correlation times of glucans and galactans (obtainable from NMR, fluorescence, radiation scattering and rheological relaxation experiments) are very sensitive to the details of molecular structure.[192]

A novel approach to the study of the elastic properties of glycan strands is atomic force microscopy (AFM).[50] Various single–molecule atomic force microscopy (AFM) experiments have probed the elastic properties of single polysaccharide strands.[51–53,121,122] The procedure uses a single molecule lifted from solution and adsorbed between a substrate and cantilever tip and measures the force required to extend the molecule by a certain

University of Cape Town



University of Cape Town

Figure 1.5: AFM force–extension curves for galactouronan (A), amylose (B) and cellulose (C) reproduced from Marszalek et al.[122] The postulated dominant conformations are illustrated along the corresponding sections of the force–extension curve. The extension x was normalised using $x_{norm} = 0.45x/l_1^c$, where 0.45 nm corresponds to the length of the residue vector $O_1O_4{}^{chair}$ in the relaxed state of the polymer and l_1^c is the estimated contour length of the molecule determined from a freely–jointed chain model fitted to the unnormalised data before the first transition.

amount. This produces a characteristic force–extension graph for each molecule, examples of which are shown in Figure 1.5. The polysaccharides amylose (α1,4 D-glucopyranose), cellulose (β1,4 D-glucopyranose), dextran (α1,6 D-glucopyranose) and pullulan ([α1,4–α1,4–α1,6]$_n$ D-glucopyranose) each generate distinctive, fingerprint force extension curves, though they differ only in the locations of their glycosidic linkages.[51–53] The α–linked molecules undergo elastic deformations in response to stress, evidenced by characteristic plateaus in their force–extension curves,[51,121] which are not present in the profile for the inelastic, β–linked cellulose.[52] The yielding behaviour of the α–linked glucans has been attributed to the sudden onset of concerted transitions of the glucose rings that lengthen the chain, specifically a force–induced elongation of the glucose rings that results in a final transition from the preferred $^4C_1$ chair to a twist–boat conformation.[53] Thus, the pyranose rings have been proposed as the structural unit controlling glucan elasticity.

1.4 Carbohydrate Simulations and Free Energy Calculations

As the experimental methods currently available for investigating carbohydrates in many cases do not provide sufficient atomistic detail, computer simulations are being increasingly used to provide insight into molecular structure and dynamic motions. Ideally, simulations serve as a bridge between physical theory and experiment, helping to resolve experimental ambiguities, providing rationales for experimental observations and ultimately allowing for the prediction of molecular behaviour.

Quantum mechanical or *ab initio* calculations are the most accurate of the simulation methods. Some *ab initio* calculations have been performed on model fragments of carbohydrates,[104,112,113,199] and a few simple sugars, such as glucose,[102–105,200] fructose[201] and some disaccharides.[67,136,202,203] However, because of their large size and flexibility, carbohydrates have been principally investigated by force field methods. These encompass Molecular Mechanics minimisation, Monte Carlo calculations and Molecular or Stochastic Dynamics simulations. Force field methods require an accurately parameterised force field for carbohydrates, and a number of these have been developed.[204,205]

Molecular Mechanics analysis usually expresses the conformational space available to a molecule in terms of a Ramachandran map of the molecular energy as a function of the principal degree(s) of molecular freedom. Methods for calculating Ramachandran maps are discussed in detail in Section 2.2.1. Ramachandran maps have been used extensively for predicting the conformational space of disaccharides, where two–dimensional maps

University of Cape Town

of the ϕ and ψ dihedral angles are typically calculated.[72,82,83] In addition, ambiguous spectroscopic data is often supplemented by calculations of these adiabatic potential energy surfaces.[42,46,57,65,67,70,71,73] Although generally used only for analysis of the preferred conformations of disaccharides, Ramachandran maps can also be used to predict the properties of an oligosaccharide. For example, Mazeau and Pérez used calculated adiabatic maps of the 18 disaccharides which form the building blocks of the rhamnogalacturonan II oligosaccharide to predict the low energy conformations of this oligosaccharide.[206] However, adiabatic maps are limited in their applicability because a slight increase in the degrees of freedom of a molecule causes a large increase in the computational complexity. Moreover, entropic and solvent contributions to the molecular energy are not incorporated into the adiabatic energy surface. While the potential energies calculated by Molecular Mechanics minimisation correspond to estimates of the enthalpy, experimental determinations of energy differences principally yield Gibbs free energies. Both enthalpy and entropy play important roles in determining the relative probabilities of conformations and the mobility of carbohydrates — enthalpy differences alone should not be used for comparison of conformer preferences.[205,207] Therefore, although adiabatic maps can give a satisfactory description of the conformational space for a glycosidic linkage in vacuum, where the entropic component of the energy can be expected to be small, they are likely to be an inadequate approximation of the glycosidic linkage conformational space in solution.

Molecular Dynamics (MD) simulations, by incorporating molecular motion into a simulation, can provide an indication of the enthalpic contributions to the molecular energy. For simulations of a carbohydrate in solution, explicit solvent molecules are usually included, though some simulations have been performed with generalised solvent models or force fields that aim to incorporate the average effects of solvent.[208,209] Although computationally more expensive, studies have shown that the explicit inclusion of water molecules is necessary for carbohydrate simulations to adequately model the conformational space available.[73,210] The first Molecular Dynamics simulation of a carbohydrate (glucose) in explicit solution was performed by Brady in 1989.[74] Since then, a variety of simulations of carbohydrates in vacuum,[82,115] water,[67,68,72,74–81,177,211] and the solid state[170] have been published. Most MD simulations of carbohydrates have concentrated on the mono– and disaccharides, such as α–D–glucose,[74] β–D–mannose, and (α,β)–D–xylose[111] as well as β–D–galactose, (α,β)–D–talose,[76] maltose,[68,69,75,212] cellobiose,[138] trehalose,[78,81,177,178] sucrose and carrabiose.[213] However, some MD studies of oligosaccharide strands have been performed,[72,80,159,214] principally to investigate the effects of solvation on conformation and to potentially extrapolate these results to the corresponding polysaccharide.

University of Cape Town

These include simulations of the oligosaccharides maltohexaose,[72] maltoheptaose[159] and a hyaluronan decamer.[80] A similar, but computationally cheaper, approach is to perform computational analysis of an oligomer using gas–phase geometries obtained from disaccharide analysis in combination with a solvation model.[160,163]

MD simulations may be used to investigate dynamic relaxation processes in oligosaccharides and thus clarify the molecular origin of the relaxation processes in polymers. However, most studies focus on determining the global energy minimum and water interactions; few are concerned with investigating conformational changes. In addition, the flexibility of carbohydrates also causes difficulties with MD simulations. In the case of large oligosaccharides, long molecular dynamics simulations (on the order of hundreds of nanoseconds) are required to ensure that phase space is adequately sampled and all possible low–energy conformations explored. With the computing power currently available, long simulations of large molecules are frequently too time–consuming to be feasible, particularly if solvent is explicitly included. There are a few alternatives for circumventing these computational limitations. One approach is to use simulations of shorter oligosaccharide strands that can be extrapolated to explain polysaccharide conformation and dynamics. For example, simulations of oligosaccharide fragments have been used to help explain the force–extension curves obtained from AFM stretching experiments on individual polysaccharide strands and thus give a partial explanation of macromolecular properties like stiffness at the molecular level.[120] Another approach is to employ computationally–cheaper reduced models of polysaccharides. In these studies, a polysaccharide is represented by rigid elements joined by flexible linkages, relying on the assumption that polysaccharide dynamics for a single chain in solution are determined principally by the glycosidic linkages between sugar residues. This approximation allows for simulation of much longer polymer strands. Reduced models have been used to investigate (1→3) and (1→4)–linked α and β–D-glucans and (1→4)–linked α– and β–D–galactans.[192]

However, although standard MD simulations are useful, ideally free energy calculations are required in order to provide a complete characterisation of the conformational space available to a carbohydrate. Free energy calculations for the important degrees of freedom in a carbohydrate allow for identification of the number of low-energy conformers, the flexibility of each conformer, the time spent in each conformation and the rate of conversion between conformers. A free energy surface along a degree (or degrees) of freedom is known as a *potential of mean force* or PMF. In solution, a PMF incorporates not only contributions from entropy, but also the averaged solvent–solute interactions. In principle PMFs can be obtained from standard simulations. However, in practice high energy barriers often prevent representative sampling of the configurational space. For

University of Cape Town

instance, none of the standard force field simulation techniques (MD, Monte Carlo simulations) is able to thoroughly explore the areas of conformational space available to a disaccharide, let alone higher carbohydrates.

Free energy calculations have long been identified as one of the major challenges to computational chemists.[215,216] However, despite the fact that methods for computing free energy changes have existed for decades, potential of mean force calculations for carbohydrates remain computationally expensive and hence infrequently attempted. Of the free energy calculations performed on carbohydrates, most have concentrated on reproducing the experimental anomeric free energy in glucose.[108,110] Few free energy studies have investigated conformational changes in carbohydrates. A ϕ, ψ PMF for sucrose in solution was calculated as early as 1995.[90] However, this procedure used very short simulation times (a total of 3.9 ns) and it is therefore unlikely that the calculations converged. Both maltose and dixylose (4–*O*–α–D–xylopyranosyl–α–D–xylopyranose) have been investigated with partial PMF calculations restricted to a central ϕ,ψ region because of sampling difficulties.[89,181] In addition, Palma et al. calculated a one–dimensional PMF for the binding of glucose to benzene in solution.[217] With the rapid increase in computer speeds, longer simulations are possible and thus free energy calculations are becoming more feasible.

University of Cape Town

1.5 Objectives

The aim of this thesis is to develop free energy calculation methods that will provide a full characterisation of both the preferred conformations and the dynamic behaviour of particular classes of carbohydrates. A description of the influence of water on carbohydrate conformations and dynamics is considered to be particularly important.

As the pyranose chair conformation is known to be fairly rigid in solution, the larger dynamic motions in a glucan polysaccharide in solution consist primarily of rotations about the dihedral angles of the *exo*-cyclic groups of the pyranose ring (hydroxymethyl and secondary hydroxyl groups) and the glycosidic linkage torsion angles. At higher temperatures and under the application of forces, chair-to–boat transitions of the glucose rings may add an additional dimension. Characterising the conformations and motions of carbohydrates therefore requires calculation of the potential of mean force along these principal degrees of freedom.

In order to build a detailed picture of the dependency of a polysaccharide's dynamics on its constituent monosaccharides and glycosidic linkages and the effects of water upon them, the aim is to employ a systematic analytical approach. The procedure begins with

calculation of the free energy along the principal degree of freedom in a monosaccharide — the hydroxymethyl group. The investigation is then extended to characterisation of the free energy about the glycosidic linkage dihedral angles in a model disaccharide. Potential of mean force calculations for a glycosidic linkage provide a complete description of the preferred conformation and the structural fluctuations that a disaccharide is likely to undergo. If the glycosidic linkages are assumed to operate independently of each other, a 2D PMF for the glycosidic linkage in a disaccharide will also provide all the information required to determine the average chain conformation of a homopolysaccharide, as well as the dynamic behaviour of the chain.[193] In addition, if performed in solution, these calculations should clarify the effect of solvent on the properties of the linkage. Finally, the conformational free energy of the corresponding oligosaccharide is explored using end–to–end PMF calculations both under standard conditions and with the application of a stretching force. Stretching simulations investigate polysaccharide dynamics more thoroughly than simple MD simulations and hence give insight into the conformational changes which determine molecular flexibility as well as the forces required to bring about conformational changes. The objective of this procedure is therefore to characterise a particular linkage and then explore its cumulative effect in a polysaccharide. Thus, this work should help to establish to what extent the behaviour of the polysaccharides can be extrapolated simply from its constituent linkages.

In the interests of simplicity, this thesis focuses specifically on characterisation of glucans connected by (1→4) linkages, which are investigated in both the α and β forms. This linkage is chosen because of its biological ubiquity and economic importance. The intention is that, once the free energy procedure has been thoroughly validated on these two linkages, a wider variety of carbohydrates can be explored.

The reliability of carbohydrate simulations is dependent on the quality of the force field employed. Simulations have shown that the chosen force field,[218] while otherwise satisfactory, does not produce hydroxymethyl rotations at a rate consistent with experimental evidence.[219] Therefore, a further objective of this work is to correct the hydroxymethyl equilibrium conformational populations and rotational frequency before proceeding with the more computationally costly free energy calculations. The hydroxymethyl group comprises the principal degrees of freedom in pyranose monosaccharides and may affect the interaction of monomer units with each other, interactions with solvent and the relative energies of the glycosidic linkage conformations.

University of Cape Town

1.6 Thesis Overview

Chapter 2 contains a brief introduction to Molecular Mechanics and Molecular Dynamics simulations of carbohydrates, with the emphasis on force fields and methods used in this thesis. This chapter may be omitted by readers familiar with these techniques.

In Chapter 3 the methods developed and employed in this thesis for potential of mean force calculations are outlined and discussed. The statistical mechanical theory of free energy and the iterative Adaptive Umbrella Sampling procedure for free energy calculations are described first. The Adaptive Umbrella Sampling method is employed as it is suited to PMF calculations while also having the benefits of gradually converging to a final solution. These sections are followed by an explanation of the Weighted Histogram Analysis Method. This method is used to combine overlapping conformational distributions generated by a series of MD simulations and thus to obtain a best estimate of the potential of mean force along a molecular degree of freedom. The techniques used to extrapolate the PMF into unexplored regions and to subsequently smooth the surface are detailed. This work represents some of the first free energy calculations to be performed on carbohydrates and the established Umbrella Sampling protocol required several modifications in order to efficiently explore the energy surface of these flexible molecules. These are discussed in the penultimate section of Chapter 3. Finally, a detailed description of the methods used to perform one– and two–dimensional PMF calculations on carbohydrates is given, including the application of Adaptive Umbrella Sampling to simulations of molecular stretching.

Chapter 4 reports the reparameterisation of the hydroxymethyl dihedral angles in a carbohydrate force field. The force field was altered by defining a new atom type for the hydroxymethyl group and then adjusting various of the dihedral angle force constants describing rotation about the C_5–C_6 bond. The new parameters are validated by determining the potential of mean force for rotation about the hydroxymethyl group in the β-D-glucose and β-D-galactose monosaccharides both in vacuum and aqueous solution. These one–dimensional PMF profiles compare favourably to those calculated for two other parameter sets and demonstrate that simulations using this force field will reproduce the experimental conformer distributions and rotational frequencies for the hydroxymethyl group. Further validation of the force field is provided by vibrational frequency calculations on α–D-glucose.

The application of the methods discussed in Chapter 3 to carbohydrates with specific glycosidic linkages is presented in subsequent chapters. Chapters 5 and 6 investigate the $\alpha(1\rightarrow4)$–linkage. Chapter 5 is concerned with characterising the effects of the *exo*–cyclic

University of Cape Town

hydroxymethyl group, solvation and additional monomers on the preferred conformations of the $\alpha(1\rightarrow4)$–linkage. These are explored by comparison of the calculated two–dimensional ϕ, ψ free energy surfaces for the $\alpha(1\rightarrow4)$–glycosidic linkage in three model compounds: maltose, dixylose (the pentose analogue of maltose) and the 4–unit oligomer, maltotetraose. Chapter 6 then extends the investigation of the $\alpha(1\rightarrow4)$–linkage to the chain conformations and dynamics of $\alpha(1\rightarrow4)$–linked oligo– and polysaccharides. The implications of the maltose vacuum and solution PMFs for amylose chain conformations in these environments are more thoroughly explored. In addition, the dynamics of this polymer are probed both by end–to–end free–energy surface calculations for $\alpha(1\rightarrow4)$–linked oligomer strands in vacuum and an extended simulation of an AFM stretching experiment on a 18–unit amylose oligomer in vacuum. This provides a new interpretation of the molecular origins of the force–induced elastic deformations in amylose evidenced by the experimental force–extension graph.

Chapter 7 deals with the $\beta(1\rightarrow4)$–linkage. Vacuum and solution ϕ,ψ PMFs of cellobiose in vacuum and solution are probed to establish the effect of water on this linkage. In addition, study of the 4–unit cellotetraose provides insight into how the $\beta(1\rightarrow4)$–linkage is affected by additional units. A stretching simulation for a 12–unit oligosaccharide fragment of cellulose is explained with reference to the cellobiose PMF.

The penultimate chapter briefly describes the application of PMF calculations to the $(1\rightarrow1)$–linkage in the cryoprotectant disaccharide α,α-trehalose as well as the $\alpha(1\rightarrow6)$–linkage in isomaltose and the polymer dextran. Vacuum and solution ϕ, ψ PMFs of trehalose are compared in order to establish the effect of solution on this molecule. A novel two–dimensional free energy surface in vacuum for the 3–bond $\alpha(1\rightarrow6)$ linkage is presented. Stretching simulations of dextran oligomers are explained in terms of this energy surface.

Finally, Chapter 9 contains conclusions and suggests future avenues of investigation.

University of Cape Town

Chapter 2

Molecular Mechanics and Dynamics Simulation Methods

University of Cape Town

Carbohydrates are particularly challenging molecules to simulate because of their flexibility, their often highly polar substituents and the importance of electronic effects in determining their conformational and configurational preferences.[205] Electronic effects are best modelled by Quantum Mechanics. However, although the electronic influences on the conformations of glucose have been extensively investigated by *ab initio* methods,[102,103,106,200,220,221] either the extension of these methods to oligosaccharides or the addition of explicit solvent molecules increases the size of the calculations to such an extent that Quantum Mechanical methods become impractical.

Force field methods, which encompass Molecular Mechanics (MM) and Molecular Dynamics (MD) as well as Monte Carlo simulations, simplify the calculations by ignoring electronic motions. The energy of the system is calculated as a function of the nuclear positions only, using a classical "ball–and–spring" model for the atomic interactions. Despite their comparative simplicity, suitable force field models are often more accurate for modelling larger carbohydrates than unreliable low levels of quantum theory and are thus an efficient and appropriate choice for complex systems. The most widely used modelling methods for carbohydrates are Molecular Mechanics and Molecular Dynamics simulations. An advantage of the latter is that it provides a detailed history of molecular interactions and conformational changes. This, as well as complementing experimental measurements, is of great assistance for investigating the thermodynamic properties of carbohydrates and an invaluable aid in elucidating the complex roles of sugar flexibility in biology.

The increasing popularity of Molecular Dynamics simulations has brought a concomitant increase in the number of available software simulation programs and packages for biomolecular modelling. Among the most frequently used are the academically developed

programs, such as the CHARMM[222] program used in this thesis, AMBER,[223] GROMOS[224] and DL_POLY.[225]

This chapter provides a brief overview of the Molecular Mechanics and Dynamics simulation methods and their application to the study of carbohydrates. The focus is on techniques used in this thesis. For a broader treatment the reader is referred to the general texts on molecular modelling, such as those by Allen and Tildesley[226] and Leach,[215] and relevant review articles.[227–229]

2.1 Force Fields

Force field methods require an appropriately parameterised force field, which in turn requires specification of both the functional form of the potential energy function and the numerical values of the functional parameters.

2.1.1 The Potential Energy Function

Models for carbohydrate simulations range from detailed atomic models to mesoscale approximations for modelling larger polysaccharides. However, force fields typically model atoms as balls having a mass and a radius, connected by bonds to other atoms. The energy of a system of atoms is then expressed as a sum of the energies of the bond and angle deformations away from specified reference values, bond rotations and non–bonded interactions. Typically, the force field takes the simple form of an effective pair potential, where all interactions are described as those between pairs of atoms. The CHARMM[222,230] force field used in this thesis and the force fields implemented in AMBER and GROMOS all employ a simple quadratic functional form for the potential energy (V) as a function of the relative positions of the N particles in the system:

$$\begin{aligned} V = & \sum_{\text{bonds}} k_b(b - b_0)^2 \\ & + \sum_{\text{UB}} k_{UB}(S - S_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{\text{dihedrals}} k_\phi(1 + \cos(n\phi - \delta)) \\ & + \sum_{\substack{\text{improper} \\ \text{torsions}}} k_\omega(\omega - \omega_0)^2 \\ & + \sum_{\substack{\text{non-bonded} \\ i<j}} [\epsilon_{ij}(\frac{Rmin_{ij}}{r_{ij}}^{12} - \frac{Rmin_{ij}}{r_{ij}}^{6}) + \frac{q_i q_j}{e r_{ij}}] \end{aligned} \tag{2.1}$$

University of Cape Town

The first term in this equation is a harmonic potential to model the energies resulting from bond deformations, where b is the bond length, b_0 the equilibrium length for bond b and k_b is the bond force constant. The second and third terms are harmonic potentials to model the energies resulting from angle deformations and Urey-Bradley deformations of the distance between atoms separated by two covalent bonds respectively. Here S is the distance between atoms separated by two covalent bonds (1,3 distance), k_{UB} the corresponding force constant, S_0 the equilibrium distance, θ is the bond angle, θ_0 is the equilibrium bond angle and k_θ is the angle force constant.

The fourth term is a torsional term that describes the energy change upon rotation about bond dihedral angles. The dihedral energy profile is obtained from the summation of cosine terms of various multiplicities ($n = (1, 2, 3, 4, 5, 6)$). Here ϕ is the torsional angle, δ is the phase angle, n is the multiplicity and k_ϕ is the dihedral force constant associated with a particular cosine term. General torsion terms only consider the types of the two central atoms: the energy profile for rotation about a bond depends solely upon the types of the atoms that comprise the central bond, and not upon the terminal atoms, while the actual torsional profile is determined by the sum of all terms. Alternatively, specific torsion terms specify the types of all four atoms and have a contribution from each bonded quartet of atoms *A–B–C–D* in the system. Where specific torsional terms are used, the energy profile for rotation about a bond is determined by the sum of all the dihedral angles incorporating the bond, each of which may contain a number of cosine terms. Therefore, although specific torsional terms are potentially more accurate, their use drastically increases the number of torsional terms and hence the number of torsional parameters required to be specified in parameterisation.

The fifth term describes improper torsion angles (ω), which are used to maintain four atoms in a plane. For each of these terms, the equilibrium value for the improper dihedral, ω_0, and the force constant associated with deviation from planarity, k_ω, must be specified.

The sixth term in this expression is a non–bonded interaction term, incorporating a Lennard–Jones potential for Van Der Waals interactions ($[\epsilon_{ij}(\frac{Rmin_{ij}}{r_{ij}}^{12} - \frac{Rmin_{ij}}{r_{ij}}^{6})]$) and a Coulomb potential for electrostatic interactions ($\frac{q_i q_j}{\epsilon r_{ij}}$). This is calculated between pairs of atoms that are either in different molecules, or separated by at least three bonds within a molecule (r_{ij} is the distance between atoms i and j, ϵ_{ij} is the Lennard–Jones well depth, $Rmin_{ij}$ is the minimum interaction radius for the atoms i and j and q_i, q_j are the charges on the atoms i and j).

Additional terms may be added to this basic functional form for a force field. Hydrogen bonding is sometimes treated explicitly with a 10–12 Lennard–Jones term between hydrogen–bonding atoms. However, many force fields, including CHARMM, rely on hy-

University of Cape Town

drogen bonding effects being reproduced by suitably parameterised non–bonded terms. Class II force fields add harmonic stretching and bending terms as well as cross terms (bend–stretch, torsion–stretch etc.) with the aim of increasing the accuracy and sophistication of the force field.

2.1.2 Parameterisation

Parameterisation of a force field involves specification of the constant parameters in the potential energy function (e.g. the force constants and equilibrium distances and angles) for each atom type pair (or quartet in the case of specific dihedral angles). Atom types are assigned according to atomic species and molecular environment. Parameters are then usually either estimated from known experimental values or optimised systematically to be compatible with data obtain from experiment and/or *ab initio* calculations on relevant model compounds. Force constants for bond–stretching and bending, the "hard" degrees of freedom, are usually obtainable from experimental vibrational data. Unfortunately, experimental data for the non–bonded interactions (torsion, Van Der Waals and electrostatic terms) are difficult to obtain and theoretical quantum mechanical calculations on relevant model compounds are the usual means of determining reference torsional profiles. The torsional parameters, in conjunction with the Van Der Waals potentials and partial charges, are then fitted to reproduce the calculated torsional barriers together with the relative energies of the different conformations. Thus, force fields are usually a combination of experimentally determined valence properties and quantum mechanically computed torsion potentials and non–bonded terms.

University of Cape Town

2.1.3 Carbohydrate Force Fields

There are numerous force fields and parameter sets available for modelling carbohydrates. Some focus specifically on carbohydrates, others aim to treat biomolecules in general and yet others claim to be applicable to any molecular class.[204] An example of the last–mentioned are the general purpose force fields MM2,[231] and its successor class II force field, MM3.[232]

Carbohydrate–specific force fields were originally developed principally as structural models to assist in interpreting ambiguous NMR data. An example is the Hard Sphere Exo–Anomeric approach (HSEA)[233,234] which uses a simple potential energy function with rigid sugar residues incorporating only Van Der Waals interactions and an *exo*–anomeric torsional potential for the ϕ angle. HSEA makes no provision for energy minimisation or monomer flexibility and hence has limited applicability.

Most of the more sophisticated biomolecular force fields currently in used were originally developed for use with proteins, but have been extended with a variety of parameter sets for modelling other classes of molecule. Specific carbohydrate parameter sets have been developed for each of the three principal biomolecular force fields: CHARMM,[181,218,235] AMBER[86,236–240] and GROMOS.[212,241,242] Attempts have been made to compare the carbohydrate force fields currently available to provide a guide as to which would be most suitable for particular calculations.[159,204,205] However, this is a difficult task, as the somewhat different functional forms of the various force fields available preclude direct comparison of parameters and often the strength of one force field is the weakness of another and vice versa. As the CHARMM simulation program is used exclusively in this thesis, only the carbohydrate force fields available for use with CHARMM will be discussed in detail.

Carbohydrate parameter sets for CHARMM

Brady and coworkers developed the first carbohydrate parameter set for use with the CHARMM force field,[181] referred to here as the HGFB force field, after the authors' initials. This force field was parameterised using the CHARMM parameter optimisation feature to fit the experimental vibrational and structural properties of a representative monosaccharide, α–D–Glc–*p*. Partial charges were based on standard CHARMM values. General ring torsional force constants were selected to optimise the agreement between computed and experimentally observed torsion values. Glycosidic angles were treated in the same way as the ring C–O–C angle and there was no specific term to treat the *exo*–anomeric effect.

Subsequently, Reiling *et al.* developed a CHARMM parameter set for carbohydrates.[235] This parameter set introduces certain specific torsional terms, parameters for which were derived from *ab initio* calculations on relevant small molecules representing fragments of carbohydrates. It also has no specific *exo*–anomeric term.

A further CHARMM force field was recently developed by Palma et al,[218] and is here referred to as the PHLB force field. This parameter set replaces the general dihedral angle terms in the HGFB force field with specific dihedral angle terms, using up to to three terms for each dihedral (1–,2– and 3–fold cosine terms). This is consistent with the parameterisation philosophy used in the recent CHARMM[222] protein force field,[243] in which important rotational profiles were represented by more than one dihedral angle potential term. The dihedral angles were parameterised to reproduce experimental vibrational frequency data and to fit *ab initio* energy surfaces for small molecules such as ethylene

University of Cape Town

glycol.[199] A considerable portion of this thesis (Chapter 4) is devoted to adjusting the PHLB force field in order to improve the dynamic behaviour of the hydroxymethyl group.

The simplified CHEAT95 parameter set for carbohydrates[208,209] is a united atom parameter set that uses extended atoms for the representation of the hydroxyl groups. This force field aims to approximate the effect of solvent, employing the assumption that intra– and intermolecular hydrogen bonding play only a minor role in determining oligosaccharide conformations.

2.1.4 Water Models

The effects of solution on a solute molecule are usually incorporated into a simulation in one of two ways: either explicitly as individual molecules that interact with the solute or implicitly either in the parameterisation of the terms of the force field or as an applied field. If an applied field is employed, the electrostatic screening effect of the solvent molecules is modelled as a dielectric continuum characterised by a dielectric constant. Dielectric constants greater than unity are often employed in vacuum simulations of carbohydrates in order to approximate the effect of solvent by reducing the effect of intermolecular hydrogen bonding. In addition, effects of solution on a solute model may be realistically modelled without explicit solvent molecules using Langevin dynamics (Section 2.3.2).

Implicit models are attractive as they are computationally cheap. However, because of the structural role that specific hydrogen bonds often play in water–polysaccharide solutions, implicit models can be inadequate for modelling aqueous states of carbohydrates. The use of explicit water in carbohydrate solution simulations has been shown to be important for reproducing NMR results[73] and to adequately model the available conformational space in simulations of oligosaccharides.[210]

The use of explicit waters in a simulation requires a suitable model. The first MD simulations of pure water were performed by Rahman and Stillinger in the early 1970's,[244–246] using the simple four point charge potential functions BNS and ST2. Since then models have developed considerably and those currently in use for MD range from simple rigid site–interaction models, through flexible models which allow for conformational changes in the water molecule, to the most complex models that explicitly include polarisation and many–body effects. Biomolecular simulations in solution have typically been restricted to the computationally cheaper site interaction models because of the large numbers of water molecules required to adequately solvate carbohydrates. Hence, the various versions of the TIP (transferable intermolecular potential) force field developed by Jorgensen and coworkers[247–250] (TIPS, TIP3P, TIP4P, TIP5P) and the SPC (simple point charge) force

University of Cape Town

	TIP3P	TIP4P	TIP5P	SPC	SPC/E
$r(OH)$ (Å)	0.9572	0.9572	0.9572	1.0	1.0
∠HOH	104.52	104.52	104.52	109.47	109.47
$A \times 10^3$ (kcal.Å^{12}/mol)	582.0	600.0	544.5	629.4	629.4
$C \times 10^3$ (kcal.Å^6/mol)	595.0	610.0	590.3	625.5	625.5
qO	-0.834	0.0	0.0	-0.82	-0.8476
qH	0.417	0.52	0.241	0.41	0.4238
qM	0.0	-1.04	-0.241	0.0	0.0
$r(OM)$ (Å)		0.15	0.70		
θ_{LOL}			109.47		

Table 2.1: Molecular geometry and parameters for the various TIP and SPC water models.

field[251] (SPC and SPC/E) are currently the most commonly–used force fields. In these models, each water molecule is maintained in a rigid geometry and represented by either three, four or five interaction sites (Table 2.1). Interactions between molecules are described using Coulombic and Lennard–Jones expressions in order to mimic hydrogen bonds in small separations and an electrostatic multipole interaction in large separations. Polarisation effects are only incorporated in an average sense.

The three–site models (TIP3P,[248] SPC[252] and SPC/E[251]) have positive charges situated on the two hydrogen atoms and a negative charge on the oxygen for computing the Coulomb energy. The only additional energy terms are Lennard–Jones functions for the Van Der Waals interactions. The SPC models and the original TIP3P model only consider Van Der Waals interactions between oxygen atoms. However, the version of the TIP3P model implemented in the CHARMM package has a small Van Der Waals radius associated with the hydrogen atoms as well. In the case of four–site models, such as TIP4P,[248] the negative charge is situated off the oxygen toward the hydrogens at a point (M) on the bisector of the HOH angle. Five site models include the original ST2[245] model and the recent TIP5P model.[250] These balance negative charges situated on the two lone–pair sites of the oxygen atom at an angle θ_{LOL} to each other, by opposite charges on the hydrogen atoms.

Water model parameters are usually optimized to reproduce the thermodynamic and structural properties of bulk water (including the energy, density, radial distribution functions and diffusion coefficients) at standard temperature and pressure (Table 2.2). However some water models, such as SPC/E, TIP4P and TIP5P, have nevertheless yielded results that show good correspondence with experimental properties over a large range

University of Cape Town

	Diffusion Constant ($cm^2.s^{-1}$)	ρ (g/cm^{-3})	Dipole Moment
experimental	2.30×10^{-5}	0.995	2.6
TIP3P	5.06×10^{-5}	1.002	2.35
TIP4P	3.31×10^{-5}	1.001	2.18
TIP5P	2.62×10^{-5}	0.999	2.29
SPC	3.85×10^{-5}	0.985	
SPC/E	2.49×10^{-5}	0.998	2.35

Table 2.2: Thermodynamic properties at standard temperature and pressure calculated for various water models.[249,250,253]

University of Cape Town

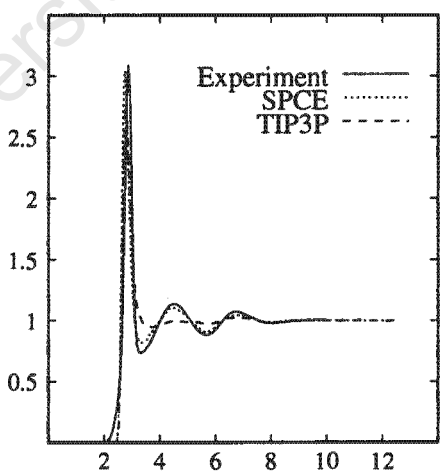


Figure 2.1: Oxygen–oxygen radial distribution functions ($g_{OO}(r)$) at 300K for the TIP3P and SPC/E models compared to the experimental profile[254] for liquid water.

of temperature and pressure.[249–251,253] A sensitive test is whether or not a water model can reproduce the second peak in the oxygen–oxygen pair correlation function, $g_{OO}(r)$ (Figure 2.1). In general, four–site models have been found to be superior to three–site models,[248,249] though the SPC/E improvement on SPC shows remarkably good results for a three–site model. Moreover, the five–site model TIP5P shows an improved diffusion constant as compared with its precursors, TIP3P and TIP4P.[253] Three–site models are, however, generally used as they are computationally cheaper and often more familiar. Of the three–site models, TIP3P exhibits less "structuring" than other water models (it fails to reproduce the $g_{OO}(r)$ second peak) and has a poor diffusion coefficient. However, the original SPC model achieved good results by over emphasising the water dimer interaction, resulting in an over estimation of the heat of vapourisation of water.[251] This was justified on the basis of the lack of polarizability in the model and self-energy corrections were applied in the subsequent SPC/E model to improve the heat of vapourisation. Though the strong water dimer interactions in the SPC/E model are not a problem for simulations of pure water, in solvated systems they result in an energetic imbalance with water–solute interactions being less favourable than they should be. The balance between solvent–solvent, solute–solvent and solute–solute interactions within a system is often referred to as the interaction triad. TIP3P is typically used as the standard water model with the CHARMM and AMBER force fields principally to ensure that the balance of the interaction triad is maintained. For this reason, the TIP3P model is used almost exclusively in this thesis, except for a comparison with the SPC/E model discussed in Chapter 4.

University of Cape Town

2.2 Molecular Mechanics: Energy Minimisation

Molecular mechanics energy minimisation is the most commonly used technique for carbohydrate conformational analysis. It is employed primarily for comparison of the energy of different molecular conformations with the goal of locating the global energy minimum conformation of a molecule.

Minimisation involves locating the atomic positions, r_n, where the force field function (Equation 2.1) takes on its minimum values (i.e. the first derivative is zero and the second derivatives are all positive). Minima are located using numerical methods, which gradually alter the coordinates to produce molecular conformations of increasingly lower energy until a minimum is reached. There are several options for minimisation algorithms, including non-derivative routines such as the simplex method, first derivative minimisation methods (e.g. the "steepest descents" and "conjugate gradients" methods) and second derivative

methods such as the Newton-Raphson algorithm.[215] These methods vary in their efficiency and reliability depending on the problem investigated. Typically, minimisation procedures begin with the more robust first-derivative algorithms and use the more accurate second-derivative algorithms for the final stages of minimisation. However, all minimisation algorithms suffer from the "multiple minima problem". This refers to the fact that the minimisation procedure will locate the minimum closest to the starting configuration, which may not be the global energy minimum for the system.

In order to find the global minimum, grid search methods are generally used to generate different starting positions for minimisation. As dihedral angles are the principal source of molecular flexibility and thus the main determinants of molecular shape, conformational analysis by minimisation typically involves a systematic search on a regular grid of all possible conformations of the molecular dihedral angles. Grid search methods systematically cover a region of phase space and therefore, in addition to pinpointing the global energy minimum, they can also provide an estimate of the potential energy surface of a molecule. Regular grid analysis involving 360 ° rotation in D° increments about N different rotatable bonds requires the calculation of $(360/D)^N$ conformations. Therefore, due to computational constraints, a grid search is only practical in a few dimensions.

University of Cape Town

2.2.1 Disaccharide Conformational Analysis

The conformational space available to a disaccharide is usually expressed in terms of two–dimensional Ramachandran contour maps of the conformational energy as a function of the glycosidic torsion angles, ϕ and ψ. These plots depict the potential energy surface for a glycosidic linkage in vacuum, giving an idea of the conformational freedom and the areas where steric clashes occur.

Ramachandran maps for a glycosidic linkage are calculated on a $n° \times n°$ grid in ϕ, ψ space. Originally, because of computational time constraints, only "rigid maps" were calculated for disaccharides, with the monomer units held in fixed conformations at each grid point. However, increases in computing power allowed for the calculation of relaxed, or adiabatic, conformational maps[82–84] where, at each grid point i,j, the ϕ and ψ angles are constrained while the other degrees of freedom are relaxed by energy minimisation. Relaxed maps typically show more conformational freedom for disaccharides than rigid maps.[83]

The actual energy ϕ, ψ energy surface for a disaccharide should represent the optimum (i.e. lowest energy) orientation of the secondary hydroxyl groups, and the hydroxymethyl groups for each ϕ, ψ point, as differences in substituent orientations can affect the energy

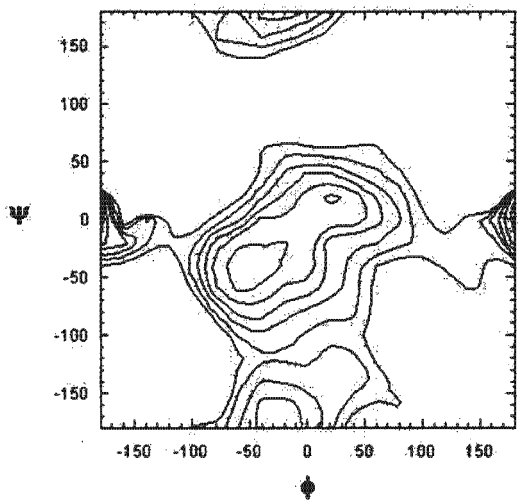


Figure 2.2: A two–dimensional relaxed adiabatic ϕ, ψ conformational energy map for β-D-maltose in vacuum.[72] The map was calculated on a 20° × 20° grid, contoured at 2 kcal/mol intervals above the global minimum, to a maximum of 12 kcal/mol.

University of Cape Town

by several kilocalories. Ideally, all other degrees of freedom would be searched while keeping the ϕ, ψ angles constrained. However, this is possible only for very small, rigid systems as it is computationally extremely intensive. For example, if one considers three staggered positions for each of the *exo*–cyclic groups in the maltose disaccharide, there are a possible (3^{10}) different starting positions for each ϕ, ψ conformation. Therefore, to save time, a disaccharide adiabatic map calculation usually considers only clockwise and anti–clockwise arrangements of the secondary hydroxyls and three orientations of each primary hydroxyl (*tg*, *gg* or *gt*) for every ϕ, ψ conformation. An example of such a map calculated for maltose is shown in Figure 2.2. Quicker, but less systematic, methods have been developed for disaccharide adiabatic map calculations, an example of which is simulated annealing.[87]

Ramachandran maps calculated for the maltose disaccharide using different force fields have produced qualitatively similar energy surfaces.[89] This can be expected to be the case for other disaccharides as well, as many of the broad outlines of the energy surface are produced by Van Der Waals repulsions arising from steric clashes between atoms on different rings, repulsions which are treated similarly by different force fields. However, the finer details in the low–energy regions of the maps are more sensitive to the subtleties of the force field and would thus be expected to differ with different force fields.

2.2.2 Normal Mode Analysis

Molecular mechanics can be used to calculate the vibrational normal modes of a molecule. The frequencies of the normal vibrational modes of a molecular system, together with the displacements of individual atoms, may be calculated from a force field using the Hessian matrix of second derivatives $\mathcal{V}''$. Suitable minimisation algorithms will supply the Hessian, which must then be converted to the equivalent force constant matrix in mass–weighted coordinates ($\mathbf{F}$):

$$\mathbf{F} = \mathbf{M}^{-\frac{1}{2}} \mathcal{V}'' \mathbf{M}^{-\frac{1}{2}} \tag{2.2}$$

$\mathbf{M}$ is a 3×3 diagonal matrix containing the atomic masses on the diagonal. Each non–zero element of $\mathbf{M}^{-\frac{1}{2}}$ is thus the inverse square root of the mass of the appropriate atom. The eigenvectors and values of matrix $\mathbf{F}$ are then usually obtained using matrix diagonalisation and the frequency of each normal mode subsequently calculated using:

$$v_i = \frac{\sqrt{\lambda_i}}{2\pi} \tag{2.3}$$

The frequencies calculated for a particular molecule can be compared with the results of spectroscopic experiments. This technique is often used in the parameterisation of a force field. In such cases, it is usually the lower–frequency vibrations that are of interest, as these correspond to the large–scale conformational motions of the molecule. In this thesis, vibrational mode analysis is used to validate parameterisation of the CSFF force field, as detailed in Chapter 4.

2.3 Molecular Dynamics Simulations

Molecular dynamics (MD) simulations aim to reproduce the dynamic behaviour of molecules, such as conformational fluctuations and interactions with solvent. MD simulations have a significant advantage over both Monte Carlo simulations and experiment, in that they provide a detailed history of the interactions and dynamics of individual particles over the course of a simulation.

The two essential elements for a dynamics simulation are the force field describing the interaction potential for the molecules involved in the system, from which the forces acting on the atoms can be calculated, and the equations of motion governing the dynamics of the system. The two principal simulation techniques used are standard Molecular Dynamics and Stochastic Dynamics. Although these methods use different equations of motion, "molecular dynamics", or "MD", is often used as a blanket term to refer to either of the techniques.

University of Cape Town

2.3.1 Molecular Dynamics

In the Molecular Dynamics approach, Newton's equations of motion are integrated simultaneously for all the atoms in the system:

$$\frac{d^2r_i(t)}{dt^2} = \frac{F_i}{m_i} \tag{2.4}$$

$$F_i = \frac{-\partial V(r_i, ...r_N)}{\partial r_i} \tag{2.5}$$

In these equations, the force on atom i is denoted by F_i and t denotes the time. The negative gradient of the potential energy, $V(r)$, are the forces, and therefore $V(r)$ must be a differentiable function of the atomic coordinates r_i.

2.3.2 Stochastic Dynamics

Stochastic dynamics is an extension of standard Molecular Dynamics, where a trajectory of a molecular system is generated by integration of the stochastic Langevin equation of motion:

$$\frac{d^2r_i(t)}{dt^2} = \frac{F_i}{m_i} + \frac{R_i}{m_i} - \gamma_i \frac{dr_i(t)}{dt} \tag{2.6}$$

This equation has added two terms to equation 2.5: a stochastic force, R_i, and a frictional coefficient, γ_i. Stochastic dynamics can either be used to mimic the effect of solvent, where the random force and frictional drag terms approximate the overall effects of solvent surrounding a solute molecule, or to control and maintain temperature by establishing a coupling of the individual atomic motions to a heat bath. The stochastic term introduces energy and the frictional term removes energy from the system.

2.3.3 Integration of the Equations of Motion

Various algorithms may be used to integrate the equations of motion. The integration is performed in small time steps, typically 1 to 10 fs. The choice of integration step size is very important: the increased accuracy obtained with a small step size must be weighed against the longer real time simulations achievable with a larger step size. Thus, the most important feature of an integration algorithm is its ability to support a long integration time step with a small numerical error. The Verlet algorithm[255] is the most widely used. This method employs the equation:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t) \tag{2.7}$$

University of Cape Town

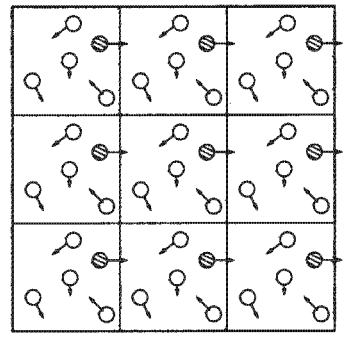


Figure 2.3: A two–dimensional illustration of cubic periodic boundary conditions.

where r is the position of an atom and a is its acceleration. A deficiency in the Verlet algorithm is that it involves adding a very small term ($\delta t^2 a(t)$) to a much larger term ($2r(t) - r(t - \delta t)$), which can lead to errors in computer floating point arithmetic. The leap–frog[256] variation on the Verlet algorithm allows for increased accuracy:

$$r(t + \delta t) = r(t) + \delta t v(t + \frac{1}{2}\delta t) \tag{2.8}$$

$$v(t + \frac{1}{2}\delta t) = v(t - \frac{1}{2}\delta t) + \delta t a(t) \tag{2.9}$$

In this method, the velocities are first calculated halfway between the time steps (Equation 2.9). The coordinates are then obtained from the previous position using the velocities calculated with Equation 2.9. The velocities and coordinates thus appear to "leap–frog" over each other. The Leap–Frog Verlet algorithm is equivalent to the standard Verlet algorithm, but avoids inaccuracies associated with terms in δt^2.

University of Cape Town

2.3.4 Simplifying Approximations

A number of simplifying approximations are often used to reduce the time required for computer simulations.

Periodic Boundary Conditions

Periodic boundary conditions are a means of simulating bulk fluids (or solids) using relatively small numbers of particles. The method involves placing the molecule and solvent system in a space–filling volume — a cube and a truncated octahedron are the most widely–used shapes — which is then treated as if it was surrounded by a periodic lattice of identical images of itself (for a cube, $3^3 - 1^3 = 26$ identical images). When a particle moves in the central shape, each of its images moves in the same way. When a particle

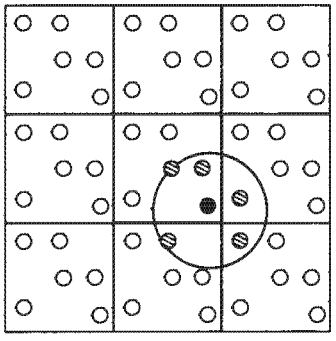


Figure 2.4: A two–dimensional illustration of a spherical cut–off and the minimum image convention.

leaves the shape, it re–enters at the opposite side in the translated image position with identical velocity, as illustrated in Figure 2.3.4. This procedure has the advantage that molecules at the surface of the shape interact with molecules in adjacent images and therefore do not diffuse away. A periodic boundary system thus removes surface effects and provides a representation of bulk solvent or solid.

When using periodic boundary conditions, it is essential that the image boundaries are treated so as to minimise the effects of the systemic periodicity. The size of the shape chosen is important, as a simulated particle should not be affected by the periodicity of the lattice, resulting in the symmetry of the cell structure being imposed on what should be an isotropic fluid. The minimum cell size required for this depends on the intermolecular potential and thus long–ranged potentials need to be dealt with correctly.

Truncating Long–Range Forces and the Minimum Image Convention

For large systems, calculation of the non–bonded interactions is the most time–consuming part of a simulation. For a pairwise interaction model, the number of non–bonded interaction terms increases with the square of the number of atoms and thus the computation time is proportional to N^2. In order to speed up calculations for large systems, the minimum image convention and a non–bonded cut–off are usually applied.

The minimum image convention requires that the interactions for a molecule i in the simulation are only calculated between that molecule and the *closest* image of every other molecule in the system. When a cut–off is used, all interactions between atoms that are further apart than the cut–off value are set to zero. For consistency with the minimum image convention, when using periodic boundary conditions for a cube of length L, the cut–off radius must not exceed $\frac{1}{2}L$.

Group–Based Cut–Offs

University of Cape Town

Simple spherical truncation of the non–bonded interactions on an atom–by–atom basis has been long known to lead to non–physical behaviour in simulations of dipolar molecules, such as water. The problem occurs because some atom–atom interactions are switched off before others as the separation between any two molecules approaches the cut–off distance. This results in truncated dipoles and thus spurious charge–dipole interactions. Group–based or residue–based cut–offs are used to avoid violent fluctuations in the non–bonded interaction energy as a molecule or group of atoms moves through the cut–off boundary. Atoms are collected into (preferably neutral) groups of adjacent atoms and interactions are calculated on a group–by–group basis, even if some of the atoms in the group are beyond the cut–off radius.

Cut–Off Smoothing

The truncation of the intermolecular potential at the cut–off distance creates some problems in defining a consistent potential and force for use in the MD simulation: the cut–off introduces a discontinuity in the potential function at the cut–off distance, which results in a gradual increase in the kinetic energy and thus the temperature of the system. This can be avoided by using smoothing functions to bring the non–bonded interactions to zero in a continuous fashion. Two of the most widely used methods are "switching" (Equation 2.10) and "shifting" (Equation 2.11) functions.

$$\begin{aligned} S_w(r_{ij}) = &\quad 1 && r_{ij} \leq r_{on} \\ = &\quad \frac{(r_{off}^2 - r_{ij}^2)(r_{off}^2 + 2r_{ij}^2 - 3r_{on}^2)}{(r_{off}^2 - r_{on}^2)^3} && r_{on} < r_{ij} \leq r_{off} \\ = &\quad 0 && r_{ij} > r_{off} \end{aligned} \tag{2.10}$$

$$\begin{aligned} S_f(r_{ij}) = &\quad (1 - \frac{2r_{ij}^2}{r_{cut}^2} + \frac{r_{ij}^4}{r_{cut}^4}) && r_{ij} < r_{cut} \\ = &\quad 0 && r_{ij} > r_{cut} \end{aligned} \tag{2.11}$$

Switching functions are polynomials in the separation distance that smoothly "turn off" the interaction potential over a given range between a "cut–on" distance and a "cut–off" distance. The interaction potential is unaffected for distances less than the cut–on distance. In contrast, a shifting function modifies the interaction potential over its entire range, going to zero at the cut–off distance.

Smoothing and truncation procedures can be based on an atom–by–atom basis or may be applied to predefined neutral groups of molecules. A switching function applied on a

University of Cape Town

group–by–group basis has been shown to be the superior method as it introduces fewer spurious effects such as artificial solvent structuring.[257]

Constraint Dynamics

The time step to be used in a molecular dynamics simulation is dictated by the highest frequency motion present in the system. However, bond vibrations involving light atoms such as hydrogen occur so rapidly that an extremely short time step is required to solve the equations of motion. These high frequency motions are usually of less interest than the lower frequency modes which often correspond to major conformational changes. Constraint dynamics enables selected coordinates to be constrained during a simulation without affecting the other degrees of freedom, thus allowing the time step size to be increased. This is possible because, in polyatomic molecules, the fast internal vibrations are usually decoupled from the rotational and translational motions.

The most common constraint method used in MD simulations is Ryckaert, Ciccotti and Berendsen's SHAKE algorithm.[258] SHAKE is a special technique developed to treat the dynamics of a molecular system where certain arbitrarily selected degrees of freedom (such as bond lengths) are constrained, while others are free to evolve under the influence of intermolecular and intramolecular forces. A difficulty is that the magnitude of the constraining force is not known. SHAKE therefore uses a set of undetermined (Lagrangian) multipliers to represent the magnitudes of the forces. The approach is to solve the equations of motion for one time step in the absence of constraint forces and subsequently determine their magnitudes and correct the atomic positions. The constraints are fulfilled exactly at each step of the integration. This method has advantages in that it produces a Cartesian form of the equations of motion with the irrelevant internal degrees of freedom eliminated.

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

2.3.5 Simulation Conditions and Ensembles

A Molecular Dynamics simulation has four phases: initialisation, heating, equilibration and production. Initialisation provides an initial position and velocity for all the atoms. Velocities are usually assigned randomly either from a Gaussian or a uniform distribution at low temperature. Heating is typically then performed and involves running an MD

University of Cape Town

simulation while gradually increasing the kinetic energy up to a final desired temperature. The simulation then continues into an equilibration stage, the process where the kinetic energy and the potential energy are allowed to distribute evenly throughout the system. Once the system has stabilised, the production phase begins. This step is used to collect the trajectory information for analysis.

It is often desirable to compare the results of a simulation with a physical experiment. Molecular dynamics is traditionally run under a microcanonical ensemble, with conditions of constant temperature, volume and energy (NVE). However, other ensembles often correspond better to experimental conditions. Constant temperature and volume (NVT) or temperature and pressure (NPT) are the most common alternative ensembles employed. The NPT ensemble corresponds to the Gibb's free energy and is therefore the ideal ensemble for comparison to experiment. However, the NVT, or canonical, ensemble is relatively easier to implement. Periodic boundary conditions result in a fixed volume for a system. The temperature is kept constant either by rescaling the velocities of the atoms at intervals, or by coupling the system to a heat bath or by using the Langevin equations of motion in Stochastic Dynamics. Where it is used in this thesis, Stochastic Dynamics fulfills this latter function, enabling the temperature of a NVT ensemble to be maintained without the usual method of velocity rescaling.

University of Cape Town

2.3.6 Parallel MD Simulations

Most of the computer programs written for MD simulations, such as CHARMM, were originally developed for serial machines. Simulation of large molecules requires enormous computing power, particularly where solvent is to be explicitly included into the simulation. One way to reduce the time required to achieve such simulations is to use parallel computers and parallelised programs. In recent years distributed memory parallel computers have been offering cost–effective computational power, particularly in the case of the increasingly popular Beowulf–class PC clusters.[259] Therefore, many traditional biomolecular programs have been gradually recoded so that most of their components can run in parallel, typically using message–passing software such as MPI. The CHARMM program was one of the first to be tackled.[260] Some of the more latterly developed packages, such as DL_POLY[225] and NAMD[261] have been developed from the start with parallel processing in mind.

The simulations performed in this work were run using a parallel version of CHARMM27 on a Beowulf PC cluster comprising 16 Pentium III dual processor machines (with a CPU speed of 900 MHz each) . This enabled PMF calculations that would otherwise have

been too computationally expensive. The methods used to perform these free energy calculations are discussed in the next chapter. However, some calculations, such free energy calculations involving explicit solution simulations of longer oligosaccharides, remain impracticably time–consuming.

University of Cape Town

University of Cape Town

Chapter 3

Free Energy Calculations Using Adaptive Umbrella Sampling

University of Cape Town

Free energy is generally considered to be the most important concept in physical chemistry.[216] Free energy differences and barriers between molecular systems determine which reactions or conformational changes are spontaneous as well as the rate at which they occur. Molecular conformational equilibria are determined by the relative free energies of different molecular conformations. Likewise, the rate of conformational transition is controlled by the free energy barriers between low–energy molecular conformations.

Free energy calculations differ as to whether they calculate pure free energy differences (such as free energies of solvation or differences between ligand–host binding energies) or rather a potential of mean force (PMF) describing a complete free energy pathway (such as that between a reactant and product or for a change in a coordinate). The principal advantage of calculating PMFs, as opposed to simple free energy differences, is that a detailed description of barrier heights and energy changes along an complete free energy pathway allows for more reliable predictions of the relative populations and rates of change between reactants and product or different conformations.

Calculation of a potential of mean force requires knowledge of the equilibrium population of the degree(s) of freedom under study. In principle, this may be obtained from a single long molecular dynamics run that samples the significantly populated minima sufficiently often. However, in practice if the free energy difference between two conformations (or the free energy barrier separating them) is greater than $2kT$, the energy surface will not be thoroughly explored in a standard MD simulation.

Special techniques have been developed to increase the sampling of conformational space in Molecular Dynamics simulations and thus enable calculation of PMFs from trajectories. The most established are thermodynamic integration (TI) and free energy

perturbation (FEP). These methods can be used to calculate free energy differences for both conformational and chemical changes. Depending on the calculation required, FEP and TI require a series of simulations along either a reactant–product mutation pathway (free energy differences) or a reaction coordinate (PMF). As free energy is a state function and both of these simulation techniques invoke reversible paths, the mutation pathway chosen need not be an actual physical pathway. However, each simulation along the path must be run for long enough to ensure thorough equilibration of the hybrid system.

Another technique widely used for PMF calculations is the adaptive umbrella sampling method. This was first developed by Mezei[262] and has been used in a number of studies of small molecules,[88,88,263–266] including disaccharides.[89] Umbrella, or non–Boltzman, sampling refers to simulations run with a biasing, or "umbrella", potential in order to enhance sampling in the high energy regions that are unexplored by regular simulations. *Adaptive* umbrella sampling is an iterative procedure that employs a series of biased umbrella simulations. The biasing potentials improve "adaptively" with successive simulations and finally converge on a potential that results in complete coverage of conformational space in a single simulation. The PMF is then calculated from this final umbrella potential. Adaptive umbrella sampling has the advantage that, unlike FEP and TI, it is clearly apparent when convergence has been achieved. An improvement of the adaptive umbrella sampling protocol employs the Weighted Histogram Analysis Method (WHAM) in order to obtain the umbrella potential for each simulation.[229,264–269] WHAM is an automated method that enables incorporation of the data from a number of simulations into the estimate for a potential of mean force and is easily extendible to multiple dimensions.

University of Cape Town

Other innovative simulation methods have been developed in order to increase coverage of phase space. Examples are the "locally enhanced sampling"[110] and the "jumping between wells" procedures.[270] These methods both use simulations run with a combination of Monte Carlo and MD techniques.

The PMF calculations presented in this work all employed the adaptive umbrella sampling method together with the Weighted Histogram Analysis Method. The theory behind these techniques and the details of their implementation will now be discussed in more depth.

3.1 Theory

Given the probability distribution function, $P(\bar{\xi})$, for a generalised multidimensional degree of freedom, $\bar{\xi}$, the potential of mean force, $W(\bar{\xi})$, can be obtained using the relation:

$$W(\bar{\xi}) = -kT \ln P(\bar{\xi}) \tag{3.1}$$

where k is the Boltzmann constant and T is the temperature.* However, it is usually not possible to obtain a satisfactory estimate of $P(\bar{\xi})$ from a standard Molecular Dynamics or Monte Carlo simulation. Adaptive umbrella sampling simulations therefore use a modified potential function, H:

$$H = H^{\circ} + U(\bar{\xi}) \tag{3.2}$$

In this expression H° is the Hamiltonian of the system and $U(\bar{\xi})$ is an applied biasing, or "umbrella", potential chosen in order to direct the system into unsampled regions. The unperturbed probability density, $P(\bar{\xi})$, may be obtained from the biased probability density generated by a simulation, $P'(\bar{\xi})$, using:

$$P(\bar{\xi}) = CP'(\bar{\xi})e^{\beta U(\bar{\xi})} \tag{3.3}$$

where C is an arbitrary constant and $\beta = -\frac{1}{kT}$.

The aim is to chose an umbrella potential which will flatten the PMF and therefore achieve uniform sampling in the entire region of $\bar{\xi}$ during a standard simulation. Clearly, the ideal choice for $U(\bar{\xi})$ would be $W(\bar{\xi})$, as uniform sampling along $\bar{\xi}$ can be achieved by setting

$$U(\bar{\xi}) = kT \ln P(\bar{\xi}) \tag{3.4}$$

which is simply the negative of the potential of mean force. However, as $W(\bar{\xi})$ is not initially known (the aim of the exercise being to calculate it), a guess is made and the approximation improved iteratively. Successive simulations are run with "adapting" umbrella potentials that will eventual converge to the inverse of the actual PMF.

3.2 The Iterative Procedure

The umbrella sampling method used in this thesis is similar to that outlined by Bartels and Karplus.[265] The degree(s) of freedom selected for study, $\bar{\xi}$, is divided into m bins of index k. Then an iterative procedure is used to converge on the potential of mean force for $\bar{\xi}$, as follows:

*The bar over ξ indicates that this parameter may be multidimensional.

University of Cape Town

1. Simulation i, of length N_i is performed with an umbrella potential $U_i(\overline{\xi}_k)$. At each step of the simulation, the umbrella potential for a specific value of $\overline{\xi}$ is calculated from a spline[271] of the grid points of $U_i(\overline{\xi}_k)$.

2. The biased probability distribution for simulation i is computed as an unnormalised histogram, $n_{i,k}$, of the values of $\overline{\xi}$ occurring during the simulation.

3. The biased distribution functions, $n_{j,k}$, for all simulations $j = 1, 2, ..., i$ are combined using the Weighted Histogram Analysis Method (detailed below) to obtain an optimal estimate for the unbiased distribution function, $\overline{p}_k^\circ$.

4. The current potential of mean force estimate, $W_{i,k}$, is calculated from $\overline{p}_k^\circ$ using Equation 3.1.

5. $W_{i,k}$ is extrapolated to as yet unexplored regions and smoothed.

6. The procedure halts if the PMF calculation has converged. Otherwise, $U_{i+1,k}$ is set to $-W_{i,k}$, i is incremented and the procedure returns to Step 1.

Simulation i=1 is usually an unbiased simulation ($U_{1,k} = 0$ for all k), unless a reliable estimate of the surface can be made. The procedure converges when a single biased simulation achieves a "uniform" distribution of $\overline{\xi}$ i.e. all conformations have been approximately equally sampled throughout the simulation.

3.3 The Weighted Histogram Analysis Method

The Weighted Histogram Analysis Method (WHAM)[265–269] is a means whereby overlapping distribution functions obtained from several successive biased simulations are combined to achieve an optimal estimate for the actual unbiased distribution function. WHAM has at least two advantages over conventional splicing methods for combining distributions: none of the data are discarded and the method is not inherently limited to one dimension.

WHAM is a generalisation and extension of the histogram technique developed by Ferrenberg and Swendsen.[272,273] The unbiased distribution function is expressed as a weighted sum of the i individual biased distribution functions and a set of coupled non-linear WHAM equations are solved to produce optimal weighting factors for each of the distribution histograms. The derivation and self–consistent solution of the WHAM equations have been detailed elsewhere.[265,267] A brief outline of the WHAM equations is given here, using the formulation and notation of Bartels and Karplus.[265,266]

University of Cape Town

After i simulations, the optimal unbiased probability distribution histogram, $\overline{p}_k^\circ$, for a degree of freedom $\overline{\xi}$ is found by iterating the following WHAM equations:

$$\overline{p}_k^\circ = \frac{\sum_j n_{j,k}}{\sum_j N_j f_j c_{j,k}} \tag{3.5}$$

and

$$\overline{f}_j = \frac{1}{\sum_k c_{j,k} \overline{p}_k^\circ} \tag{3.6}$$

with

$$c_{j,k} = e^{-\beta U_j(\overline{\xi}_k)} \tag{3.7}$$

$$N_j = \sum_k n_{j,k} \tag{3.8}$$

In these equations, f_j is the weighting factor for the jth simulation, $\overline{\xi}_k$ denotes a value of $\overline{\xi}$ that falls in the kth bin, $n_{j,k}$ is the histogram of the reaction coordinate from the kth window in the jth simulation and N_j is the number of configurations stored during the jth simulation. Equations 3.5 and 3.6 are iteratively applied until convergence is achieved. This is usually judged to occur when the maximum difference between the weighting coefficients f_j from the previous and current iterations, $\max |f_j^i - f_j^{i-1}|$, is less than a certain tolerance (typically 0.001). However, this criterion for convergence causes accuracy problems with floating point arithmetic when the differences in the shifting coefficients f_j are of large orders of magnitude. This can occur when the free energy surface has very large differences between maxima and minima. In such cases, convergence is judged to be achieved when the *ratio* of the difference between current and previous iterations and the current weighting coefficient ($\frac{f_j}{\max |f_j^i - f_j^{i-1}|}$) is less than the tolerance factor.

3.4 Extrapolation and Smoothing.

Once calculated, the umbrella potential for the next simulation must be extrapolated into areas not yet explored. Ideally, extrapolation should direct the simulation into unsampled regions without introducing additional local minima into which the system could become trapped.[266] Two methods of extrapolation were implemented. The first emulates the system of Bartels and Karplus,[265] with the unsampled regions set to the maximum value of the umbrella potential. However, in order to have a smooth transition from explored to

University of Cape Town

unexplored areas, the "border" regions were handled in the following manner. Unsampled bins in the umbrella potential grid that were directly adjacent to sampled bins were set to the average value of the surrounding cells. Unsampled cells adjacent to these border cells were set to the average of the surrounding cells plus half the difference between the average and the maximum. In this way, sharp discontinuities in the umbrella potential were avoided. The second extrapolation method uses an estimate of the PMF for extrapolation to the unexplored areas. This is typically applied where a PMF calculated in vacuum can be used as an estimate for the PMF in solution. However, though an accurate estimate of a PMF can potentially aid convergence of the adaptive umbrella sampling procedure, in practice the first method was found to be generally the most effective as it has less potential for introducing artificial minima.

Subsequent to extrapolation, any discontinuities in the umbrella potential introduced by extrapolation were limited by smoothing the potential map at least twice in each dimension using the following function:[265]

$$U'_{k,l} = \frac{1}{3}(-0.3U_{k-2,l} + 1.3U_{k-1,l} + U_{k,l} + 1.3U_{k+1,l} - 0.3U_{k+2,l}) \tag{3.9}$$

The umbrella potential surface must be smoothed to ensure correct calculation of the forces on the atoms.

3.5 Umbrella Sampling Methods and Considerations

The various 1D and 2D umbrella sampling calculations undertaken in this thesis were implemented via modifications to the CHARMM USERE routine (version 27b1).[222] The USERE routine was adjusted in order to add the energy and forces from the umbrella potential to the specified atoms involved in the degree(s) of freedom under investigation. Two general types of PMF calculation were performed. The first type encompasses both one– and two–dimensional PMFs for rotation about specified dihedral angles. In this case, the forces determined from the umbrella potential are applied to the four atoms of each torsion angle in order to drive rotation about the central bond. The second type of PMF calculation is of one-dimensional end–to–end distance PMFs. The end–to–end distance is defined from a selected atom on one end of a molecule to another on the opposite end. The force calculated from the applied umbrella potential is added to the two specified atoms in order to adjust the extension of the molecule. End–to–end adaptive umbrella sampling aims to converge on an umbrella potential that will result in uniform sampling of the chosen range of end–to–end distances (distances that are so short as to cause steric clashes or so large as to stretch the molecule far beyond its usual range of motion are

University of Cape Town

excluded). End–to–end PMFs are thus used to describe the energy required to extend and compress a molecule. They are frequently calculated for polymer strands,[229] but have not yet been reported for oligo– or polysaccharides.

Umbrella sampling simulations were performed using the CHARMM program with the adjusted USERE routines. The CHARMM program was compiled to run in parallel on a Linux Beowulf cluster of computers. Each biased simulation consisted of an equilibration period (usually in the range of 100ps to 1ns) followed by a production period (in the range of 500ps to 20 ns) used to generate the $\overline{\xi}$ population histogram. The umbrella potential to be applied was specified in a file as a grid of energy points. At each time step during the simulation, the specific umbrella potential to be applied was calculated from a spline of the grid points (a cubic spline in the case of two-dimensional umbrella potentials).[271] In most cases, long simulations of around 10 ns were required so that the unconsidered conformational degrees of freedom in the molecule (such as the hydroxyl or hydroxymethyl groups) were thoroughly sampled. The slower their equilibration, the longer the simulations required and hence the slower the convergence of the whole procedure.

During each simulation, the modified USERE routine generated data consisting of the value of the degree of freedom $\overline{\xi}$ (e.g. angle, length etc.) and the associated value of the applied umbrella potential. Histograms of these data were created using a binning program. The latest estimate of the PMF was then generated from the histograms by a WHAM program, which incorporates the extrapolation and smoothing routines.

After initial experimentation, an automated procedure was developed in order to manage the extremely time–consuming series of simulation, histogram binning, WHAM application, smoothing and extrapolation of the umbrella potential. The procedure (in the form of a Linux shell script) could generally run for several weeks without interference. However, user intervention was sometimes required if the procedure was not converging. During the course of this work, a few general heuristics for achieving convergence were identified. The first point concerns the mesh size for the umbrella potential grid. Coarser grids are attractive in that they require less computer memory and result in faster computation of the splines. However, too large an interval between grid points hinders convergence, as a number of conformations of potentially different energies are merged into one bin. This is particularly the case in the steeper regions of the PMF. On the other hand, too fine a mesh is not only computationally expensive, but requires longer simulations in order to populate each bin sufficiently. Therefore, some effort must to put into identifying the optimum mesh size for the degrees of freedom under consideration.

Another point is that the very high energy regions in the PMF surface involving steric

University of Cape Town

clashes (peaks, not saddle points) are physically interesting only if one is concerned with dynamics at extreme pressures and temperatures. If explored, these areas vastly increase the convergence time and it is more efficient to avoid sampling them. Therefore, regions in the PMF maps with energy greater than a predetermined truncation point above the global minimum were excluded from the umbrella surface by setting any bin in the PMF greater than the cut–off to the cut–off value. The resultant plateaus in the umbrella potential were not sufficient to flatten the very sharp peaks in the steric clash regions and thus the subsequent simulation is in effect elegantly restricted to the lower energy regions. This approach resulted in a valuable increase in the speed of convergence of the two-dimensional glycosidic linkage PMF calculations. The location of the truncation point depends on the system being investigated, but must not be so low that the transition states crucial for equilibration of the system and for the correct sampling of local minima are excluded from the simulation.

It was found to be particularly important that the initial simulations were long enough to give a reliable estimate of the umbrella potential in the regions explored. If they were not, the convergence behaviour of the entire procedure was found to be adversely affected. Further to the point, the umbrella potentials obtained from the first few simulations in a procedure often cannot be trusted to be a reliable estimate of the PMF in the higher energy regions. Therefore a good method is to gradually increase the energy range of the umbrella potential as the PMF estimate improves. This was done by limiting the range of the umbrella potential by truncation to a few kilocalories per mol in the first few simulations and gradually increasing its range as the surface was more thoroughly explored. In this way, poor guesses of the PMF, which can lead to the trapping of the system in artificial minima, are avoided.

For faster coverage of the energy surface, it was found to be best if the simulations were begun from a variety of different starting structures. An easy way to do this is to use the final structure from the last simulation as the starting structure for the next simulation, but to reassign the velocities of the atoms. However, with this approach, it is possible that the system will become "trapped" in an unusual conformation. Therefore the original equilibrated starting position was reinstated after a specified number of simulations (here usually 8). It was also found to be a good practice to restrict the number of biased simulation distributions used to calculate the next PMF estimate to no more than 50. Therefore, as the number of simulations increased beyond this, the older, more inaccurate distributions were gradually discarded.

In the case of the 2D dihedral angle PMFs, convergence of the whole procedure was defined to occur when, for a single simulation, every bin k in $n_{j,k}$ in the ϕ,ψ range of

University of Cape Town

interest was occupied at least once. Tighter convergence criteria were used for both types of one-dimensional PMF, where the entire phase space was more easily sampled. For these PMFs, the final simulation was required to produce a ratio of most populated to least populated histogram bin of at most 5. This is a stricter requirement than that used by Kumar et al.[267,269] Finally, where convergence does not occur, detailed analysis of the system is required. Some simulations may, due to insufficient equilibration, deviate significantly from equilibrium behaviour. Other studies have found that the way in which deviations are handled has an important effect on the convergence behaviour, although less so on the final results obtained after very long simulations.[265] Convergence was found to be accelerated when data that appeared to involve large deviations were discarded. However, if this happens repeatedly, it is likely that longer simulation times are required in order to bring about crucial conformational changes and equilibration of the unconsidered degrees of freedom.

3.6 Stretching Simulations

The adapted USERE routine for calculating end-to-end PMFs can be equally well applied to other problems that require modification of the molecular potential. In this work, the end–to–end PMF protocol was used to simulate atomic force microscopy stretching experiments on single oligosaccharide molecules. AFM stretching experiments represent the ideal conditions for the application of molecular simulations in that experimental data that requires interpretation is available for reference. Once this data has be satisfactorily reproduced, simulations can be inspected in detail to provide insight into the dynamic processes occurring as a response to the applied force.

For the stretching simulations, a straight line with a constant slope was employed as the applied umbrella potential. This results in a constant force (equal to the negative of the slope) being applied to the ends of the molecule. Previously reported simulations of forced elongation used a different method, with a molecule fixed on one side and extended by a spring.[51,119]

During each stretching simulation, the force was ramped at regular intervals with subsequent equilibration. Data was collected during the subsequent production phase. Ramping the force involved increasing the slope of umbrella potential by a fixed amount. Each simulation produced a series of distributions corresponding to the extension of the molecule at specific forces. These distributions were used to generate simple force–extension profiles by plotting the force applied at each step against the average molecular extension determined from the corresponding distribution. The force-extension graphs

University of Cape Town

were then compared to those produced by AFM experiments on the same molecules. If the two curves were sufficiently similar the simulations were then analysed to determine which molecular events were responsible for the principal features of the curve. In addition, in some instances PMFs for stretching the oligosaccharides were obtained by combining the overlapping distributions using WHAM. A more accurate force–extension profile was then calculated from the derivative of this curve and compared to the first graph. However, the application of WHAM was limited to the initial stages of the stretching simulations because of floating point number overflow errors that occurred with the calculation of the exponential terms ($e^{-\beta U_j(\bar{\xi}_k)}$) with large applied umbrella potentials.

Once an experimental stretching curve is satisfactorily reproduced, these simulations are valuable in that they can be thoroughly inspected in order to clarify which conformational transitions contribute to the stretching behaviour. This helps to resolve ambiguities in the molecular mechanisms involved during polysaccharide stretching and thus provides insight into macromolecular properties like stiffness and elasticity. Stretching simulations also provide insight into the dynamical behaviour of the polysaccharides when heated, as the application of a force resembles the application of heat in that it reduces the activation energy of conformations not occupied at room temperature.[119]

University of Cape Town

Chapter 4

Force Field Parameterisation

4.1 Introduction

The principal degrees of freedom for the relatively rigid hexopyranose rings are the *exo*–cyclic hydroxymethyl and the secondary alcohol groups. For accurate computer simulations of pyranose carbohydrates in solution, a force field that reproduces the experimental hydroxymethyl conformational distribution (defined and discussed in Section 1.1.1) and rate of rotation is required. However, this has not always been achieved. Hydroxymethyl sampling problems have been a general concern in carbohydrate simulations[129] and adjustments to the hydroxymethyl torsion angles have been found to be necessary for a number of force fields.[109,214,242,274] The CHARMM parameter sets are no exception. The original CHARMM parameter set for carbohydrates,[181] HGFB, is known to produce incorrect populations of the primary alcohol rotamers in most simulations, as it favours the *tg* conformation over *gt* and *gg* in both vacuum and solution.[74,88] Palma et al. produced a new CHARMM parameter set,[218] PHLB, which aimed to correct the relative energies of the hydroxymethyl, as well as the unrealistic flexibility of the HGFB carbohydrate model.* Adiabatic map calculations for PHLB indicated correct relative energies for the three hydroxymethyl conformations ($gg < gt < tg$). However, the PHLB parameter set has been found to exhibit very infrequent conformational transitions of the hydroxymethyl, to the extent that nanosecond MD simulations do not achieve equilibrium populations.[72,275]

Parameterisation of carbohydrate force fields has typically focused on obtaining agreement with *ab initio* conformational calculations; reproduction of the solution conformational distributions has been largely ignored in the parameterisation process. X–ray structures and *ab initio* results have traditionally been used for parameterisation and validation

*see Section 2.1.3.

University of Cape Town

of the dihedral angles in carbohydrate force fields, with an emphasis on obtaining correct relative conformational energies. For example, the specific dihedral terms in the PHLB force field were parameterised to reproduce experimental vibrational frequency data and *ab initio* dihedral angle rotational energy profiles for small molecules such as ethylene glycol.[199] This was also the method used by Kony et al. for the OPLS–AA force field: the torsional barriers were adjusted and scaling factors for the electrostatic interactions were applied in order to reproduce the calculated *ab initio* rotational profiles.[242] However, comparison of the relative energies of hydroxymethyl conformations in vacuum does not take into account the contributions of entropy and solvation to the equilibrium conformer distribution. In addition, little attention has been paid to the time scale of transition between the hydroxymethyl conformations, which is determined by the height of the free energy barriers between conformational minima. An accurate force field for molecular dynamics calculations in solution should reproduce not only the experimental equilibrium conformational distribution, but also the experimental time scale for the primary and secondary alcohol group rotations in solution.

The equilibrium hydroxymethyl conformational distributions produced by a force field are usually estimated from the populations obtained from extended MD simulations.[242] However, hydroxymethyl transitions are fairly infrequent on the time scale of molecular dynamics simulations. There have been several reports of MD simulations with various force fields exhibiting very few hydroxymethyl rotational transitions.[110,141] In addition, difficulties with achieving a representative sample for (1→6)–linked oligosaccharides using a CHARMM parameter set have been recently reported.[57] This is in agreement with experimental measurements — exchange between the different conformations of the hydroxymethyl groups has been estimated to occur on a time scale of between 10^{-10} and 10^{-3} seconds,[123,141] with recent studies placing rotational isomerisation more specifically on the nanosecond time scale.[61,188] Therefore it is difficult to judge from a standard nanosecond MD simulation whether or not the hydroxymethyl rotameric populations are correctly represented by a given force field, as the system will not have reach equilibrium.

The equilibrium populations produced by a parameter set for the hydroxymethyl group can only be completely established by calculating the PMF profile for rotation about the C_5–C_6 bond. A PMF provides not only the relative energies of the three hydroxymethyl conformers, but also the barriers to rotation and the equilibrium rotamer distributions. In this study, the behaviour of the hydroxymethyl group in the PHLB and HGFB parameter sets was quantified by calculation of potential of mean force for rotation about the primary alcohol dihedral, ω, for both β–D–glucose and β–D–galactose (Figure 4.1). PMFs were calculated in both vacuum and solution to investigate the effects of water on

University of Cape Town

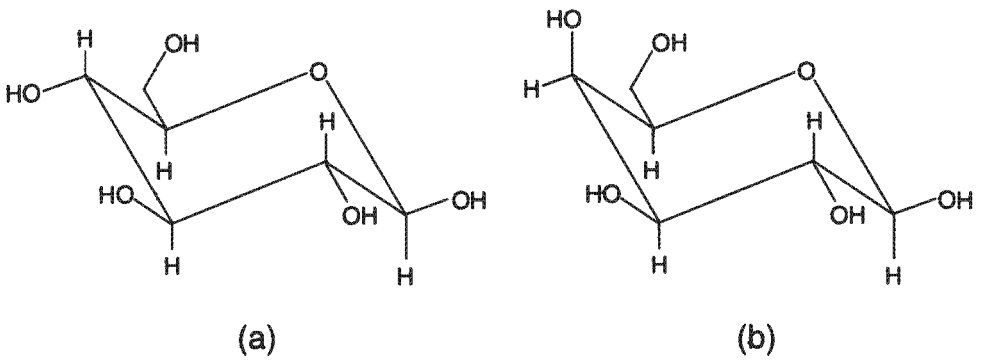


Figure 4.1: Line diagrams of β–D–glucose (a) and β–D–galactose (b).

the conformational distribution. Based on the results obtained, adjustments were made to the PHLB parameter set in order to alter the relative free energies of the *gg*, *tg* and *gt* hydroxymethyl conformations and lower the rotational energy barriers for both the hydroxymethyl groups and the secondary alcohols. For ease of reference, this modified force field is termed the Carbohydrate Solution Force Field (CSFF). Solution potential of mean force profiles for β–D–glucose and β–D–galactose were calculated for the CSFF in order to confirm that it will exhibit hydroxymethyl conformer equilibrium distributions and rotational frequencies in agreement with experimental observations. The validity of the force field was further established by a comparison of the calculated and observed normal modes of vibration of α–D–glucose. Lastly, the PMF ω profile for glucose in SPC/E solution was calculated. This assists in quantifying the differing effects of the SPC/E[251] and TIP3P[248] water models on the hydroxymethyl conformational preferences.

University of Cape Town

4.2 Methods

4.2.1 Parameterisation of the Torsional Terms

In order to adjust the primary alcohol rotation without affecting the parameterisation of the ring atoms in the force field, a new atom type for the C_6 carbon (CPS) was introduced into the CSFF parameter set. This atom type allowed for isolation of the hydroxymethyl dihedral angle parameters from those of the pyranose ring carbons and can be considered to be a generic non–ring carbon.

The PHLB parameter set contains specific dihedral terms. Therefore, aside from energetic contributions from non-bonded interactions, the energy profile for rotating about an A–B bond is determined by a sum of all the dihedral angles terms for every quartet of atoms X–A–B–Y around the bond. In particular, the rotational energy profile about

the C_5–C_6 bond has contributions from nine dihedral terms, each of which can have up to six multiplicities. The force constants of the various cosine terms were adjusted in order to lower the energy barriers to rotation about the C_5–C_6 bond, with particular emphasis on the *gg–gt* barrier. In order to adjust these dihedral terms with predictable results, it was found that careful analysis of the phases of the dihedral terms in relation to each other was required. Some indication that correct adjustments were made was obtained by graphical plotting of the sum of the dihedral terms. However, as the torsional profile is determined by a host of non-bonded interactions as well as the dihedral terms, calculation of the PMF for C_5–C_6 bond rotation was required to establish correct parameterisation. Alterations to the dihedral force constants were restricted to the O_5–C_5–C_6–O_6, O_5–C_5–C_6–H_6 and C_4–C_5–C_6–O_6 dihedral angles, to ensure that the ethylene glycol rotational profile remained parameterised to closely approximate the profile obtained from *ab initio* calculations.[199] The k_θ values for all other hydroxymethyl dihedral angles remain the same as the CTS atom type in the PHLB force field. The new dihedral term parameters incorporated into the CSFF force field are listed in Table 4.1.

Dihedral	k_n	**n**	θ_0
Hydroxymethyl			
OES – CTS – CPS – HAS	0.2086 (0.1618)	3	0.0
OES – CTS – CPS – OHS	-4.0193(-3.7993)	1	0.0
	-1.2688 (0.5686)	2	0.0
	-0.9704 (0.4202)	3	0.0
CTS – CTS – CPS – OHS	-1.7139 (-1.9139)	1	0.0
	-1.0239 (-0.3739)	2	0.0
	-0.0340 (-0.0340)	3	0.0
Secondary alcohol			
HOS – OHS – CTS – HAS	0.0677 (0.1686)	3	0.0
HOS – OHS – CBS – HAS	0.0677 (0.1686)	3	0.0
HOS – OHS – CPS – HAS	0.0677 (0.1686)	3	0.0

Table 4.1: New dihedral force constants (k_n) added to the CSFF parameter set, showing the periodicity (n) and phase angle (θ_0) for the cosine terms ($E_\theta = \sum k_n(1+cos(n\theta-\theta_0)), n = 1,2,3$). Units are in kcal/mol. The original PHLB values appear in brackets.

In addition to the primary alcohol dihedral angle adjustment, force constants for all the *H–O–C–H* dihedral angles were reduced to 40% of their original values to speed up the slow rotation rate of the secondary hydroxyls. Except for the O–H bond force constant

University of Cape Town

which was lowered to 460.5 $Å^{-2}$ mol^{-1} to better match the experimentally measured O–H stretching frequency, all other parameters were transferred unchanged from the PHLB set. The complete CHARMM topology and parameter files for the CSFF are listed in Appendices B and C respectively.

4.2.2 PMF Calculations

Using the adaptive umbrella sampling method described in Section 3.5, the potential of mean force surfaces about the β–D–glucose ω dihedral angle were calculated using the HGFB, PHLB and CSFF carbohydrate parameter sets in both vacuum and TIP3P[248] water solution. In the case of the CSFF, a glucose ω solution PMF using the SPC/E[251] water model was also calculated. In addition, ω dihedral PMF profiles for β–D–galactose were calculated for the PHLB and CSFF parameter sets, in vacuum and TIP3P solution.

In each case, the umbrella potential surface was represented as one–dimensional grid of points, with a grid separation of 2.5°. The first simulation in each series was started from a minimised and equilibrated coordinate set and run with no applied umbrella potential surface. During a simulation, the biasing potential energy at a particular position was calculated from a spline of the umbrella energy surface. Each biased simulation consisted of alternating periods of equilibration (typically 500 ps) and production (varying from 500 ps to up to 10 ns). After each simulation, the biased distribution function histograms was calculated for the 2.5° bins over the entire range of the hydroxymethyl dihedral angle by summing the number of configurations in each bin over the production phase of the trajectory. The WHAM procedure was used in order to obtain an optimum combination of all the $n_{i,k}$ distributions and thus the next estimate of the PMF, using a tolerance value of 0.001. This PMF surface was extrapolated into unexplored regions by setting unsampled bins to the maximum value of the sampled bins and then smoothed three times using Equation 3.9. The umbrella potential for the next simulation was set to the negative of this PMF estimate and no truncation procedures were employed. Convergence of the PMF calculations was defined to have been achieved when a single simulation explored the entire range of the ω dihedral angle and the resultant population histogram had a ratio between the least and most occupied histogram bin of less than 5.

4.2.3 Simulation Conditions

Molecular dynamics simulations were performed using the program CHARMM[222] (version 27b1), with modifications incorporated into the USERE module in order to implement the one-dimensional adaptive umbrella sampling PMF calculations for the ω dihedral angle.

University of Cape Town

For the solution simulations, minimised structures of the monosaccharides were each placed in an equilibrated cube of 512 TIP3P or SPC/E water molecules. Solvent water molecules that overlapped with the solute molecules were removed and the system was equilibrated for 500 ps. The solution simulations for both β–D–glucose and β–D–galactose incorporated one monosaccharide molecule surrounded by 499 TIP3P or SPC/E water molecules in a cube of length 24.79 Å with periodic boundary conditions. This gives a density of 1.013 g/cm^3, which is the approximate experimental density of an aqueous glucose solution of this weight percentage. The cube was subjected to minimum image periodic boundary conditions to eliminate edge effects.

Initial velocities for the atoms were selected at random from a Boltzmann distribution at 300 K. The molecular dynamics trajectories were integrated as microcanonical (constant n, V, E) ensembles using a leap–frog Verlet integrator[256] with a 1 fs time step at 300 K. The non–bonded interactions were truncated using a switching function applied on a neutral group basis between 10.0 and 12.0 Å. The groups corresponded to electrically neutral collections of atoms in the carbohydrate molecules and entire water molecules for the solvent. The SHAKE algorithm[258] was used to fix the water molecule geometry and covalent bonds between hydrogen and the heavy atoms.

University of Cape Town

4.2.4 Normal Mode Analysis

The normal modes for α–D–glucose were calculated for the CSFF and PHLB parameter sets using the CHARMM program, which diagonalises the mass–weighted Hessian (Section 2.2.2). The α–D–glucose structure for the normal mode calculation was taken from the neutron–diffraction data of Brown and Levey,[117] with subsequent minimisation. Three stages of Newton–Raphson minimisation were employed: the first with all the hydroxyl and the hydroxymethyl torsion angles constrained to their experimental configuration, the second with constraints only on the secondary hydroxyl dihedral angles and the final with all constraints released. The resultant conformation was closer to that reported in the crystal structure than one obtained using constraint–free minimisation. Vibrational frequency assignments were made on the basis of the potential energy distributions for each mode and used to match the CSFF and PHLB modes. The calculated frequencies were compared with the experimental frequencies measured by Dauchez et al.[276] using the vibrational assignments for their calculated normal modes to match frequencies.

glucose		*gg*	*gt*	*tg*
	exp.[234]	52	41	7
	exp.[124]	53	45	2
	HGFB	24	27	49
	PHLB	85	15	0
	CSFF	66	33	1
galactose		*gg*	*gt*	*tg*
	exp.[123]	12	56	32
	exp.[125]	18	61	21
	PHLB	13	86	1
	CSFF	4	75	21

Table 4.2: Hydroxymethyl equilibrium conformer populations in β–D–glucose and β–D–galactose for the HGFB, PHLB and CSFF parameter sets compared to experimental values.

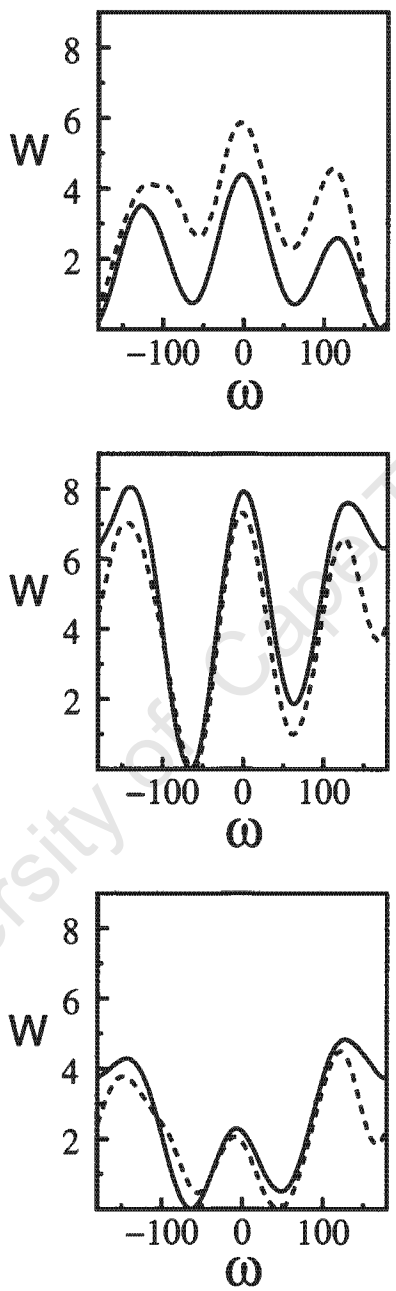
4.3 Results and Discussion

The ω PMF profiles calculated for β–D–glucose using the HGFB, PHLB and CSFF parameter sets in both vacuum and solution are shown in Figure 4.2. Table 4.3 lists the corresponding conformer populations calculated from the PMF curves (using Equation 3.1).

The ω solution PMF for the HGFB force field clearly indicates that the hydroxymethyl rotational profile for this force field does not agree with experimental observations: the *tg* conformation is lowest in energy in both vacuum and solution. This is further shown to be the case by the solution equilibrium conformer distribution calculated from the PMF curve, which is not compatible with experimental populations.

The PHLB parameter set shows a free energy order of $gg < gt << tg$ — a clear improvement on the HGFB set. Moreover, the equilibrium distribution calculated from the PMF profile for β–glucose is *gg*:*gt*:*tg* = 85:15:0. This ratio is in better agreement with experiment than the HGFB parameter set, although the *gg* conformation is populated somewhat more than is generally predicted. However, the energy barriers between conformations are 7 to 8 kcal/mol in height, which is more than 10×kT. These high barriers make it improbable that transitions will occur between the different conformations during nanosecond room temperature (298K) molecular dynamics simulations. This is illustrated by the ω time series for a β–D–glucose TIP3P solution simulation (Figure 4.3): a 10 ns molecular dynamics simulation started from *gg* maintained this conformation throughout and equilibrium was not achieved.

University of Cape Town



University of Cape Town

Figure 4.2: β–D–glucose: vacuum (– – –) and TIP3P (—) solution PMF profiles (kcal/mol) obtained for the hydroxymethyl ω dihedral using the three parameter sets: HGFB (top), PHLB (middle) and CSFF (bottom).

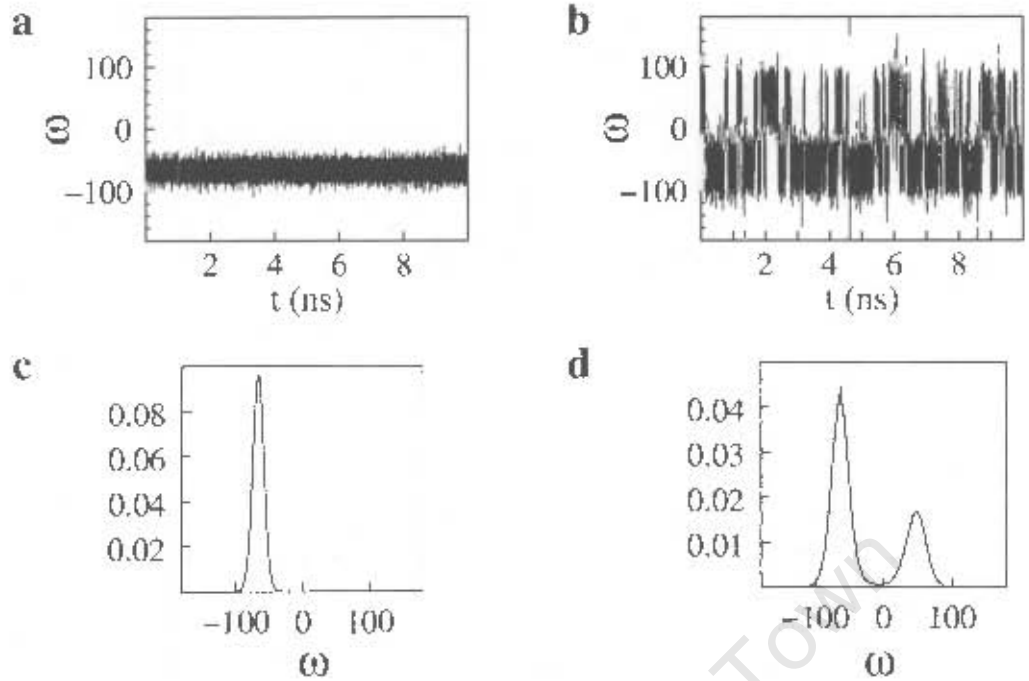


Figure 4.3: β–D–glucose: times series and statistics from 10 ns MD simulations using the PHLB (left column) and CSFF (right column) parameter sets. Top: Hydroxymethyl ω dihedral time series plots (PHLB (a) and CSFF (b)). Bottom: Probability distribution profiles for the ω dihedral (PHLB (c) and CSFF (d)).

The CSFF ω PMFs are significantly different from the corresponding PHLB profiles. In the case of glucose (Figure 4.2), the *gg* conformation is still favoured in solution, though the relative energies of the *gt* and *tg* conformations have been lowered by about 1.5 and 2.0 kcal/mol respectively. The calculated solution equilibrium conformational distribution of *gg*:*gt*:*tg* $\approx$ 66:33:1 shows a better match with estimates from NMR experiments than that obtained using the PHLB parameter set. Further, the barriers to rotation about the hydroxymethyl dihedral have been significantly lowered compared to those from the PHLB parameter set, which is reflected by the increased frequency of transition seen in Figure 4.3 (b). For the CSFF parameter set, rotation of the hydroxymethyl group takes place on a nanosecond time scale, in agreement with recent experimental estimates.[61,188] In addition, a 10 ns simulation of β–D–glucose produced a rotameric distribution of *gg*:*gt*:*tg* $\approx$ 69:31:0 (Figure 4.3), indicating that hydroxymethyl equilibrium populations were achieved in this time frame.

The ω free energy profiles for galactose in water and vacuum (Figure 4.5) show a similar decrease in the CSFF ω rotational barriers compared with those in the PHLB force field. As with glucose, the lowered barrier heights in CSFF facilitate rotation about the C_5–C_6 hydroxymethyl bond, which results in more rapid equilibration of this group (Figure 4.4).

University of Cape Town

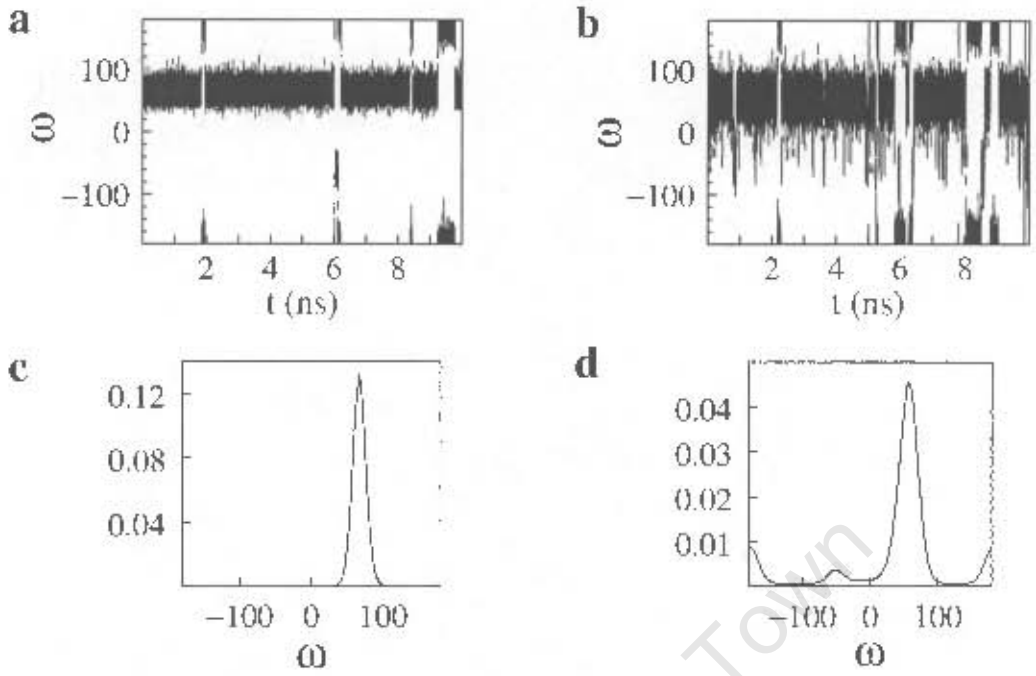
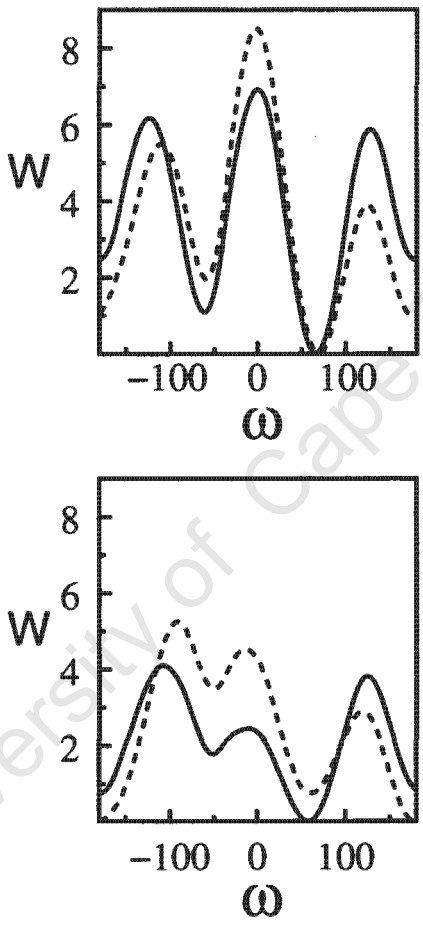


Figure 4.4: β D galactose: times series and statistics from 10 ns MD simulations using the PHLB (left column) and CSFF (right column) parameter sets. Top: Hydroxymethyl dihedral ω time series plots (PHLB (a) and CSFF (b)). Bottom: Probability distribution profiles for the ω dihedral (PHLB (c) and CSFF (d)).

Calculated population distributions for the galactose PMF curves using these force fields are both comparable with experimental results (Table 4.3). Though the CSFF parameter set favours the *gt* conformation more than is predicted, the relative preference for the *tg* conformation over *gg* compares well with experimental observations. The β-D-galactose distributions calculated from 10 ns solution simulations (Figure 4.4) are PHLB: *gg*:*gt*:*tg* $\approx$ 1:91:8 and CSFF: *gg*:*gt*:*tg* $\approx$ 6:80:14, indicating that CSFF, in contrast with PHLB, has in fact achieved equilibrium.

The effect of solvent on the relative sampling of the three different rotamers is revealed by comparisons of the vacuum and solution PMF profiles. For each of the parameter sets, the effect of the TIP3P solvent is to raise the energies of the *gt* and *tg* conformations relative to *gg*, by approximately 1 kcal/mol and 2.5 kcal/mol respectively for the PHLB and CSFF parameter sets, and 0.5 and 2 kcal/mol for the HGFB case. Solvation thus favours the *gg* conformation over *gt* and the *gauche* over the *trans* ω conformations. This accords with quantum mechanical studies of glucose in aqueous solution by Cramer and Truhlar, where unfavourable solvation of *tg* was found to destabilise this conformation relative to *gt* and *gg*.[102] Destabilisation of the *tg* conformation is more pronounced for the glucose PMF calculation performed with the SPC/E water model (Figure 4.6). Though

University of Cape Town



University of Cape Town

Figure 4.5: β–galactose: vacuum (– – –) and TIP3P (—) solution PMF profiles (kcal/mol) obtained for the hydroxymethyl ω dihedral using the PHLB (top) and modified CSFF (bottom) parameter sets.

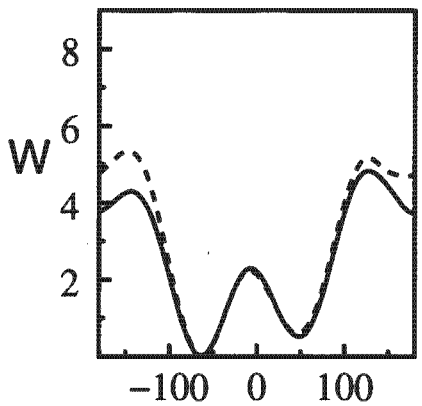


Figure 4.6: β–D–glucose: SPC/E (– – –) and TIP3P (——) solution PMF profiles (kcal/mol) obtained for the primary alcohol ω dihedral using the CSFF parameter set.

Dihedral	Average value (degrees)
O5 - C1 - O1 - HO1	-63.7
C1 - C2 - O2 - HO2	62.6
C2 - C3 - O3 - HO3	-70.4
C3 - C4 - O4 - HO4	81.1
C5 - C6 - O6 - HO6	180

Table 4.3: Average values for the glucose hydroxyl dihedral angles obtained from a 10 ns simulation in TIP3P water.

University of Cape Town

the *gg* and *gt* conformations have the same relative free energies in TIP3P and SPC/E solution, the *tg* conformation is 0.5 kcal/mol higher in energy in the SPC/E solution. The SPC/E water model has very similar atomic charges to the TIP3P water model, but differs in its geometry, which results in the better reproduction of water structure reported for this model (Section 2.1.4). As discussed in Section 2.1.4, the SPC/E model also features stronger water–water interactions, which results in the water–solute interactions being relatively disfavoured. The fact that the *tg* conformation is more disfavoured in SPC/E solution lends weight to the argument that the *tg* conformation is disfavoured in solution because it disrupts the three dimensional structure of water.[76]

The secondary hydroxyls showed increased frequency of rotation for both glucose and galactose with the new CSFF parameter set. As glucans are the primary subject of this work, the hydroxyl behaviour in glucose was analysed further. The time series histograms for the β-D-glucose hydroxyl groups obtained from the 10 ns CSFF simulation in TIP3P water are shown in Figure 4.7. The profiles for the O_2 and O_3 hydroxyls are essentially mirror images of each other. The O_4 dihedral, though more restricted than O_2 and O_3,

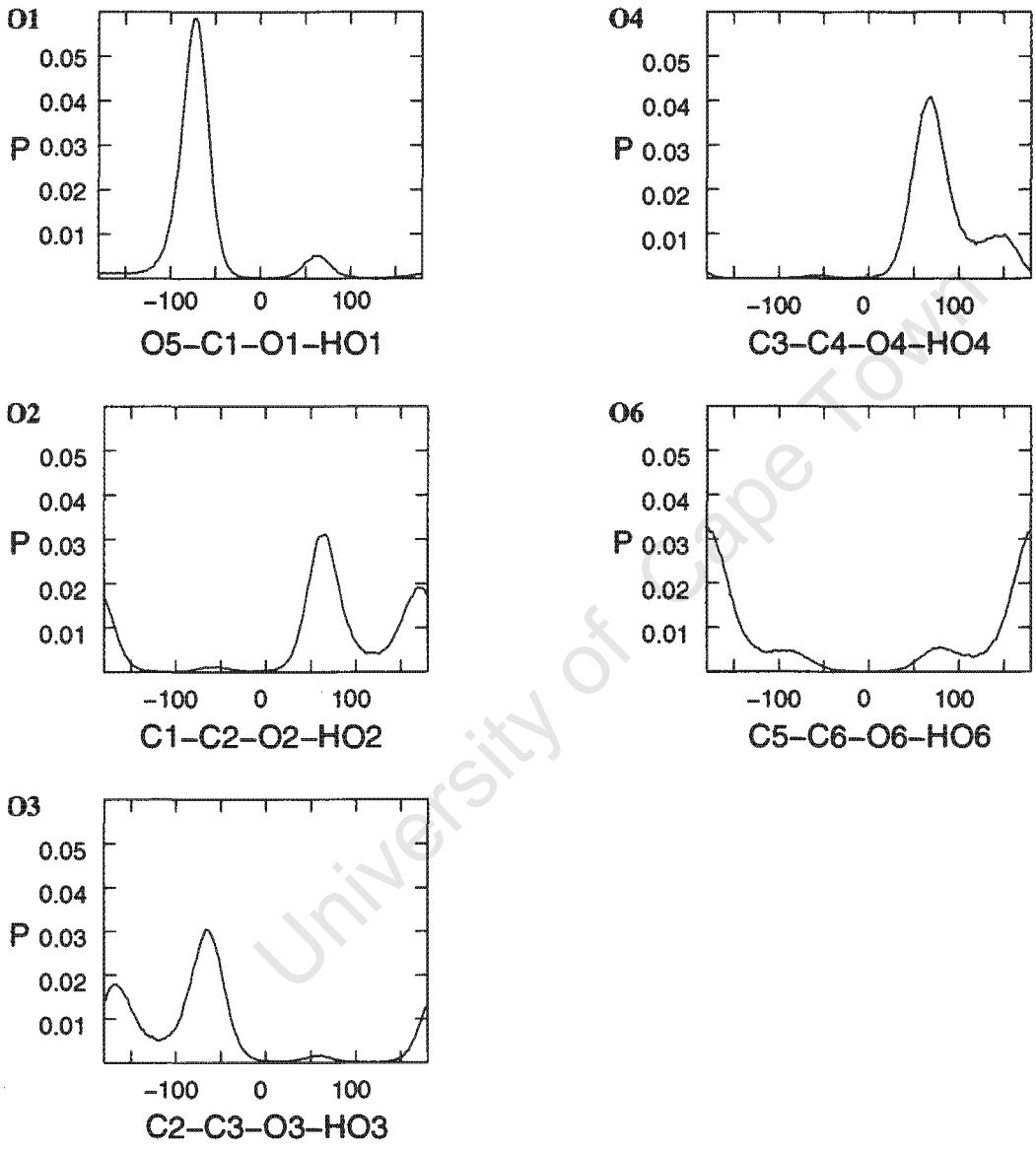


Figure 4.7: Probability distributions for the glucose H–O–C_X–C_{X-1} dihedral angles obtained from a 10 ns simulation in TIP3P water.

University of Cape Town

has a similar distribution. Though these ring hydroxyls are principally in a clockwise arrangement in solution, there is a significant population of conformations with anti-clockwise hydroxyl arrangements. This is not the case for the O_1 dihedral. The least mobile of the ring hydroxyls, it persists almost exclusively in the clockwise orientation. This is unexpected, as O1 has greater freedom to rotate than the other ring hydroxyls. Its relatively fixed conformation is probably the result of very favourable interactions with water, such as strong single hydrogen bonds and cooperative water bridges to the ring oxygen and the hydroxymethyl group.

The normal modes of α–D–glucose calculated for both the CSFF and PHLB parameter sets are reported in Appendix A along with the experimental frequencies[276] and an approximate description of each mode in terms of bond stretches, angle bends and torsion angle rotations. Both parameter sets show good agreement with observed frequencies. This is particularly encouraging as the PHLB force field was not specifically parameterised to fit saccharide vibrational frequencies. The fit with experiment is especially close in the mid– to low–frequency range. These are the most important modes for a carbohydrate force field to reproduce, as this is the region of torsional vibrations and complex motions of the pyranose ring. The normal modes calculated for PHLB and CSFF are similar and are generally within a few wave numbers of each other. The CSFF force field matches the experimental frequencies better in the very high frequency region associated with O–H and C–H stretches. This was achieved by lowering the O–H bond force constant, as mentioned earlier. However, it should be noted that these stretches are irrelevant in most MD simulations as a result of the use of the SHAKE algorithm to constrain bonds involving hydrogen atoms (see page 45).

Aside from the O–H stretching region, the CSFF and PHLB parameter sets differ most in the modes associated with the C_5–C_6 and hydroxyl dihedral angles, as is to be expected. Encouragingly, CSFF modes in these regions compare better with experiment than do the PHLB predicted modes. However, both CSFF and PHLB exhibit seven modes in the 1500 to 1800 cm^{-1} region that have no counterparts in the experimental spectra. These modes involve bond stretches and angle bends for the ring atoms and are a reflection of the high bond and angle force constants for the ring atoms in the CSFF and PHLB parameter sets. High force constants were used in the parameterisation of the PHLB force field in order to avoid the extreme ring flexibility observed for HGFB force field.[217,277]

University of Cape Town

4.4 Conclusions

The dihedral angle force constants in the PHLB force field were adjusted to correct the hydroxymethyl behaviour and the new Carbohydrate Solution Force Field (CSFF) validated by calculating potential of mean force profiles in solution and vacuum for β–D–glucose and β–D–galactose. The PMF profiles confirm that that Molecular Dynamics simulations using this force field will produce equilibrium population distributions and rotational frequencies in solution that are consistent with experimental evidence. The rotational frequency of the secondary hydroxyl groups has also been increased, which should in turn aid equilibration of these groups in solution. This force field is thus a marked improvement on the PHLB parameter set. In addition, normal modes calculated for α–D–glucose in vacuum compare favourably with experimental data.

Moreover, as the frequency of the hydroxymethyl rotation has been increased for the CSFF, the hydroxymethyl group equilibrates faster and equilibrium populations of the hydroxymethyl conformations are now achievable with nanosecond simulations. This is necessary for carbohydrate simulations where rotation of the hydroxymethyl is important for the molecular conformations and dynamics, such as for the PMF calculations reported in subsequent chapters of this thesis. In addition, the CSFF force field should display improved equilibration and dynamics for $\alpha(1\rightarrow6)$–linked saccharides, such as isomaltose and dextran. These are molecules are investigated in Chapter 8.

University of Cape Town

University of Cape Town

Chapter 5

The $\alpha(1\rightarrow4)$ Glycosidic Linkage I: Linkage Properties

University of Cape Town

5.1 Introduction

Chapter 4 detailed the adjustments that were made to the PHLB carbohydrate parameter set to improve the relative conformer populations and rotational dynamics of the hydroxymethyl C_5–C_6 bond. Using this new force field, free energy studies of selected glycosidic linkages are now described, with the principal aim of quantifying the influence of a water solvent on their conformer population distributions and the rotational dynamics. This has implications for the conformations and behaviour of polysaccharides in water as well as the effects of water on carbohydrate glasses.

In this chapter, the subject of investigation is the important Glc–$\alpha(1\rightarrow4)$–Glc glycosidic linkage, which is the backbone glycosidic linkage in three abundant biological storage polysaccharides: amylose, amylopectin and glycogen. Although this linkage has not been completely characterised, a number of experimental and computational investigations have provided some insights into its conformation and dynamics. Firstly, the $\alpha(1\rightarrow4)$ glycosidic bond in glucans is known to be conformationally restricted compared to other linkages: adiabatic map calculations of the ϕ, ψ angles in maltose have shown that only a small domain of ϕ, ψ space is energetically accessible to maltose at ordinary temperatures.[34,134] Secondly, the $\alpha(1\rightarrow4)$–linkage is known to be sensitive to water. Various experiments have indicated that, when solvated, the maltose glycosidic angle conformation shifts from the crystal structure values[19,21] of $\phi, \psi = 5°$, 13° to conformations closer to the ϕ, ψ = -30,-30 region.[31,42–44,47] This finding has been corroborated by molecular dynamics simulations[68,72,86,181] and an *ab initio* study on a model compound for the

$\alpha(1\rightarrow4)$–linkage, 2–methoxytetrahydropyran.[128] The effects of solvation on the $\alpha(1\rightarrow4)$–linkage are of interest not only for solution conformations, but also because of the known dramatic effect of water on the physical properties of both maltose[37,164,278] and amylose glasses.[6,12–14,278,279]

In addition to interactions with the solvent, there are other factors which may influence the behaviour of the $\alpha(1\rightarrow4)$–linkage. The *exo*–cyclic hydroxymethyl group is thought to affect the rotational dynamics, relaxation behaviour and preferred conformations of glycosidic linkages in $\alpha(1\rightarrow4)$–linked glucans, by both unfavourable steric clashes and favourable hydrogen bonding interactions.[4,60] However, the effect of the primary alcohol on disaccharide conformations has not yet been quantified.[89] Moreover, the influence of adjacent glucose units on the conformational space and relative energy of the $\alpha(1\rightarrow4)$–linkages is unpredictable. Bulky extra units would be expected to restrict the conformational space, but cooperative intramolecular interactions may contribute to altering the relative energies of the minima and perhaps fix a polysaccharide into a particular shape.

This study aims, within the limits of the force field used, to provide a full characterisation of the $\alpha(1\rightarrow4)$–glycosidic linkage, including the effects of aqueous solvation, the hydroxymethyl group and additional monomers upon its mobility and dynamics. The behaviour of the $\alpha(1\rightarrow4)$–linkage is investigated by free energy studies on three model carbohydrates — the disaccharides maltose and dixylose and the 4–unit maltotetraose. The ϕ, ψ potential of mean force surfaces for β–maltose and β–dixylose in both vacuum and solution and maltotetraose in vacuum are presented.

Maltose (4–*O*–α–D–glucopyranosyl–D–glucopyranose), shown in Figure 5.1 (a)) is the simplest conceptual model for the $\alpha(1\rightarrow4)$ glycosidic linkage in glucans. The effects of water on the $\alpha(1\rightarrow4)$-linkage are to be elucidated by calculation and comparison of the ϕ, ψ PMF surfaces in vacuum and aqueous solution. Dixylose (4–*O*–α–D–xylopyranosyl–α–D–xylopyranose, Figure 5.1 (b)) is a pentose disaccharide analogue of maltose, lacking the complication of the *exo*–cyclic hydroxymethyl groups. Although not investigated experimentally, dixylose has previously been simulated as a convenient model for maltose.[89] However, here dixylose is investigated as a model for the behaviour of an $\alpha(1\rightarrow4)$–linkage without the hydroxymethyl group. Comparison of PMFs calculated for maltose and dixylose in both vacuum and solution is expected to shed light on the effect of this group on the preferred conformations of the $\alpha(1\rightarrow4)$–linkage. Although both maltose and dixylose have been investigated previously with partial PMF calculations,[89,181] this work represents the first complete ϕ, ψ PMFs calculated for maltose and dixylose. In addition, novel one–dimensional potential of mean force surfaces for rotation around the virtual C_1–C_4' bond

University of Cape Town

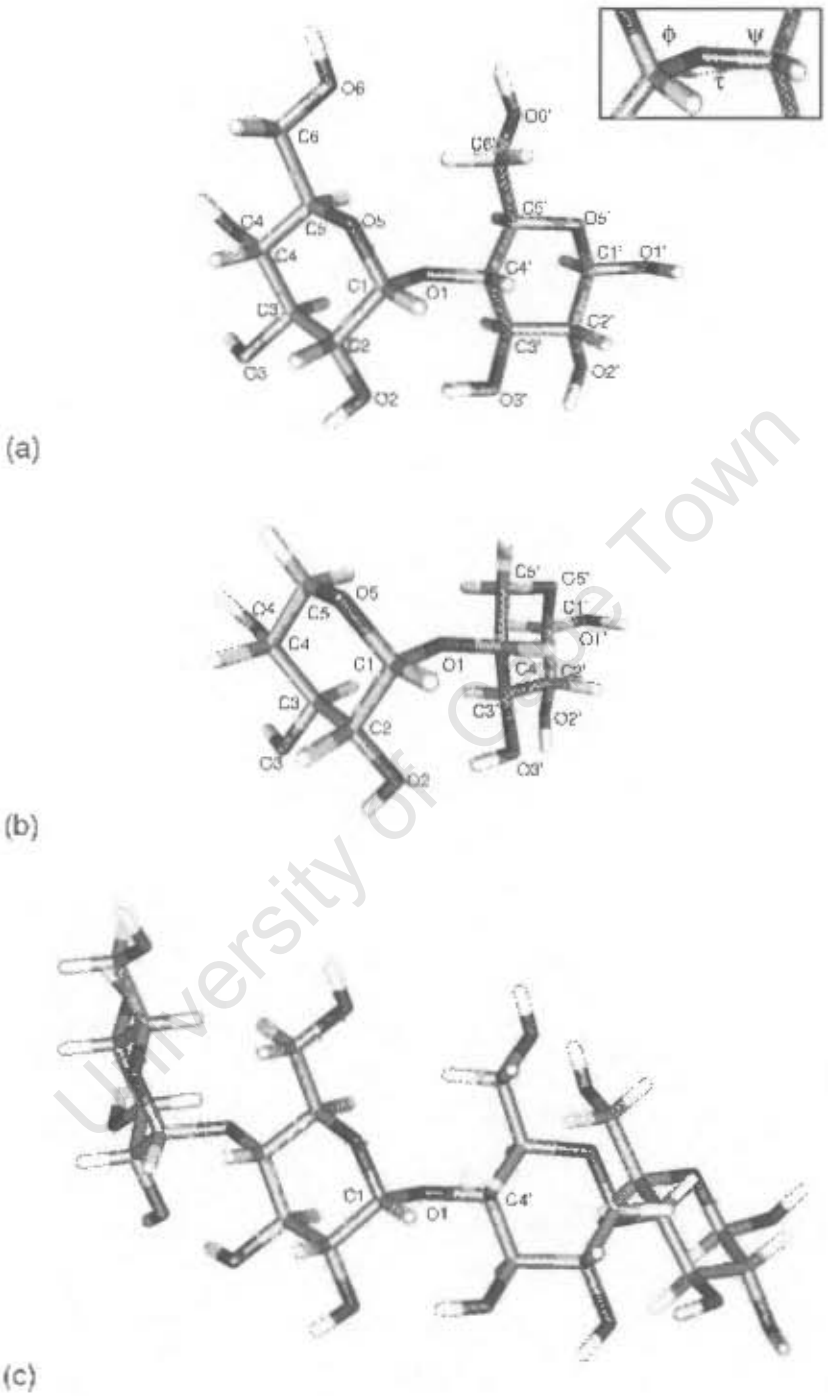


Figure 5.1: Three-dimensional structures of maltose (a), dixylose (b) and maltotetraose (c), showing the atomic naming convention used. The inset shows the bonds of rotation for the ϕ, ψ and τ dihedral angles.

University of Cape Town

described by the dihedral angle τ (Section 1.1.2) are calculated from the solution and vacuum PMFs for maltose and dixylose. These give a clear one–dimensional representation of the minima and energy barriers in the ϕ, ψ PMFs. Finally, the effect of adjacent glucose residues on the $\alpha(1\rightarrow4)$-linkage is clarified by investigation of maltotetraose (Figure 5.1 (c)), a four unit $\alpha(1\rightarrow4)$ linked oligosaccharide. The tetramer calculation was only carried out in vacuum, because of the extremely long simulations (on the order of tens of nanoseconds) required in order to ensure that the low energy regions of the other degrees of freedom are properly sampled. A calculated PMF in vacuum for the central glycosidic linkage in maltotetraose is expected to quantify the effect of additional monomers on the conformational freedom of the central linkage, and thus provide a better estimate of the $\alpha(1\rightarrow4)$–linkage in the amylose polymer.

5.2 Methods

5.2.1 2D ϕ, ψ PMF Calculations

Using the iterative adaptive umbrella sampling method described in Section 3.5, ϕ, ψ PMFs were calculated for β–maltose and β–dixylose in vacuum and aqueous solution and for β–maltotetraose in vacuum only. The β–conformations of these molecules were simulated as they are expected to be favoured in solution, in accordance with the anomeric ratio for glucose in water.[94, 96] The ϕ and ψ glycosidic linkage dihedral angles in each of these molecules were defined in Section 1.1.2. In the case of maltotetraose, the PMF was calculated for only the central glycosidic linkage (as shown in Figure 5.1) and the other two linkages were allowed to relax freely.

In each case, the $360°\times360°$ umbrella potential surface for the ϕ, ψ dihedral angles was represented as a two–dimensional grid of points, with a grid separation of 2.5°. Although in two dimensions a $2.5°\times2.5°$ grid is computationally expensive, too coarse a grid was found to provide an inadequate estimate of the disaccharide PMFs, particularly in the steeper barrier regions. At each integration step of a simulation, the biasing potential energy corresponding to the current ϕ, ψ position was calculated from a cubic spline of the umbrella energy surface. After each simulation, a two-dimensional ϕ, ψ distribution histogram was calculated by summing the number of configurations in each $2.5°\times2.5°$ bin over the production phase of the trajectory. The WHAM procedure was then applied, using a tolerance value of 0.001, to obtain an optimum combination of all the simulations distributions and thus the next estimate of the PMF. This PMF surface was extrapolated into unexplored regions by setting unsampled bins to the maximum value of the sampled

University of Cape Town

bins, treating the border regions as described in Section 3.4. The resultant PMF surface was smoothed three times using Equation 3.9. Subsequently, regions in the 2D ϕ, ψ PMF associated with high-energy steric clashes of the atoms were removed. This was done by setting any bin in the PMF with an energy value greater than a predetermined cut-off of 20 kcal/mol to the cutoff value. The umbrella potential for the next simulation was set to the negative of this PMF estimate.

In each umbrella simulation, an equilibration period of 500 ps was followed by a data collection period used to obtain the ϕ, ψ distribution histogram. The data collection phase was in the range of 1 to 10 ns for the disaccharides, in order to allow the hydroxymethyl groups to equilibrate. The increased frequency of rotation for the hydroxymethyl in the CSFF force field facilitated calculation of the PMFs, as the PHLB force field would require much longer simulation times for in order to achieve equilibrium populations of the hydroxymethyl groups. In the case of maltotetraose, simulations had to be extended to up to 20 ns in order to achieve full sampling of the non-central ϕ, ψ angles.

Successive simulations in a series were gradually run for longer periods as the PMF estimates improved and more of the phase space was explored. The first simulation in each series was started from a minimised and equilibrated coordinate set and performed with a flat umbrella potential surface. Subsequent simulations were begun from the final conformation of the previous simulation, to facilitate gradual exploration of the whole of conformational space. After every 8 simulations, the next simulation was begun from the original starting configuration, but with the velocities reassigned. Convergence of the whole adaptive umbrella sampling procedure was defined to occur when, for a single simulation, every bin k in $n_{j,k}$ in the ϕ, ψ range of interest was occupied at least once. Up to 200 simulations were used to obtain the final PMF, but at any time not more than the latest 50 distributions were used to estimate the next PMF; earlier simulations were gradually discarded as new, more extensive distributions were produced. A total of approximately 500 ns of simulation time was required to produce each of the final PMF surfaces. This is considerably more than in the previous partial PMF calculation for dixylose (8.7 ns).[89] For convergence of the disaccharide PMF calculations in solution, approximately 4 months of CPU time were required, using simulations run in parallel on eight 900 MHz Pentium II processors of a Beowulf PC cluster.

Equilibrium ϕ, ψ populations were obtained from the calculated PMF curves using the relation $P = e^{-\frac{W}{kT}}$. One-dimensional PMFs for the ψ angle only were calculated by adding all the equilibrium ϕ ψ populations to obtain the total population for a particular value of ψ. From the resultant ψ population histogram, a PMF was calculated. The forces required to overcome a barrier could be obtained from this curve from the negative

University of Cape Town

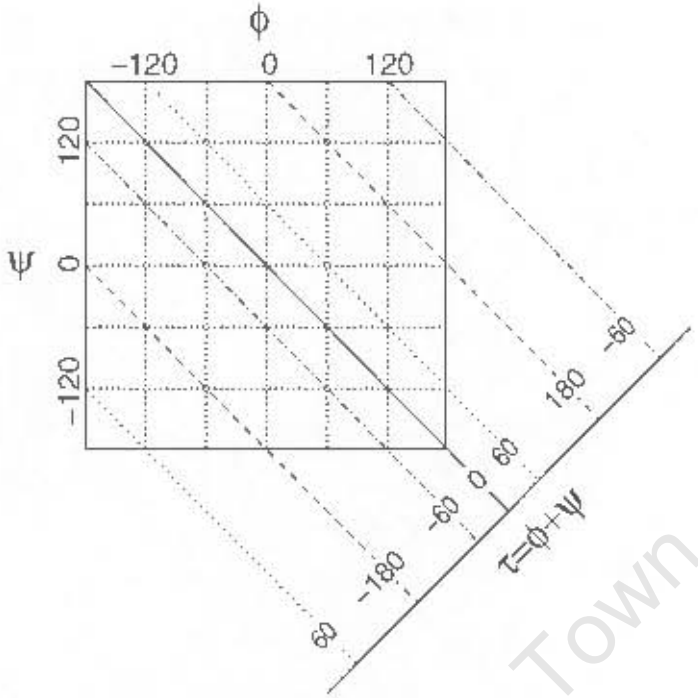


Figure 5.2: Diagram of the relationship between the τ and ϕ and ψ dihedral angles ($\tau = \phi + \psi$).

University of Cape Town

of the maximum slope of the barrier (using $F = -\frac{\delta E}{\delta \psi}$). In addition, novel 1D τ PMFs were calculated from the ϕ, ψ PMFs by a simple procedure. As the value of the τ dihedral angle is simply a sum of ϕ and ψ, the population of a particular τ point is obtained by summing all the contributing ϕ, ψ populations. This is illustrated in Figure 5.2 — the ϕ, ψ histogram bins corresponding to grid points falling on the diagonal line representing a particular τ bin are summed to produce the τ population. The τ PMF can then be calculated from the normalised τ population distribution.

5.2.2 Adiabatic Map Calculation

A relaxed adiabatic ϕ, ψ map for β maltose was calculated for comparison with the vacuum and solution PMF maps, on a $5° \times 5°$ grid using the method described in Section 2.2.1. At each point in ϕ, ψ space, all other degrees of freedom were relaxed by energy minimisation. In addition, for each ϕ, ψ conformation, clockwise and anti-clockwise arrangements of the secondary hydroxyls and three orientations of the primary hydroxyl (*tg*, *gg* or *gt*) were considered. For each analysed structure, the steepest descent method was used for 500 steps of initial minimisation, followed by up to 2000 steps of conjugate gradients minimisation.

5.2.3 Simulation Conditions

Molecular dynamics simulations were performed using the program CHARMM[222] (version 27b1), with modifications incorporated into the USERE module in order to implement the two-dimensional adaptive umbrella sampling PMF calculations. The CSFF parameter set (Chapter 4) for the CHARMM force field was used to model the carbohydrate molecules in all the simulations. The TIP3P model was used to represent water.

For the solution simulations, minimised structures of the disaccharides were each placed in an equilibrated cube of 512 TIP3P water molecules. Solvent water molecules that overlapped with the solute molecules were removed and the system was equilibrated for 500 ps. The solution simulations for β-maltose and β-dixylose employed one disaccharide molecule surrounded by 489 TIP3P water molecules[248] in a cube of length 24.64 Å (maltose) and 24.53 Å (dixylose). This gives a density of 1.013 g/cm^3, which is the approximate experimental density of an aqueous maltose solution of this weight percentage (2.2%). The cube was subjected to minimum image periodic boundary conditions to eliminate edge effects.

Initial velocities for the atoms were selected at random from a Boltzmann distribution at 300 K. All simulations were performed in the canonical ensemble (constant n, V, T), using stochastic Langevin dynamics with a frictional coefficient of 62.5 to maintain a constant temperature of 300 K. The equations of motion were integrated using a Leap-Frog Verlet integrator[256] with a step size of 1 fs. The SHAKE algorithm[258] was used to fix the length of bonds involving hydrogen atoms and the water molecule geometry throughout each simulation. Non-bonded interactions were truncated using a switching function applied on a neutral group basis between 10.0 and 12.0 Å. The groups corresponded to electrically neutral collections of atoms in the carbohydrate molecules and entire water molecules for the solvent.

5.3 Results and Discussion

5.3.1 Solvent Effects: Maltose

Contoured plots of the calculated ϕ, ψ adiabatic map and the PMF energy surfaces for β maltose are shown in the left column of Figure 5.3. As expected, the principal features of the various maltose energy landscapes are similar. All the energy surfaces are characterised by two principal valleys, whose general location is consistent with published adiabatic maps for maltose.[35,57,72,82,83,135,148,280] As the ϕ angle is considerably more restricted than the ψ angle, these valleys can be defined approximately as $\psi = 0^\circ$ (marked

University of Cape Town

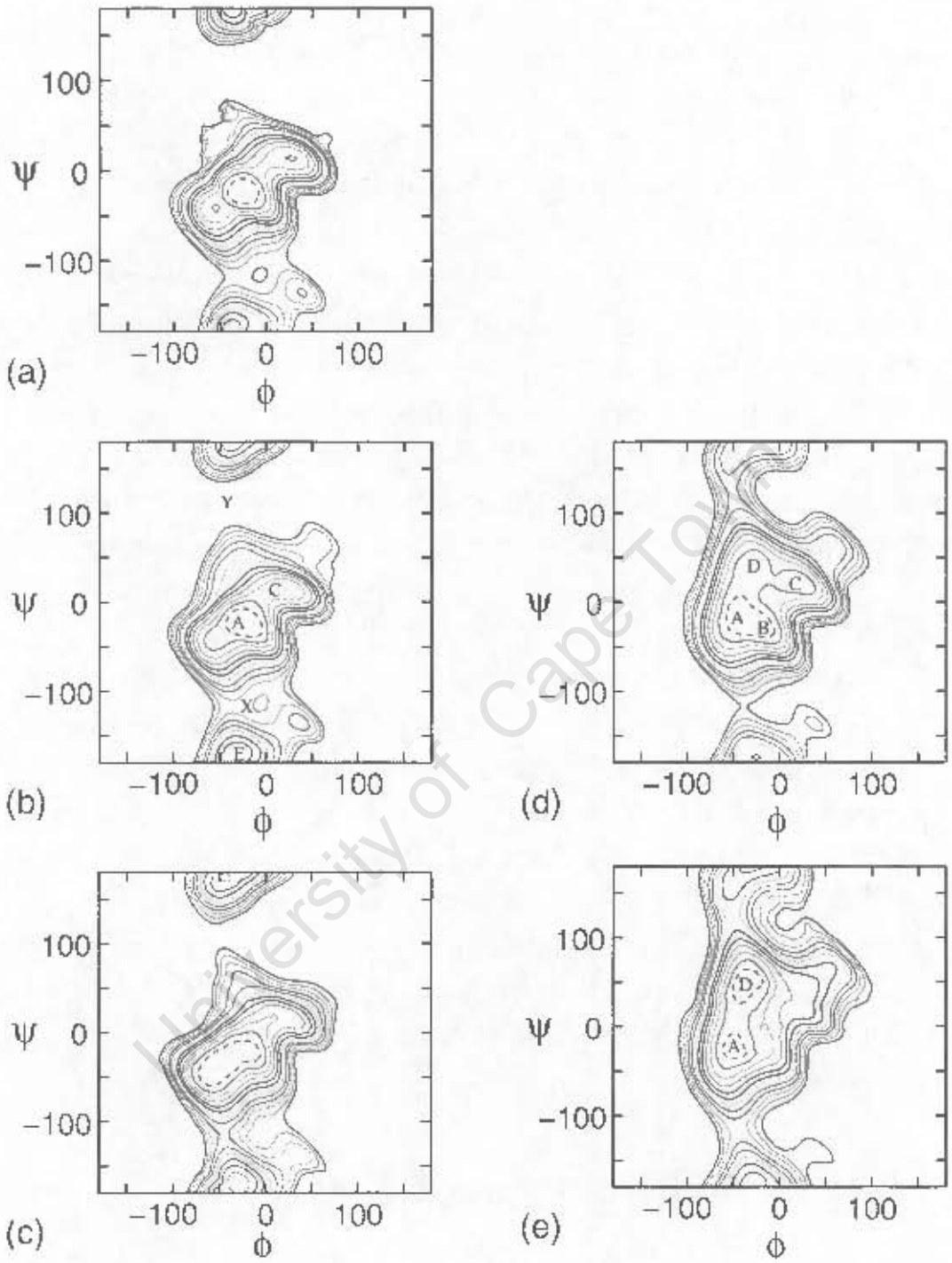


Figure 5.3: Contoured ϕ, ψ energy surfaces for maltose (left column) and dixylose (right column). (a): maltose adiabatic map (b) and (d): vacuum PMF surfaces for maltose and dixylose respectively (c) and (e): PMF surfaces in TIP3P solution for maltose and dixylose respectively. Contours are at 1 kcal/mol intervals above the global energy minimum, to a maximum of 12 kcal/mol. The first contour at 1 kcal/mol is dashed.

University of Cape Town

	Maltose		Dixylose	
	Vacuum	Solution	Vacuum	Solution
Global minimum (ϕ, ψ)	(-22.5,-25.0)	(-47.5,-35.0)	(-37.5,-22.5)	(-50,-25)
ΔG (A→E)	4	3	6	5
ΔG (A→X)	10	8	11	9
ΔG (A→Y)	15	12	8	6
ΔG (E→X)	6	6	5	6
ΔG (E→Y)	11	9	2	2

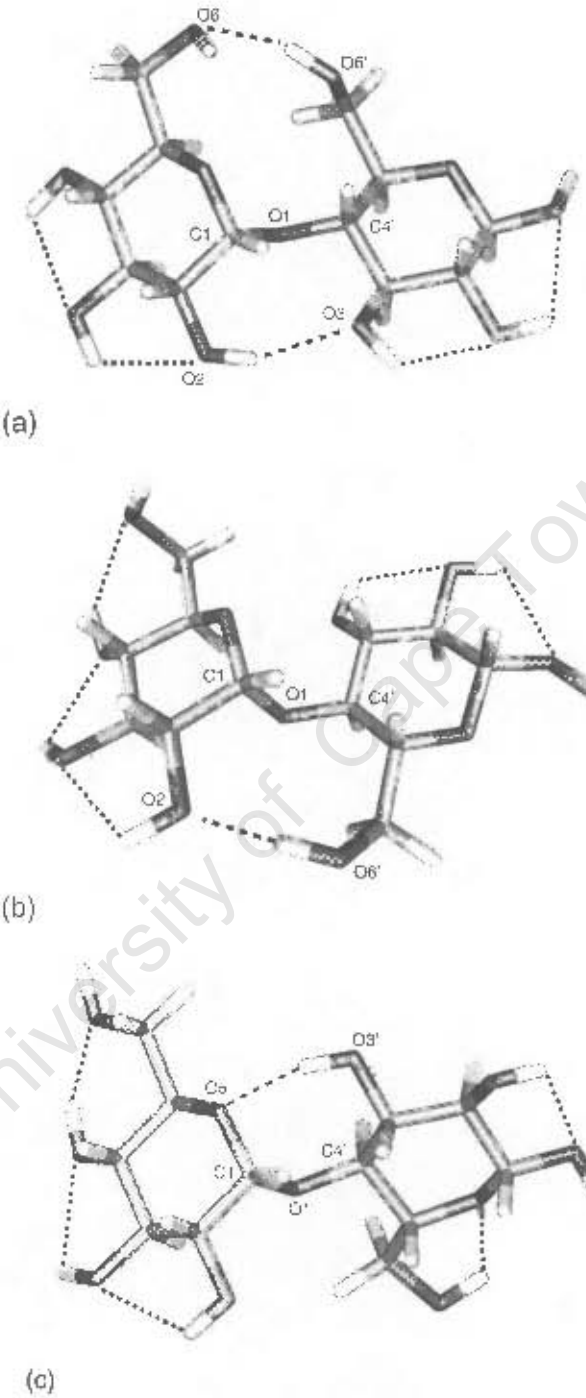
Table 5.1: Location of the global minimum and approximate free energy differences (in kcal/mol) for transitions between the A– and E–minima and the X– and Y–barrier regions on the maltose and dixylose ϕ, ψ potential of mean force surfaces in vacuum and solution.

A on the maltose vacuum PMF) and $\psi = 180°$ (E). The global minimum is in all cases found within the A–valley (see Table 5.1). However, although the adiabatic map global minimum energy conformation at ϕ, ψ =-23.90, -21.79 is close to the vacuum PMF global minimum at ϕ, ψ =-22.5, -25.0, in solution the global minimum shifts to -47.5, -35.0. The maltose crystal structure[19,21,22] is located in the region indicated by C and the lowest energy saddle points of the barriers separating the A– and E–well are labelled X and Y.

The slight differences apparent between the adiabatic map (5.3 (a)) and vacuum PMF surface (5.3 (b)) can be attributed to the following factors. Firstly, the more exhaustive umbrella sampling procedure is likely to locate minimum energy conformations not discovered by the adiabatic map procedure. This is clearly the case in the areas where the vacuum PMF surface is obviously smoother than the adiabatic map. It also occurs in the ϕ, ψ = -70,-70 region of the map, where a stabilising O'_3–O_5 hydrogen bond was not located by the adiabatic map procedure. Secondly, the maps will differ where entropy, in the form of molecular flexibility, makes a contribution to the free energy. This appears to occur in both the A– and E–wells, which, though maintaining the same relative energies, are broadened in the vacuum PMF as compared to the adiabatic map.

In vacuum, disaccharide conformations can be stabilised by both intraresidue hydrogen bonds involving hydroxyl groups on the same pyranose ring as well as interresidue hydrogen bonds between hydroxyls on different rings. Intraresidue hydrogen bonds around a pyranose ring typically require suboptimal angles for the hydrogen bonds of between 100° to 130°. Therefore, interresidue hydrogen bonds are expected to be more energetically favourable as the hydroxyls are relatively free to align in an ideal configuration. In the A– and E–well conformations, both of these types of hydrogen bond occur, examples of

University of Cape Town



University of Cape Town

Figure 5.4: Examples of minimum energy configurations for maltose in vacuum. Stabilising intramolecular hydrogen bonds are shown with dashed lines. (a): A–well conformation (ϕ, ψ = -25 °,-25°). (b): E-well conformation (ϕ, ψ = -35°,-170°). (c): X saddle–point conformation (ϕ, ψ = -35°,-115°).

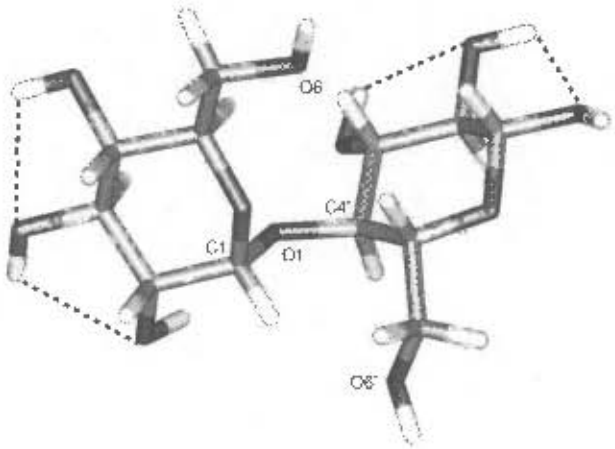


Figure 5.5: Example of a maltose conformation from the Y–barrier region, showing the relatively few numbers of intramolecular hydrogen bonds.

which are shown in Figure 5.4. In the A–region, two interresidue hydrogen bonds can form simultaneously: one between the O_2 and O_3' hydroxyls and the other between O_6 and O_6'. Distance constraints mean that conformations from the E-region can form only one interresidue hydrogen bond at a time, which can be either an O_2–O_6' or an O_1–O_6' interaction. Both of these regions are also entropically favoured, as the hydrogen bonds may alternate as to which are the donor and acceptor hydroxyls and may have a number of energetically similar configurations. In contrast, the C region, though able to form two interresidue hydrogen bonds, is disfavoured by entropic factors, as was found for the partial maltose vacuum PMF calculated by Brady and coworkers.[88]

Judging from the difference in the relative heights of the X and Y barriers ($\approx$ 5 kcal/mol, Table 5.1), transition between the A– and E–minima in vacuum is expected to occur chiefly via the saddle point at X. This assumption is corroborated by the forces required to bring about transition, as calculated from the maximum slopes of these hills in the one-dimensional ψ-only PMF shown in Figure 5.7. Rotation of the ψ dihedral angle from the A– to the E well via the X–barrier requires 0.12 kcal.mol^{-1}.degree^{-1} compared to 0.14 kcal.mol^{-1}.degree^{-1} for Y. The energetic difference between the X– and Y–barriers can be explained as follows. Although both of these routes involve some degree of steric clash between atoms, maltose conformations from the saddle point at X can be stabilised by a single intraresidue hydrogen bond between O_5 and O_3' (Figure 5.4 (c)). However, in conformations corresponding to the Y barrier (Figure 5.5), steric clashes between the non-reducing ring and the bulky hydroxymethyl group on the reducing ring prohibit a similar bond between O_5 and O_6'. In addition, steric clashes also prevent the O_6' hydroxyl from forming any stabilising intramolecular hydrogen bonds, such as an O_6' O_5' interaction.

University of Cape Town

	Maltose		Maltotetraose	Dixylose	
	Vacuum	Solution	Vacuum	Vacuum	Solution
A well region (%)	99.52	98.93	99.77	99.96	99.96
E–well region (%)	0.48	1.06	0.23	0.04	0.04

Table 5.2: Comparison of the relative populations at 300K of the A–well ($\psi \approx 0^{\circ}$) and E–well ($\psi \approx 180^{\circ}$) valley regions calculated from the maltose, maltotetraose and dixylose PMFs.

Thus, steric clashes combined with the absence of intermolecular hydrogen bonds in the Y-conformations account for the height of this energy barrier in vacuum.

The influence of water on the preferred conformations of the $\alpha(1\rightarrow4)$–linkage is apparent in a comparison of the vacuum and solution PMFs for maltose, shown in Figure 5.3 (b) and (c) respectively. Aqueous solvation affects both the relative energies of the two minimum energy wells and barrier heights between them, as listed in Table 5.1. The energy difference between the A and E wells is lowered by approximately 1 kcal/mol (see Table 5.1). This has the result that the population of *anti* (ψ=180°) conformers doubles in solution as compared to vacuum (Figure 5.6 (b) and Table 5.2). Moreover, as the heights of the X– and Y–barriers are lowered, transition between the A– and E–wells becomes more probable in aqueous solution. A further effect of the water solvent is to broaden and extend the A–well toward more negative ϕ, ψ values, away from the C conformation. This change is demonstrated by the ψ population histograms for the A region at 300K (Figure 5.6 (a)): the solution histogram is broader with a median at ψ=-35, in contrast to the solution value of ψ=-25. This is consistent with various experimental[31,42,44] and molecular dynamics simulation studies,[68,72,159] which found similar shifts in the maltose ϕ, ψ populations upon solvation.

The ψ-only PMFs shown in Figure 5.7 give an impression of the effect of solution on the maltose conformations. However, the influence of solution is perhaps clearer in the one dimensional representation of the ϕ, ψ PMFS given by the τ PMFs (Figure 5.9). Figure 5.8 illustrates how the τ PMF relate to the ϕ, ψ PMFs. Unlike ψ, each value of the τ dihedral angle represents a particular total conformation of the glycosidic linkage. Therefore, this torsion angle is more useful for extrapolation to oligosaccharide conformations. In addition, as each τ angle represents a particular orientation of the glycosidic linkage, where multiple minima are not present this graph provides a good estimate of the minimum energy path for rotation. In the case of maltose, the τ plots are particularly useful because the energy surface is almost one dimensional: the ϕ angle is so restricted that a particular τ value corresponds to only one ϕ, ψ minimum region.

University of Cape Town

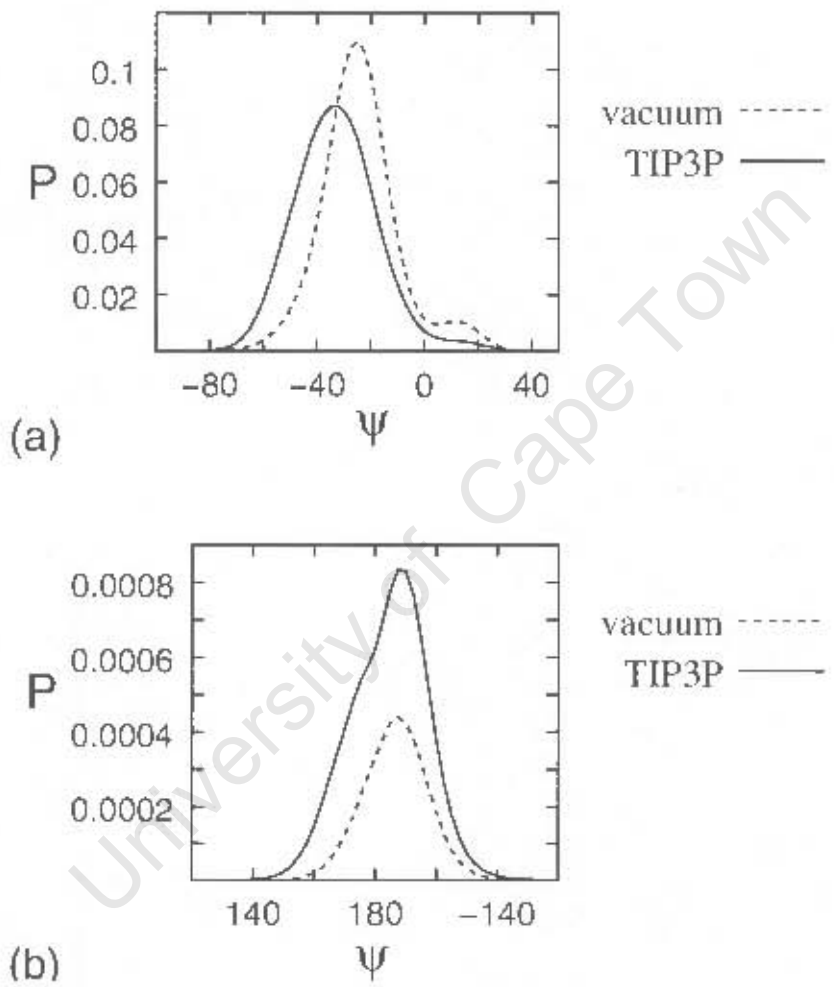


Figure 5.6: Normalised equilibrium populations of the maltose ψ dihedral in vacuum and solution at 300K in (a): the A–well and (b): the E–well.

University of Cape Town

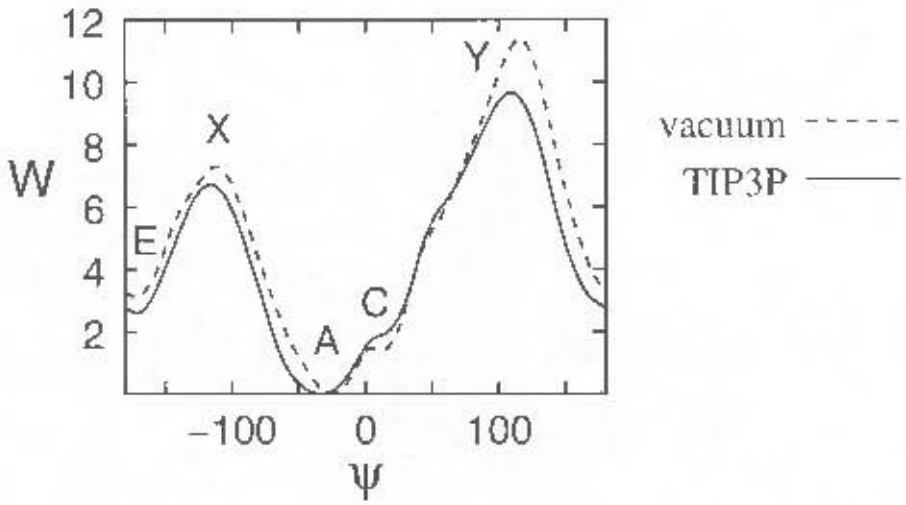


Figure 5.7: One-dimensional ψ PMFs calculated for maltose in vacuum and TIP3P solution. The curves are shown relative to their global energy minima, with the energy in kcal/mol. Regions corresponding to the A– and E–wells and the X– and Y–barriers on the ϕ, ψ map are indicated.

The only real exception to this are the X– and Y–barriers, both of which in the τ graph incorporate lower energy regions e.g. the 60° to 90° τ range encompasses the top of the C–region as well as the Y–barrier. Thus, these barriers in the τ graphs appear to be lower than they actually are. The τ PMFs show a lowering of almost the entire τ region, excepting the upper A– and C–regions. The A–region corresponds to left–handed helical conformations, whereas the C–region results in right–handed helices. The A–well is dramatically broadened toward more negative τ values in solution, with the minimum shifting from around $\tau = -50$ to $\tau = -90$. In an $\alpha(1\text{-}\!>\!4)$-linked oligosaccharide, this corresponds to a change from seven glycosidic linkages required for a full 360° helical turn to only four linkages.

The effect of aqueous solution can be explained by the fact that the water molecules provide alternative hydrogen–bonding partners for the sugar hydroxyl groups. A multitude of hydrogen bonding arrangements between water molecules and a sugar solute are possible. For example, water molecules can form "bridging" hydrogen bonds between two sugar hydroxyls, as shown in Figure 5.11. However, more complicated stabilising networks of hydrogen bonds involving chains of interactions between two to five molecules were found to be more common. The presence of water thus reduces the enthalpic advantage of the vacuum intramolecular and, particularly, interresidue hydrogen bonds. Favourable interactions between the solution and solute molecules have the effect of lowering the relative energy of maltose conformations that are unable to form interresidue

University of Cape Town

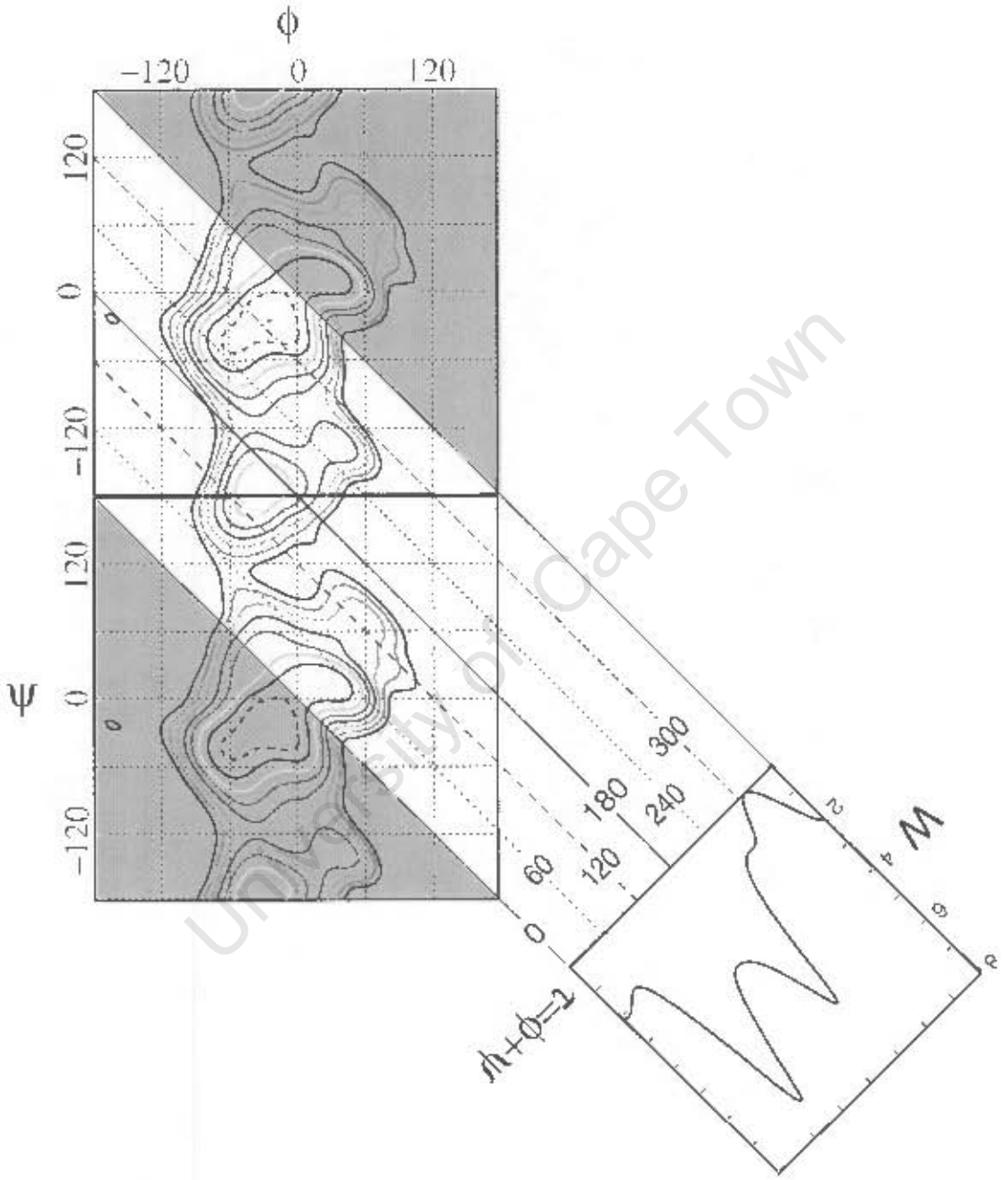


Figure 5.8: Diagram of the relation between the maltose 2D ϕ, ψ PMF and the 1D τ PMFs.

University of Cape Town

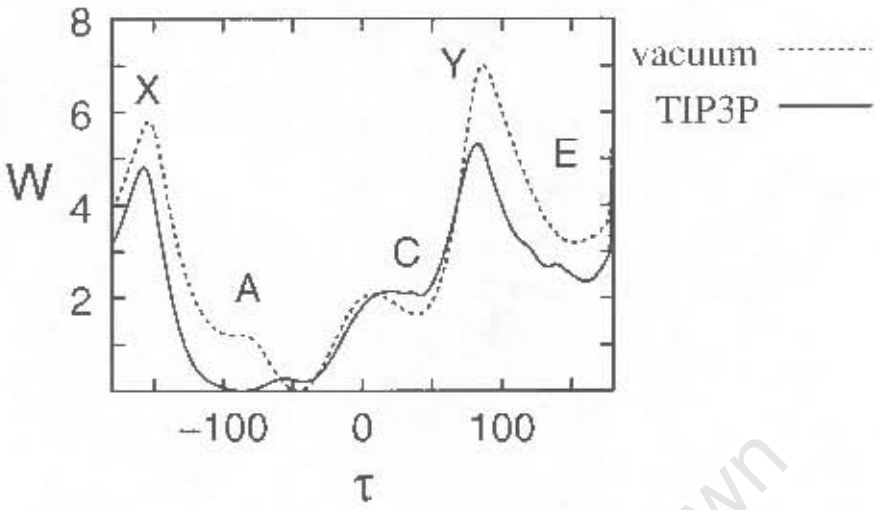


Figure 5.9: PMF surfaces for τ dihedral in maltose obtained from the ϕ, ψ populations. The graphs are shown relative to their global energy minima, with the energy in kcal/mol. Regions corresponding to the A and E wells and the X- and Y-barriers on the ϕ, ψ map are labelled.

University of Cape Town

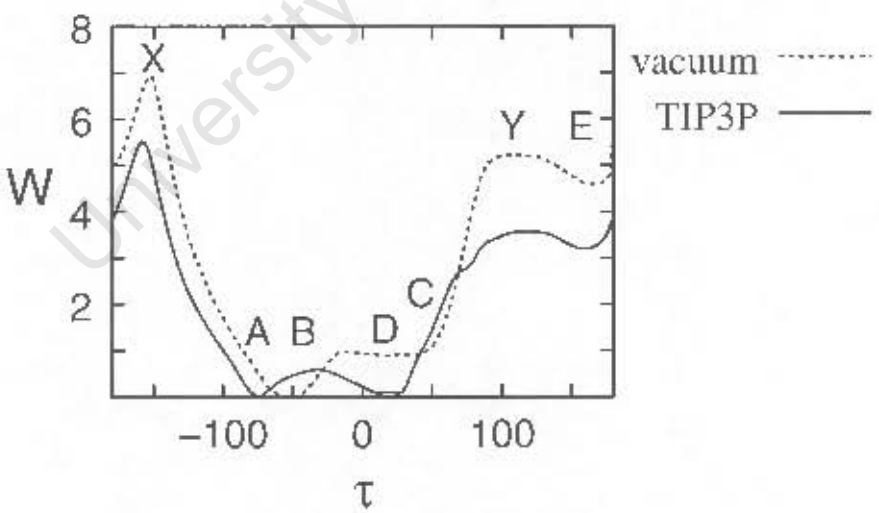


Figure 5.10: PMF surfaces for τ dihedral in dixylose obtained from the ϕ, ψ populations. The graphs are shown relative to their global energy minima, with the energy in kcal/mol. Regions corresponding to the A, B, C, D and E wells and the X- and Y-barriers on the ϕ, ψ map are labelled.

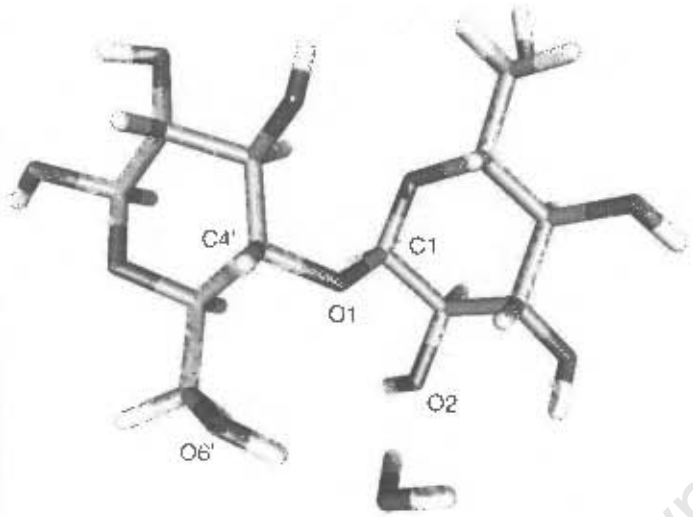


Figure 5.11: Snapshot from an MD simulation, showing maltose in an E–well conformation ($\phi = \quad 43°, \psi = -162°$), with a water molecule bridging between the O_6' hydroxyl (in the *gg* conformation) and the O_2 hydroxyl.

hydrogen bonds. As the A– and C–well configurations are stabilised with two intraresidue hydrogen bonds compared to one in both the E- well and X saddle point and none for the Y–saddle point, the addition of water as a solvent results in a change in the relative energies of these regions. This has the result that the A–well broadens toward conformations that have a greater separation of the primary alcohols. In addition the A–well becomes shallower in solution relative to the E- well as well as the X– and Y–saddle points (Table 5.1). This reasoning also explains why the energy required for a E→X transition is unaffected by solvation, whereas the E→Y barrier lowered. As discussed earlier, the E– and X–regions have approximately the same potential to form internal hydrogen bonds and are therefore affected similarly by solution. However, the Y-conformation can form relatively few intermolecular hydrogen-bonding interactions and thus gains the greatest energetic advantage in solution. This translates to an increase in energy of 2 kcal/mol for the A–well and 1 kcal/mol for the E– and X–regions relative to the Y–barrier.

The increase in the flexibility and range of motion for the α(1→4) linkage in maltose provides an explanation for the effect of water on the glass transition temperature: T_g lowering of maltose glasses in the presence of water can be explained not only by disruption of the hydrogen bond network between maltose molecules,[37] but also by the increased mobility of maltose. In addition, the flattening of the maltose PMF in solution and the change in the relative energy of the minima has implications for the structure and dynamics of α(1→4) linked oligo- and polysaccharides: by increasing the range of motion

University of Cape Town

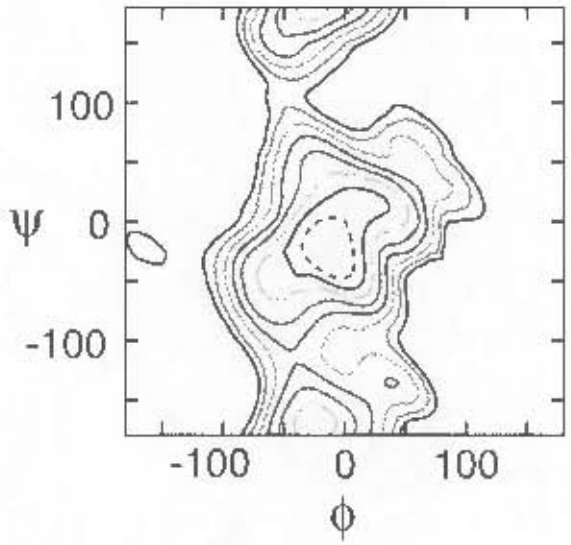


Figure 5.12: Contoured ϕ, ψ free energy surface for maltotetraose in vacuum. Contours appear at 1 kcal/mol intervals above the global energy minimum. The first contour at 1 kcal/mol is dashed. Contours higher than 12 kcal/mol are not shown.

University of Cape Town

of the glycosidic linkage, solvation should result in a more flexible oligosaccharide chain, with a wider range of low energy conformations.

5.3.2 Chain Length Effects: Maltotetraose

The maltotetraose vacuum PMF (Figure 5.12) gives an indication of the effect of contiguous glucose monomers on the $\alpha(1\rightarrow4)$–linkage. The principal change compared to the maltose PMF is that the A–well has narrowed slightly. A–well conformations in maltotetraose correspond to a curved section of a helix, as shown in Figure 5.1 (c). The narrowing of the A well is brought about by the series of intraresidue O_2 O_3' hydrogen bonds that occur in this conformation, which amplify the strength of the interaction. Hydrogen bonds of this sort are known to occur in crystal structures of cyclodextrins. However, despite this, the relative energies of the A– and E–wells are unaffected by the additional monomers in maltotetraose. As the maltose and maltotetraose PMFS are so similar, one can conclude that there are no dominant interactions affecting the $\alpha(1\rightarrow4)$ linkage in maltotetraose that are not present in maltose. Therefore, it seems that as far as maltotetraose is concerned, the maltose dimer is an adequate model of the $\alpha(1\rightarrow4)$–linkage and that the left handed helical structure proposed for amylose is correct.[16,151,152] The implications of these PMFs for the conformation and dynamics of amylose are explored in more detail in Chapter 6.

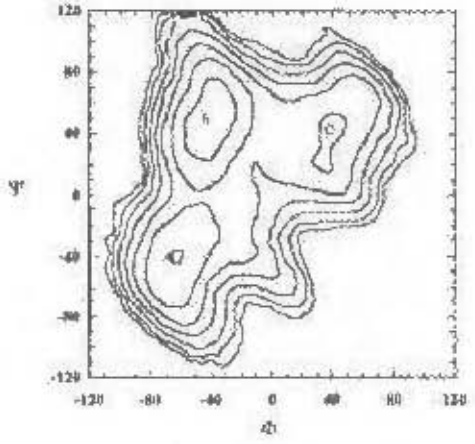


Figure 5.13: The previously calculated partial ϕ, ψ PMF for dixylose in solution, reproduced from Naidoo and Brady.[89] Energy minima are shown at 1 kcal/mol intervals above the global minimum at ϕ, ψ ~-37°, 52°.

5.3.3 Effect of the Primary Alcohol: Dixylose versus Maltose

Comparison of the vacuum PMFs for maltose and dixylose (Figures 5.3 (b) and (d) respectively) clarifies the contribution of the hydroxymethyl to the configurational free energy of the maltose disaccharide. Removal of the primary alcohol lowers the energy of the C well and creates a new minimum region marked D. In addition, the energy of the E-minimum is raised. The effect is most apparent in a comparison of the τ vacuum PMFs for dixylose (Figure 5.10) and maltose (Figure 5.9). The A-well broadens into the D region (ϕ, $\psi \approx$ -40,40) in dixylose because of the absence of the steric repulsions between hydroxymethyl groups that increase the energy of this area in maltose. This also explains the slight lowering in energy of the C–well ($\phi, \psi \approx 30, 30$). The general A–B–C–D region in dixylose predominates in vacuum because of a stabilising interresidue O_2–O_3' hydrogen bond, which is particularly favourable in the A well global minimum region. Dixylose lacks the potential for interresidue stabilising bonds when it is in the *anti* (ψ=180°) conformation, with the result that in vacuum the energy of the E–well is raised with respect to the A–well. However, in comparison to maltose, removal of the hydroxymethyls lowers the Y–barrier in vacuum by 7 kcal/mol (Table 5.1). This considerable change comes about because the steric clash that occurs between the hydroxymethyl groups on this pathway in maltose is absent in dixylose. In contrast, the X route is slightly raised in energy. Therefore, in dixylose, transitions between the A– and E–wells are likely to occur via the Y–barrier — the opposite situation to maltose.

The solution PMF for β–dixylose (Figure 5.3 (e)) is similar to the partial PMF previously calculated by Naidoo and Brady (shown in Figure 5.13).[89] However, the global

University of Cape Town

minimum appears within the B well in their PMF map and the C–well is rather more favoured in solution than is the case here. The effects of solution on the xylose conformations predominating in the minimum energy wells have been discussed in detail in the Naidoo and Brady dixylose study. Although some of their work is repeated here, the focus is specifically on a comparison of maltose and dixylose.

Aqueous solution lowers the energy of the D–region and raises the energy of the B– and C–regions for dixylose, as can be seen in a comparison of the solution and vacuum τ PMFs (Figure 5.10). The change in the glycosidic linkage energy surface is principally a result of the hydrophobic effect. Removal of the hydroxymethyl group and its replacement with a hydrogen atom results in a hydrophobic site where previously it was hydrophilic. Conformations which result in close proximity of the two pairs of hydrophobic H_5, H_5' atoms are favoured over those which separate them. Therefore, the C– and B–regions, which both separate the H_5 and H_5' hydrogens more than the D– and A–regions, are raised in energy in solution. As part of the same effect, the global energy minimum moves from the A–well to the D–well in solution. The 1 kcal/mol energy drop in the D–well is due to the fact that this conformation brings the hydrophobic hydrogens into closest proximity, while maintaining the O_2-O_3' hydrogen bond. The ϕ rotation in moving from A to B or D to C moves these hydrogens further apart, while the ψ rotation in going from A to D brings them closer together. As in maltose, the energy of the E–well is lowered with respect to the A–well in solution. This is due once more to favourable hydrogen bonding interactions with the solvent. The X barrier is raised for the same reason. In addition, the Y barrier is lowered further because of a favourable hydrophobic arrangement of the H_5' hydrogens with H_1.

Comparison of the τ PMFs of maltose and dixylose (Figures 5.9 and 5.10) also gives a general impression of the differences between these molecules. The E–region is obviously considerably less favoured in dixylose than in maltose. In addition, the A-region is affected quite differently by solvation in these two molecules. Conformations taken exclusively from the C or D regions would result in a dixylose polymer forming a broad, right–handed helix, while the A– and B–regions correspond to a left–handed helix. In maltose, it is clear that the A-conformations dominate in both vacuum and solution. However, as these conformations are close in energy for dixylose (particularly in solution), a mixture of these angles is likely to occur in a dixylose polysaccharide. Therefore, a dixylose polymer seems to be less likely to form a regular, identifiable secondary structure than is amylose.

University of Cape Town

5.4 Conclusions

Adaptive umbrella sampling used together with the Weighted Histogram Analysis Method has enabled calculation of the first vacuum and solution PMFs for maltose and dixylose that cover all of the low energy regions of ϕ, ψ phase space. These PMFs allowed for a quantitative analysis of the effects of water, the hydroxymethyl groups and additional pyranose rings on the α(1→4)-linkage. In addition, the one–dimensional approximation of the glycosidic linkages provided by the τ PMFs was found to be useful for providing an overall picture of a glycosidic linkage conformation.

In maltose, aqueous solvation broadens the range of low–energy structures, lowers the transition barriers between the two minima and thus flattens the PMF surface. In addition, the relative energies of the *syn* and *anti* minima are altered, so that *anti* conformations of the α(1→4)–linkage are likely to be more prevalent in solution than previously thought. This has implications for the physical properties and dynamics of the amylose polysaccharide, as the E well is narrower and more restricted than the A–well. From the broadening of the A–well, one can predict that the amylose polymer forms extended, left handed helices in solution, interrupted by occasional *anti* conformations.

Comparison of the maltotetraose PMF in vacuum with the maltose PMF leads to the conclusion that the extra glucose units in this molecule does not affect the PMF surface for the α(1→4) linkage dramatically. This suggests that maltose is therefore a valid model for the α(1→4)–linkage in oligo– and polysaccharides.

The favoured conformations of dixylose in solution are principally determined by hydrophobic interactions, in contrast to the hydrophilic interactions that predominate in maltose. When the hydroxymethyl is not present, the central minimum region resolves into two narrow energy wells in solution and the E-well is less populated. The fact that dixylose in solution shows two minima of almost equal energy in the central region (where maltose has only one broad minimum) implies that polymers comprising repeat units of dixylose should have a less regular conformation in solution than malto–oligomers.

University of Cape Town

University of Cape Town

Chapter 6

The $\alpha(1\rightarrow4)$ Glycosidic Linkage II: Oligosaccharide Properties

University of Cape Town

6.1 Introduction

The two-dimensional ϕ, ψ potential of mean force surfaces for the $\alpha(1\rightarrow4)$ linkages discussed in Chapter 5 have implications for the preferred conformations and dynamics of $\alpha(1\rightarrow4)$–linked polysaccharides. In this chapter, investigation of the $\alpha(1\rightarrow4)$–linkage is extended to its effect on the conformation and dynamics of oligosaccharide strands. The backbone glycosidic linkages comprise the principal degrees of freedom in a polysaccharide. If these linkages function independently of each other, ϕ, ψ PMFs calculated for the corresponding disaccharide(s) are sufficient to establish the average conformations and dynamics for oligo– and polysaccharide chains. Even if successive linkages are not completely independent (as is likely to be the case), the dimer PMFs should nevertheless still provide valuable insight into the structural and dynamic properties of the polysaccharides.

Amylose, one of the two polymeric components of starch, consists exclusively of $\alpha(1\rightarrow4)$–linkages. As discussed in Chapter 1, many problems have been encountered with the experimental determination of the three–dimensional structure of polysaccharides in both solution and the solid state. Amylose is no exception: despite being one of the first carbohydrates to be investigated, its conformation in different environments has not yet been satisfactorily determined.

Biologically, starch forms compact, microscopic granules that vary in their shape, size, composition and mechanical properties according to their origin. Starch granules are composed of a mixture of amylose and its $\alpha(1\rightarrow6)$ branched partner, amylopectin. Cross-sections of the granules (Figure 6.1) show concentric rings of alternating crystalline

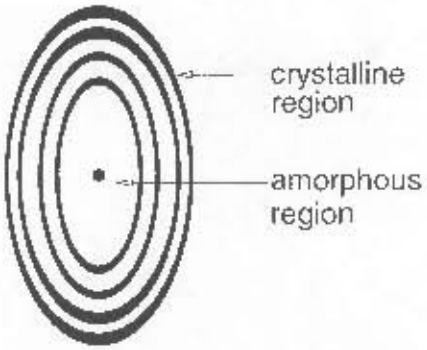


Figure 6.1: A diagram of the structure of a starch granule.

lamella (mostly amylopectin) and amorphous regions (containing the branch points of amylopectin plus amorphous amylose).[281]

Pure amylose adopts various helical forms in the solid state. These forms, designated A, B and V, have periodicities of 5–7 residues. V–amylose is the only fully characterised crystalline form.[15,17] It is a single stranded, left–handed helix with 6–glucose units per turn and a helical pitch of 7.91 to 8.17 Å. This helix has a central channel that is able to include molecules such as "iodine blue" used in the well–known test for starch. In the hydrated form of V amylose (V_A), disordered water molecules occupy the central cavity. Other helical forms of amylose are not well characterised and their helical parameters differ, with helical repeats of 6 to 8 units being the most common.[282] Within starch granules, amylose chains are folded into the double–helical A (mostly in cereals) and B (mostly in tubers) conformations. These two forms are similar: they have an increased axial spacing compared to V, with no central channel and they differ principally in their packing arrangement and water content. In the A–type, the double helices are closely packed, probably without water molecules, whereas in the B–type the packing is more open and water molecules fill the spaces.[152] X–ray diffraction studies indicate that these forms both have 6×2 glucose residues per turn and are either left–handed[36,151,152,283] or right–handed[284] double helices with intertwined chains. To further complicate the issue, a single–crystal X–ray study of an amylose–type oligosaccharide, the complex (*p*–nitrophenyl α–maltohexaoside)$_2 \cdot Ba(I_3)_2 \cdot 27H_2O$, revealed an anti–parallel left handed double helix that has no resemblance to A–, B– or V–amylose.[151,152]

A variety of theoretical and experimental studies have suggested two possible structures for amylose in a neutral aqueous solution: either a random coil[153,154,158,193] or an "interrupted helix".[285] Though similar macroscopically, these two models differ in that an interrupted helix has short range order that is absent in the random coil. The suggested structure of the interrupted helix comprises a combination of loose, helical sections, each

University of Cape Town

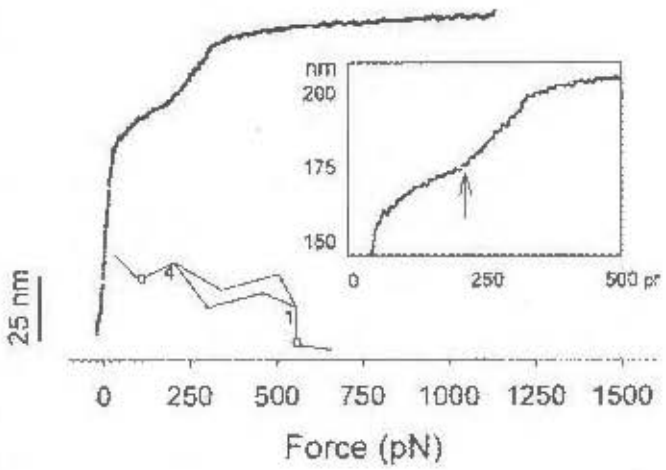


Figure 6.2: The force extension curve for carboxymethylamylose, reproduced from Marszalek et al.[53] The inset shows the α–D–glucopyranose ring and the transition region at high resolution, with the arrow indicating the start of the deformation region.

containing about 100 glucose units, which alternate with short, flexible, random coil segments.[285] In contrast, amylose in dimethylsulfoxide (DMSO) solution seems to have a more regularly helical conformation.[158,285] The conformational change from a tight helix to a random coil proposed to occur when amylose is moved from DMSO to an aqueous solution has been assumed to reflect changes in the torsion angles of the glycosidic linkages.[32,158]

The characteristic motions of an amylose chain determine the relaxation behaviour of this molecule and hence the response of an amylose–containing material to external stresses. In the solid state, gradual structural relaxations are thought to cause the physical ageing of amylose, which results in brittleness.[9–11] The ageing process can be reversed by heating starch in the presence of water. Water molecules are thought to facilitate dynamic transitions, a process which would also explain the effect it has of plasticising amyloses glasses and hence lowering the glass transition temperature.[8,11–14] The pronounced effect of water on the physical properties of starch–based materials hinders the otherwise feasible use of high–amylose starch as a cheap and biodegradable plastic.[6–9]

The dynamic behaviour of amylose and α(1→4) linked oligomers are not well characterised experimentally and the molecular origin of the macroscopic physical properties of amylose glasses remains to be elucidated. Indirect evidence for the range of possible dynamic transitions for amylose strands in solution is given by the force–extension relationship explored by Atomic Force Microscopy stretching experiments.[52,53,121] The force–extension curve for a single molecule of carboxymethylamylose recently reported by Marszalek et al. is shown in Figure 6.2.[53] This curve has a characteristic shape, with

University of Cape Town

four identifiable regions. These regions roughly correspond to the number of different structural processes occurring. In first stage, at forces below ≈ 50 pN, the curve rises steeply. The second stage, between 50 and 200 pN, shows a flatter slope and hence lower elasticity, which ends with an abrupt increase in slope at around 200 pN. This is shown by the arrow on the inset in Figure 6.2. The slope flattens again at 350 pN, where the final stage of low elasticity begins. This curve has been interpreted as follows. The first stage is associated with rapid elongation of the carboxymethylamylose polymer. This elongation slows in the second stage as the extension of the molecule approaches its contour length. Then, a sudden onset of concerted chair–to–boat transitions in the pyranose rings results in the increase in elasticity at the third stage. This continues until all the rings have converted to the boat conformation. The final stage, with its greatly reduced elasticity, involves deformations of the rings and bending of the glycosidic linkages.

The chair to boat conformational changes of the pyranose ring on stretching were a surprising idea at first, as boat conformations are not typically seen in crystal structures. The transition has been postulated to occur as a result of a torque produced by the axial glycosidic bonds on stretching, as the number of transitions in the force–extension curves of pyranose polymers with 1 4 linkages is correlated with the number of glycosidic bonds in the axial conformation.[122] As discussed in Section 1.1.1, calculations have shown that, when O_1 is in the α–configuration, the O_1–O_4 distance is longer in the boat (≈ 5.6 Å) than in the chair conformation(≈ 4.6 Å).[119] A chair–to–boat transition is therefore a plausible response to stretching of the pyranose ring. Indeed, MD stretching simulations of 16–unit oligomers have shown chair–to–boat transitions to occur in the α–linked amylose but not in the β–linked cellulose.[120] However, these simulations required the ψ angles to be constrained in the *anti*–parallel conformation in order to produce a shoulder in the stretching curve similar to that seen in experiment. Constraining the ψ dihedral angles was justified by an assumption that previously unaccounted–for bulky or charged side groups caused steric hindrance and thus favoured the *anti* conformation.[120] Further corroboration for the chair–to–boat mechanism was provided by experiments that showed the characteristic transitions in amylose and pectin to be absent after oxidative cleavage of the rings with periodate.[121]

This chapter aims to establish the preferred conformations and characteristic dynamics of $\alpha(1\rightarrow4)$–linked glucans. To some extent, the preferred conformation of amylose in solution and vacuum can be determined from the maltose ϕ, ψ PMFs presented in the previous chapter. Therefore, as a first step, the average conformation of an amylose polymer strand in vacuum and solution is estimated from the ϕ, ψ PMFs. The influence of the $\alpha(1\rightarrow4)$ glycosidic linkage on the average conformation and dynamic behaviour of

University of Cape Town

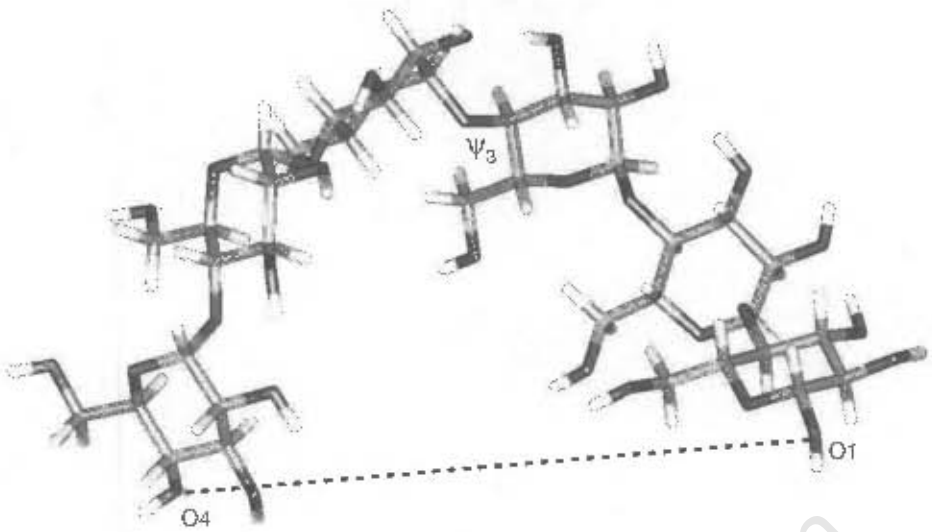


Figure 6.3: Three–dimensional structure of the maltohexaose oligosaccharide, showing the glycosidic linkage dihedral angles and the O_1O_4 end–to–end distance (r). The glycosidic linkages are numbered from the nonreducing and to the reducing and of the chain.

University of Cape Town

α(1→4)–linked oligosaccharides is then further explored by calculation of the end to end distance (r) PMF for a 6 unit amylose oligomer (maltohexaose) in vacuum. Current computational restrictions precluded the calculation of an end to end PMF in solution. The PMF calculation compresses and elongates the oligomer chain, with the result that all configurations corresponding to the various chain lengths are explored. Although six glucose units correspond to a relatively short strand, maltohexaose molecules have been shown to reproduce the relaxation behaviour of the amylose polysaccharide.[56] Finally, the dynamic processes occurring in amylose under strain are investigated using simulated force–ramp AFM experiments on an 18–unit amylose strand. An 18–unit strand corresponds to three helical turns in amylose. These studies should help to establish the validity of dimer model for the glycosidic linkage conformations in a carbohydrate chain.

6.2 Methods

6.2.1 Chain Conformations

The minimum–energy helical conformations predicted by the maltose PMF for an α(1→4)–linked oligosaccharide in solution and vacuum were investigated by constructing two maltododecaose molecules with ϕ, ψ values corresponding to the global minimum energy A–well conformations in vacuum and solution respectively. The structures were minimised

with the glycosidic linkages constrained to the assigned values, using 500 cycles of "steepest descents" minimisation followed by 2000 cycles of "conjugate-gradients" minimisation. In addition, the average chain conformation for a 300 unit amylose polysaccharide was estimated by random assignment of the values of ϕ, ψ angle pairs in accordance with the population distribution calculated from the maltose solution PMF.

6.2.2 End–to–End PMF Calculations

A one dimensional end to end PMF for a 6-unit β-maltohexaose molecule was computed as described in Section 3.5. The end–to–end distance for this molecule was defined as that from the O_1 on the reducing end of the strand to the O_4 on the non–reducing end, as shown in Figure 6.3.

In each case, the one-dimensional umbrella potential surface for the end-to-end distance was represented a grid of points, with a grid separation of 0.25 Å. At each integration step of a simulation, the biasing potential energy to be applied to the ends of the molecule was calculated from a spline of the umbrella energy surface. After each simulation, a two-dimensional distribution histogram was calculated for the end-to-end distance by summing the number of configurations in each 0.25 Å bin over the production phase of the trajectory. The WHAM procedure was then applied, using a tolerance value of 0.001, to obtain an optimum combination of all the simulation distributions and thus the next estimate of the PMF. This PMF surface was extrapolated into unexplored regions by setting unsampled bins to the maximum value of the sampled bins and then smoothed three times using Equation 3.9. Subsequently, the PMF was truncated by setting all regions with an energy value greater than a predetermined cut–off of 25 kcal/mol to the cutoff value. This was done to avoid applying so much force that the pyranose ring conformations were altered. The umbrella potential for the next simulation was set to the negative of this PMF estimate.

The adaptive umbrella sampling procedure was started with a flat PMF and a conformation with all glycosidic linkages set to ϕ, ψ =-22,-25. In each simulation, an equilibration period of 500 ps was followed by a data collection period of 2 ns to 20 ns. After every 8 steps, the next simulation was begun from the original starting structure. Convergence of the adaptive umbrella sampling procedure was defined when a single simulation produced a histogram where the ratio of most occupied to the least occupied bin was less than 5.

During the course of the PMF calculation, the central ψ_3 dihedral angle intermittently rotated into a 180° conformation. This had a marked affect on the dynamic behaviour of

University of Cape Town

the strand and hence adversely affected the convergence of the PMF calculation. Therefore, it was decided to calculate two additional end–to–end distance PMFs for maltohexaose in order to separate the dynamics of the helical conformations (ψ_3 *syn*) of the strand from the ribbon conformations (ψ_3 *anti*). This was done by apply one of two alternative umbrella potentials, depending on whether the ψ_3 angle was in the *syn* or *anti* conformation. These two umbrella potentials then converged separately on two different PMFs. The calculated PMFs are actually a two–dimensional map of end–to–end distance versus ψ, but are very coarse–grained in the second (ψ) dimension. For the sake of simplicity, only the central dihedral was chosen to discriminate between ribbon and helical structures, as this has the greatest effect on conformation. The conformations of the other glycosidic linkages were ignored. Thus, "helical" conformations encompass not only pure helices, but also those with *anti* ψ dihedral angle orientations at the ends of the stand.

6.2.3 Oligosaccharide Stretching Simulation

An 18–unit amylose oligomer was used for the simulated stretching experiments, with the reducing residue in the β-configuration. A strand of this length corresponds to three helical turns of V–amylose. Shorter strands were found to generate population data that was too noisy for the calculation of smooth force-extension curves. The simulation was begun with an initial minimised helical structure with all the glycosidic linkages set to the maltose solution PMF values of ϕ, ψ =40, 50. This corresponds to the edge of global energy minimum valley for maltose (Chapter 5) and produces a fairly extended helix.

The 18–unit amylose oligosaccharide strand was stretched by applying equal, but opposite, forces to the terminal O_1 and O_4 atoms at the reducing and non-reducing ends of the chain respectively, using the method outlined in Chapter 3. The simulation was performed in vacuum, as vacuum calculations can be extended to far longer times than is feasible with a solution calculation. Rupture strengths for weak biochemical interactions are not constants, but are dependent on both the rate and duration of force application.[286] Indeed, it was found for preliminary simulations that the faster the pulling rate an amylose fragment was subjected to, the larger the forces required to reach a particular molecular extension. Moverover, if the force was ramped too quickly, the distinctive features of the experimental force extension curve for amylose were not reproduced. Therefore, the final stretching protocol used ramped the force by 27.79 pN (0.4 kcal.mol^{-1}.Å^{-1}) every 15 ns to 20 ns during the simulation, up to a maximum force of 2225 pN. This corresponds to a ramping rate of 1.85 to 1.39 pN/ns, which, although comparatively long for computer simulations, is still six orders of magnitude faster than the experimental ramp rate of 2.5

University of Cape Town

pN/ms.[53]

The simulation was begun with a force of 55.58 pN, to prevent the molecule from folding in on itself in the vacuum. Each simulation step consisted of 1ns of equilibration, followed by 14 to 19 ns of data collection. A total of 81 ramping steps were applied over the course of the simulation, which equates to over 1200 ns of simulation time. A force–extension relationship was obtained by plotting the average extension during each production step against the applied force. This curve was normalised by dividing by the maximum extension of the oligomer achieved by the end of the simulation: 99 Å. The curves were subsequently adjusted by subtracting the initial force of 55.58 pN. Using the WHAM procedure, an end–to–end PMF was calculated for the first 30 stretching stages from the time–series histograms generated at the production phase of each of the steps. A force–extension relationship was then calculated from the derivative of this PMF for comparison with the force average extension graph.

6.2.4 Simulation Conditions

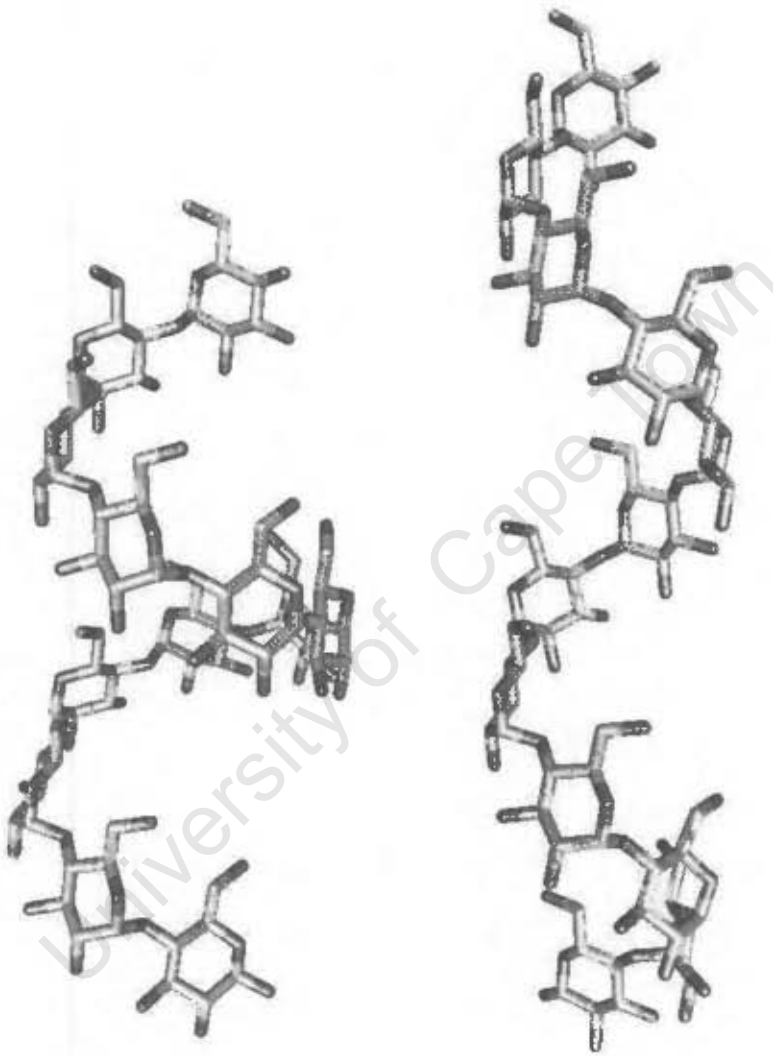
Molecular dynamics simulations were performed using the program CHARMM[222] (version 27b1), with modifications incorporated into the USERE module in order to implement the one dimensional end to end PMF calculations as well as the stretching simulation. The CSFF parameter set (Chapter 4) for the CHARMM force field was used to model the carbohydrate molecules. All simulations were performed in vacuum under the canonical ensemble (constant n, V, T) and using stochastic Langevin dynamics with a frictional coefficient of 62.5 to maintain a constant temperature of 300K. Initial velocities for the atoms were selected at random from a Boltzmann distribution at 300 K. The equations of motion were integrated using a Leap–Frog Verlet integrator[256] with a step size of 1 fs. The SHAKE algorithm[258] was used to fix the length of bonds involving hydrogen atoms. The non–bonded interactions were truncated using a switching function applied on a neutral group basis between 10.0 and 12.0 Å. The groups corresponded to electrically neutral collections of atoms in the carbohydrate molecules.

6.3 Results and Discussion

6.3.1 Oligosaccharide Chain Conformation

In Chapter 5 ϕ, ψ PMF surfaces for the $\alpha(1\rightarrow4)$ glycosidic linkage in maltose (Figure 5.3) were shown to have two principal energy valleys: an A–well at $\psi \approx$ -20°, and another,

University of Cape Town



University of Cape Town

Figure 6.4: Minimised 3D helical structures for a 12–unit maltododecaose strand with glycosidic linkage conformations corresponding to the maltose global minimum in (a): vacuum (ϕ, ψ = -22.5,-25.0) and (b): solution (ϕ, ψ = -47.5,-35.0).

higher valley at $\psi \approx 180°$ (the E–well, *anti*–parallel conformation). The A–well contains the global energy minimum conformation and was found to be by far the most populated in both vacuum and solution. In general, the effect of solution on the maltose glycosidic linkage was to broaden the A–well into more negative ϕ, ψ values and to increase the population of the E–well.

The typical structures of $\alpha(1\rightarrow4)$–linked saccharides in solution and vacuum may be inferred from these PMFs. If repeated for an oligosaccharide, glycosidic linkages taken exclusively from the A–region result in a left handed helical secondary structure. Figure 6.4 shows two helical structures of a maltododecaose strand corresponding to the global minimum energy conformations in vacuum and solution of the maltose $\alpha(1\rightarrow4)$–linkage. The vacuum conformation has 7 glucose units per helical turn and a helical pitch of about 14 Å, compared to a helical turn of about 5.5 units and a helical pitch of around 15 Å for the solution conformation. In water the shift of the global minimum in the maltose PMF into more negative ϕ,ψ regions, results in a more slender and extended helical structure for an oligosaccharide. In addition, the broadening of the A–well upon aqueous solvation implies that $\alpha(1\rightarrow4)$–linked oligosaccharide helices will become more flexible and disordered in solution, with a broader range of low energy helical conformations. However, assumptions of the structure of amylose polysaccharides should not be made on the basis of the A well alone — all the low energy conformations of the $\alpha(1\rightarrow4)$ glycosidic linkage must be considered. The impact of E–well conformations are of particular interest. Because of their low incidence, these conformations are not likely to have a pronounced effect on the properties of maltose and have been largely ignored in modelling studies of $\alpha(1\rightarrow4)$–linked oligomers.[159] However, if the maltose solution populations are extrapolated to large polymers, 1% of the glycosidic linkages will be in an *anti* conformation. These conformations, termed "band–flips", have been recently shown to occur in large cyclo–malto–oligosaccharides, where they are associated with kinks in the strand that are thought to reduce steric strain.[17,29]

Figure 6.5 displays an example of the conformation of a 300 unit amylose strand produced from a random assignment of ϕ, ψ torsion angle values in accordance with the population distribution for maltose in solution. This tertiary structure consists of straight helical sections interspersed with non–helical bend regions and is in agreement with the "interrupted helix" proposed for amylose in solution.[285] Interestingly, the bends are not simply random coil regions, but are each associated with an *anti*–parallel conformation of a ψ dihedral angle. The bends disrupt the helical predisposition of the A–well conformations, rotating the helix by roughly 90°. A similarly constructed conformation based on the maltose vacuum PMF (not shown) has more regular and compact helical conforma-

University of Cape Town

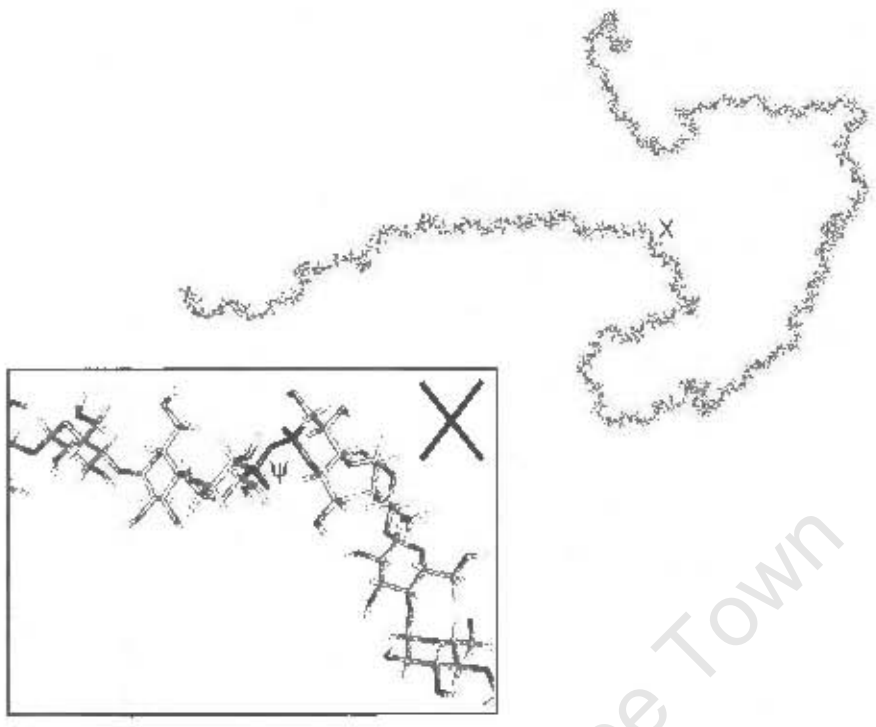


Figure 6.5: Example of a conformation for a 300–unit amylose strand in solution, based on a random assignment of ϕ, ψ angles in agreement with populations calculated from the maltose solution PMF surface. Inset shows enlargement at point X. The indicated ψ dihedral angle is in the *anti* conformation.

University of Cape Town

tions, with fewer *anti* conformation bend regions. Thus, comparison of the vacuum with the solution structure supports the suggested transition of amylose from an interrupted helical conformation in water to a tighter helix in DMSO.[158]

There has been some debate on the nature of the helical twist in the solid state forms of amylose.[16] Conformations taken from the C–well, which is the region of the maltose crystal structure, generate right–handed helices. However this region was found to be unpopulated in both vacuum and solution. A–well configurations of the glycosidic linkages result in left–handed helical conformations of the amylose polysaccharide in vacuum and solution. Therefore, unless packing effects dramatically alter the preferred conformations of the glycosidic linkage, amylose is likely to also form left–handed helices in the solid state. The more hydrated forms of amylose (e.g. B and V_H) should have more disordered, flexible helices with a greater incidence of *anti* orientations of the glycosidic linkages. However, because *anti* conformations result in large bends, steric crowding from neighbouring molecules in a glassy or crystalline solid state could result in lower *anti* populations than is the case in solution or vacuum.

6.3.2 Maltohexaose End–to–End PMFs

The calculated end–to–end PMF for maltohexaose in vacuum is shown in Figure 6.6. In vacuum the stability and preferred conformations of the amylose helix are dependent on stabilising interresidue hydrogen bonds as well as the intrinsic preferences of the glycosidic linkages. The PMF shows two minima. The global minimum occurs in a narrow well around 6 Å. This is the region of nearly circular conformations, which are stabilised by favourable interactions between the reducing and non reducing chain ends. The second minimum region is the plateau ranging from 10 to 20 Å. This area encompasses a host of extended conformations which can be broadly classified into helices (all the ψ dihedral angles are in a *syn* conformation) and ribbon–like shapes (some *anti* ψ dihedral angles occur). The minimum for this well occurs at 16.5 Å, corresponding to a favourable helical conformation with ϕ, ψ =-20,15. This is approximately the global minimum on the maltose vacuum PMF. The maltohexaose chain is thus quite "floppy" and stretches and compresses easily in the 4–20 Å range. Extrapolating from the maltose PMF, the effect of solution should be to extend the second well to longer extensions and flatten it further.

Conformational transitions in $\alpha(1\rightarrow4)$–linked oligosaccharides are likely to be primarily hydroxymethyl group and glycosidic linkage rotations. Aside from increased oscillations and range of movement in the A–well, oscillations of the glycosidic linkages can potentially also involve ψ transitions. It is expected that the glycosidic linkages will undergo occasional ψ transitions from the A– to the E–well if sufficient energy is provided to overcome the energy barrier between them. At 300K, most of the glycosidic linkages are expected to be in the lowest energy *syn* conformation, but transitions between the *syn* and *anti* minima are more likely upon solvation and upon heating or the application of a force. Indeed, ψ dihedral angle transitions occurred with regularity in the maltohexaose simulation when the chain was extended to lengths greater than 25 Å. The vacuum PMF confirms that stretching to this length will supply 8 to 10 kcal/mol of energy to the chain, which is sufficient to overcome the 10 kcal/mol A to E well barrier in the maltose vacuum PMF. The ψ transition typically occurred in the third glycosidic linkage, as this is the central point of the chain and thus the focus of the stretching force. As discussed earlier, *anti* ψ configurations interrupt the helical structure of a chain. For a short oligosaccharide such as maltohexaose, this alters the molecular conformation from a helical turn to a flatter, "ribbon" shape. These two conformations are illustrated in Figure 6.8.

A combination of ribbon and helical populations contribute to the final PMF. As described in the Methods section, the total PMF can therefore be separated into two

University of Cape Town

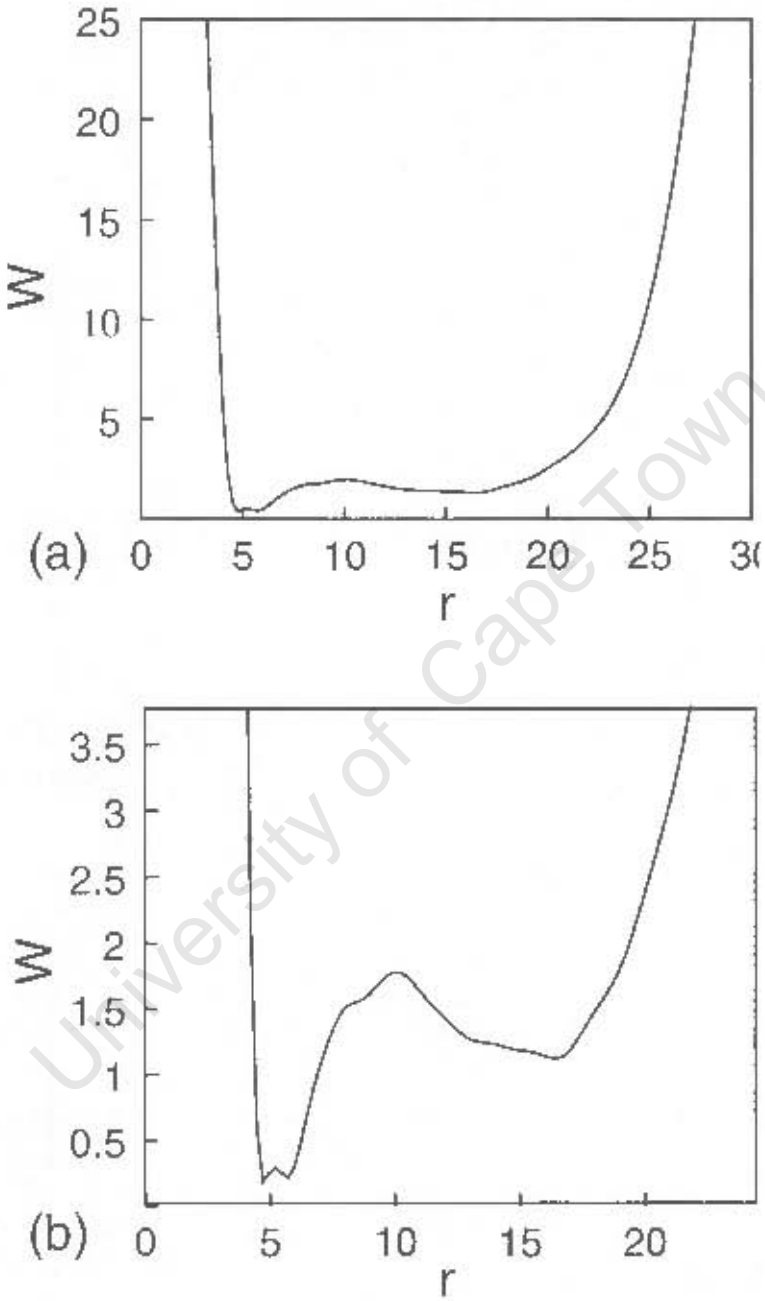


Figure 6.6: (a): The calculated end–to–end potential of mean force energy surface for the maltohexaose strand in vacuum. A magnification of the low–energy regions is shown in (b). The end–to–end distance (r) is in Å.

University of Cape Town

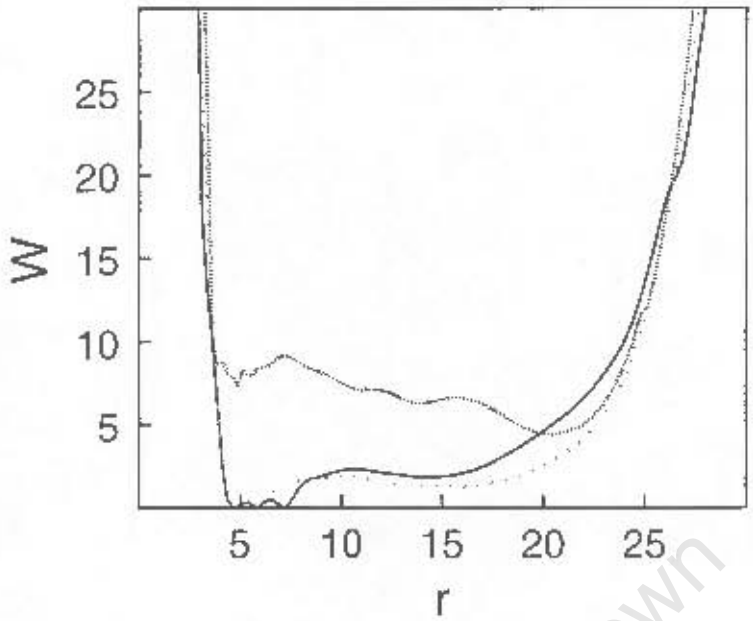


Figure 6.7: Comparison of three end-to-end PMFs for stretching and compressing maltohexaose in vacuum: a PMF for helical (—), ψ_3-*anti*-parallel ribbon (- - -) and all conformations (···). The end to end distance (r) is in Å.

University of Cape Town

ψ_3-dependent end to-end distance PMFs one PMF describing the energy required to stretch and compress helical conformations with ψ_3 in a *syn* orientation and the other the "ribbon" conformations with ψ_3 in an *anti* orientation. These are compared in Figure 6.7. At short extensions the PMF for the helical conformations is similar to the total PMF. This indicates that these conformations dominate in this region. The total PMF follows the contour of the ribbon PMF at longer extensions. The ribbon PMF is quite different from that for the helix. The preferred ribbon conformation is much more extended and has a global minimum at 20.4 Å. The minimum energy well is narrower than for the helix and conformations with the ends of the strand in close proximity are high in energy. This indicates that the ribbon, in contrast to the helix, does not compress easily. The depth and width of the ribbon minimum energy well as compared to that for the helix corresponds to the different dimensions of the A- and E-wells in the maltose vacuum map. The dynamics of the ribbon are thus less flexible than the helix, as the links are "stiffer" (there is a steep energy rise moving out of E well). The PMF of the ribbon in the 20 to 26 Å region is lower than the helix PMF, indicating that stretching the maltohexaose strand to this length results in significant populations of the ribbon conformers. Interestingly, the ribbon and helix PMFs cross again at 26.6 Å, resulting in the helical PMF being once more favourable at very long extensions. Thus, upon extreme stretching, the ribbon conformation should revert back to an extended helix.

University of Cape Town

Figure 6.8: Superposition of a three-dimensional maltohexaose helical conformation (H) onto a ribbon conformation (R). The ribbon conformation has the third ψ torsion angle (ψ_3) in an *anti* conformation, while in the helix all ψ dihedral angles are in the *syn* conformation. Hydrogen atoms are not shown.

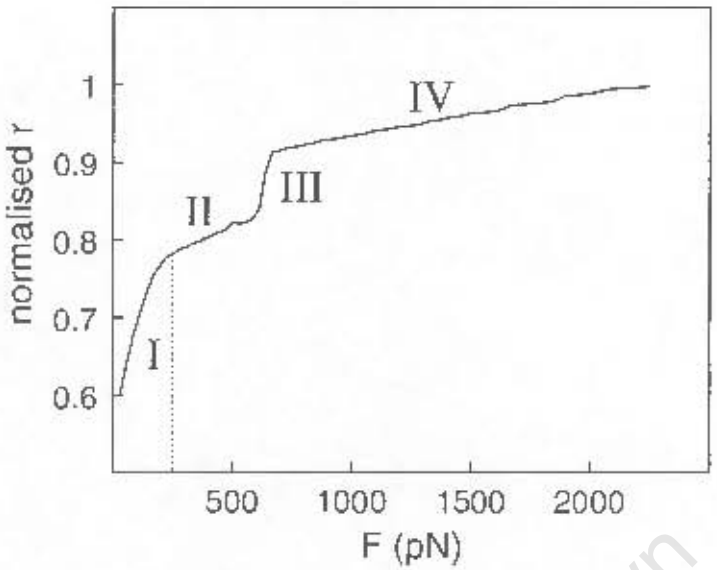


Figure 6.9: The force–extension curve produced by a stretching simulation of an 18–unit amylose oligomer. The curve was adjusted by subtracting the initial force required to prevent collapsing of the helix in vacuum. The O_1 to O_4 end–to–end distance (r) is normalised by the final end–to–end distance at 2250 pN (= 99 Å). The dotted line shows the estimated zero point for a simulation in aqueous solution.

Long amylose strands therefore consist of flexible helical sections joined by inflexible *anti* regions. This has implications for the properties of starch, particularly thermally–treated starch, which is expected to have more *anti* ψ conformations. Once a ψ transition occurs, there is an high energy barrier of $\approx$ 10 kcal/mol (from the maltose PMF) to traverse for a return to the $\psi \approx -30°$ conformation. If the applied force or heat that induced the transformation is subsequently rapidly removed (e.g. by quenching), the strand could potentially be kinetically trapped in a metastable ribbon conformation. This would have implications for the physical properties of amylose that has been subjected to heat or stretching forces. This accords with experiment, in which the thermal history of a starch polymer has been shown to affect its physical properties.[9,13]

6.3.3 Oligosaccharide Stretching Simulation

The force–extension profile obtained from the simulated stretching of an 18–unit amylose oligomer is shown in Figure 6.9. This curve reproduces the distinctive transition seen in force–ramp AFM stretching experiments on carboxymethylamylose molecules (Figure 6.2). The shoulder in the amylose curve was found to be reproducible only with prolonged equilibration at each step, in agreement with the experimental assumption that polysaccharides stretch in equilibrium.[53] If shorter simulations were used, the system

University of Cape Town

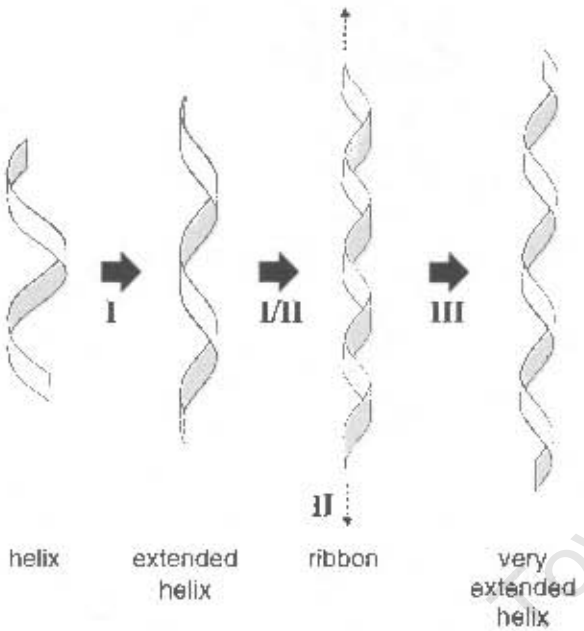


Figure 6.10: Schematic diagram of the processes involved in the first three regions of the force-extension curve.

University of Cape Town

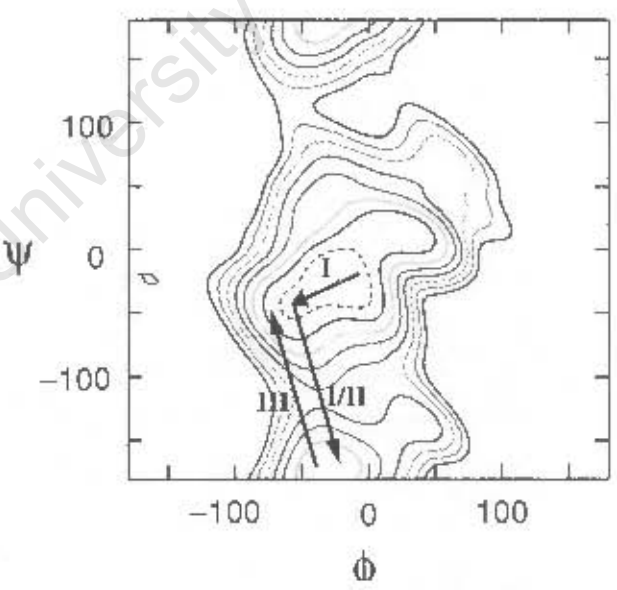


Figure 6.11: The maltose ϕ,ψ PMF in vacuum, with the glycosidic linkage transitions involved in the first three stages of the amylose stretching curve indicated by arrows. Contours appear at 2 kcal/mol intervals, with the lowest energy contour appearing as a dashed line.

did not equilibrate sufficiently and the resultant curves were similar to those produced by the unconstrained simulations of Heymann et al.[120] However, the force required for this transition is at least 300 pN higher than in the experimental curve. This can be explained by the fact that the AFM experiment was performed using a molecule lifted from solution, whereas the simulated stretching experiment was performed in vacuum. In vacuum, strong intramolecular interactions between successive residues as well as neighbouring helical turns increase the force required to extend the strand. In solution, the effect of these interactions is expected to be considerably reduced due to the hydrogen bonding capabilities of the water molecules as well as the general electrostatic masking effect of the solvent.

There are four stages apparent in the stretching profile, labelled I to IV. Figure 6.10 shows a schematic diagram illustrating the process involved at each stage. The initial stage (I) is a steep slope reflecting extension of the oligomer strand from a compact to a more extended helical conformation. In terms of the $\alpha(1\rightarrow4)$ glycosidic linkage ϕ, ψ torsion angles, this involves a shift to the more negative quadrant of the A well, as indicated in Figure 6.11. The slope of region I gives the elasticity of this process as 73 pN/nm. Visually, this region can be seen to be less steep in the simulation curve compared to experimental curve. This is probably because of the absence of solvent molecules. If the extension of the maltose A–well region in solution is extrapolated to amylose, aqueous solvation should result in easier extension of the helix and therefore a steeper slope for this region.

In the gradual transition from stage I to stage II (the gentle bend between 150 and 250 pN), the amylose strand switches from a helical conformation to a ribbon conformation, as illustrated in Figure 6.10. This change is brought about by rotation of most of the ψ angles from *syn* (ψ=0°) to *anti* (ψ=180°) conformations, as can be seen from the time series for the ψ angles from this region of the simulation (Figure 6.12). The end–to–end distance PMFs for the ribbon and the helix show the ribbon conformation to be more stable in a moderately elongated conformation. Thus stretching the $\alpha(1\rightarrow4)$–linked oligosaccharide in this region favours ribbon over helical conformations. Although the time series in Figure 6.12 shows that not all the ψ torsion angles convert into the *anti* conformation, enough do to disrupt any helical character. This is shown by a snapshot of a representative ribbon conformation in Figure 6.13. In aqueous solution the helix–to–ribbon conversion is expected to occur at lower energies, as Chapter 5 showed solution to lower the energy of both the maltose E–well and the barrier between the A– and E–wells.

Stage II begins properly near 275 pN and involves slight lengthening of the ribbon conformation. The ribbon conformation is stiffer than the helical conformation and stretches

University of Cape Town

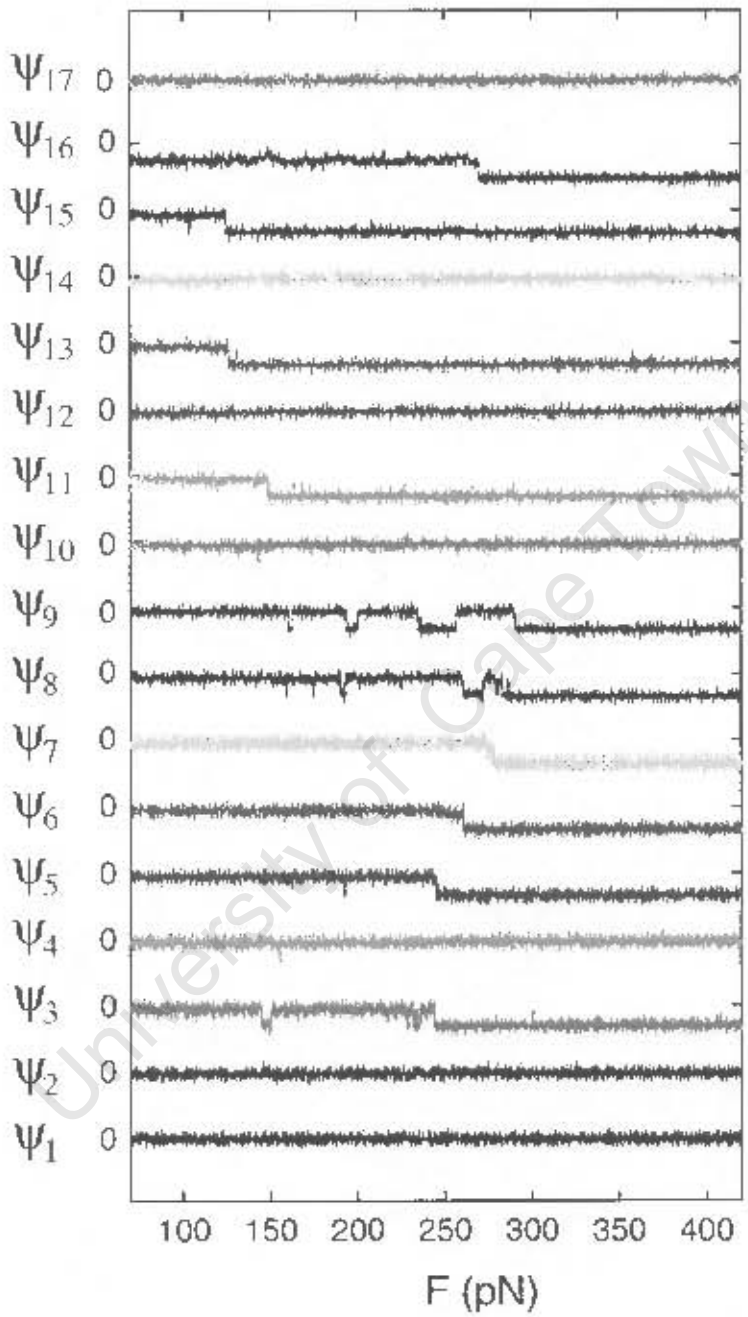
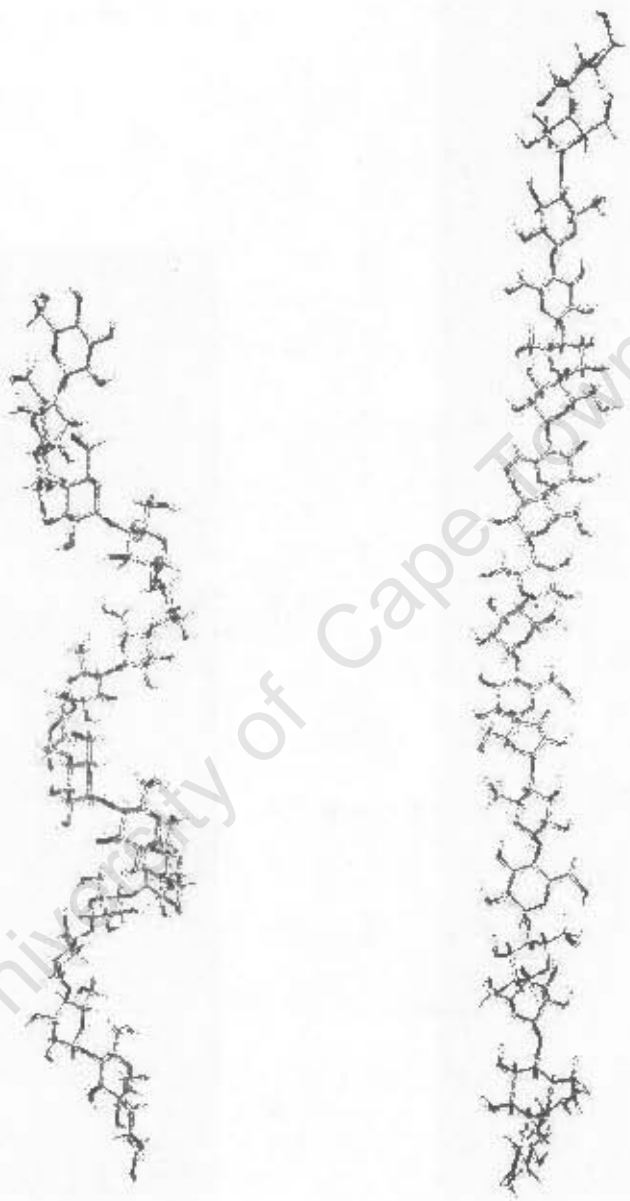


Figure 6.12: A section of the ψ dihedral angles time series from the stretching simulation of the 18 unit amylose strand, showing transitions to *anti* conformations that occur when moving from region I to region II in the force–extension curve.

University of Cape Town



University of Cape Town

Figure 6.13: Snapshots from the stretching simulation of an 18–unit amylose oligomer showing (a) a helical conformation from Stage I of the force–extension curve and (b) a ribbon conformation from Stage II of the force–extension curve.

less readily. This results in a lower elasticity of 800 pN/nm for this region.

Stage III shows a very steep slope, which results in the low elasticity of 53.5 pN/nm. In the rapid onset of Stage III, the pyranose rings in 2 of the 18 residues were found to convert from a chair to a boat conformation. However, chair–to–boat transitions are not the dominating process. The rapid increase in length at this stage is caused primarily by the conversion of all the ψ angles that were in the *anti* conformation back to the *syn* conformation. This produces a very extended helical conformation for the strand, as illustrated in Figure 6.10. This constitutes valuable illustration of the effect seen in the maltohexaose PMFs where, although moderate stretching was found to favour ribbon conformations over helical conformations, the situation was reversed under more severe stretching forces.

At the end of Stage III, the extension of the amylose molecule approaches its contour length and further rotations of the ϕ, ψ glycosidic dihedral angles can no longer lengthen the chain. It is at this point that the majority of the chair–to–boat transitions begin. Stage IV has an elasticity of 1666 pN/nm and involves gradual, anti–cooperative chair to boat conversions of the pyranose rings. This process finally halts at around 2300 pN. Thus the chair to boat conversions occur at higher forces and at a different stage of the force–extension curve than previously expected – Stage IV, not Stage III. These transitions therefore do not produce the characteristic "shoulder" in the amylose stretching curve. Instead, it is largely rotations of the glycosidic torsion angles that account for the elastic behaviour of amylose. This new interpretation is supported by experimental evidence. For example, the removal of the chain elasticity of amylose and pectin upon ring cleavage with periodate can be explained by ring cleavage removing steric hindrances to the glycosidic linkage rotation and thus considerably altering the chain dynamics. In addition, the previously proposed location of the chair–to–boat transition in amylose has a force threshold (275 ± 45 pN) which is well below the values predicted by theory.[51,119] Moreover, previous simulations of stretching simulations on 16 unit oligomer of amylose were found to produce cooperative transitions of the pyranose rings only if the ψ angles were constrained in *anti* conformations.[120] If the angles were left in the *syn* conformation, the chair–to–boat transitions started at a higher force and proceeded in an anti–cooperative fashion, requiring increasingly higher forces for each successive transition. Finally, the stretching mechanism delineated here is more consistent with the complete reversibility of the stages I to III observed in experiments, which suggests that the mechanism only involves traversal of relatively low energy barriers.[53]

A PMF calculated from the first 30 steps of the stretching simulation is shown in Figure 6.14. The PMF has an initial region of high elasticity which gradually converts

University of Cape Town

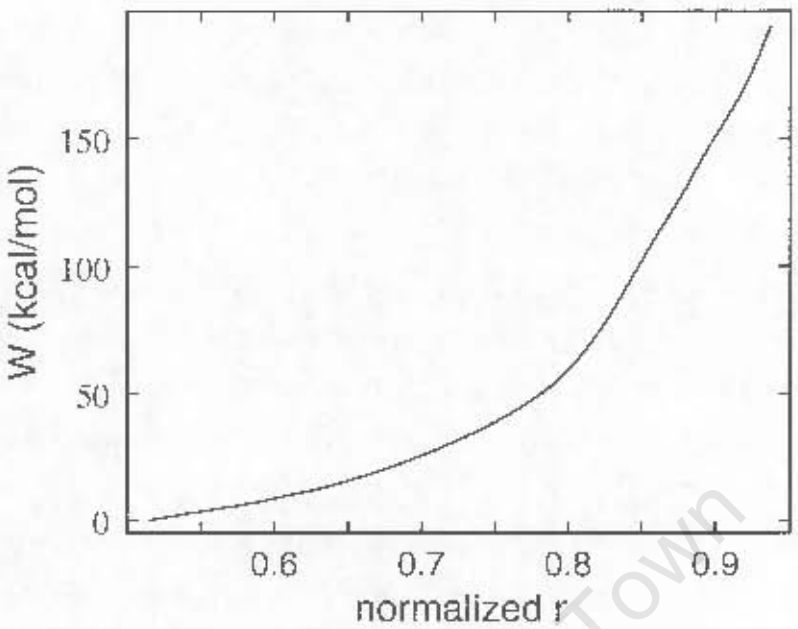


Figure 6.14: PMF calculated from the first 30 steps (up to an applied stretching force of 830 pN) of the stretching simulation of an 18–unit amylose oligomer. The end–to–end distance (r) is normalised as in Figure 6.9.

University of Cape Town

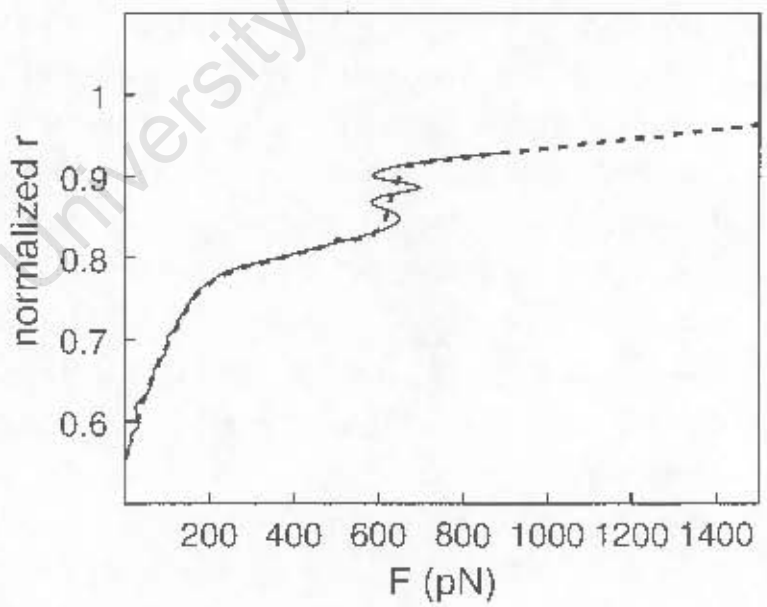


Figure 6.15: The force–extension curve calculated from the derivative of the amylose end to–end PMF (—) compared to that obtained from the averages taken for each stage of the simulation ( ). The curves are normalised as in Figure 6.9.

to a stiffer region. The details of the change in slope in the PMF can be more easily seen in the corresponding force extension curve calculated from the derivative of the PMF, which appears in Figure 6.15. This is a good match with the force–extension curve calculated from the average length of the molecule at each step. However, the derivative curve provides more detail of the complex force–extension relationship remarked on by Marszalek et al.[53] Both Stage I and Stage II show evidence of non–linear force–extension relationships which are caused by rotation about the glycosidic linkages.

The elasticities calculated from the curves of each stretching region are considerably lower than those estimated by Marszalek et al compare 800 pN/nm for Stage III with their value of 3973 pN/nm and 1666 pN/nm for Stage IV with their value of 27800 pN/nm. These estimations were based on a two–state freely jointed chain model which required a prediction of the number of monomers in the strand being stretched and neglected to take into account possible helical conformations of the chain. Therefore, as helical conformations were ignored, it is likely that the two–state model underestimated the actual contour length of the chain. This in turn would have resulted in an underestimate of the number of monomers present in the molecule, and thus the high values for the calculated elasticities.

The force–extension curve produced by simulation differs from the experimental curve in that the forces required for the onset of each stage are at least 300 pN too high. Aside from errors in force field parameterisation, there are a number of possible contributions to this discrepancy. Firstly, the forces required to break weak biochemical bonds are not constant, but are dependent on the rate at which the force is increased.[286] As mentioned in Section 6.2.3, faster ramping rates were found to increase the force required to reach a particular extension of an amylose fragment. The ramping rate used in the 18-unit amylose fragment simulation (1.85 to 1.39 pN/ns) is six orders of magnitude faster than the experimental ramp rate of 2.5 pN/ms.[53] Therefore, it is expected that the forces measured should be somewhat too high. Secondly, the simulations were carried out in vacuum while the experiments used a wet sample. In the previous chapter, an aqueous solvent was shown to extend the range of the A–well as well as to lower the barrier between the A– and E–wells. Stage I and the onset of Stage II in the amylose curve are affected by the extent of the A–well and the height of the energy barrier between the wells respectively. By extension, water should lower the force required to extend a helix and bring about *syn* to *anti* transitions on the amylose stretching curve, resulting in a much more rapid progress to Stage II. If this is assumed to shift the entire curve by ≈ 250 pN (as indicated by the dotted line in Figure 6.9), this would result in a force–extension curve with a better match to the experimental curve. Finally, the possiblilty that the simulated

University of Cape Town

curve is not converged due to sampling limitations should also be considered.

6.4 Conclusions

This study provides evidence that the tertiary structure of the $\alpha(1\rightarrow4)$–linked amylose polysaccharide is comprised of left–handed helical regions interrupted by bends which are associated with *anti*–parallel ψ conformations. The physical properties of the strand should alter depending on the ratio of *syn* to *anti* conformations in the strand. PMF studies on maltohexaose showed conformations containing *anti* ψ dihedral angles to be more extended, stiffer and less compressible than the helical conformations. This has implications for the properties of starch, particularly thermally treated starch, which is expected to have more *anti* conformations due to kinetic trapping of the torsion angles upon rapid cooling. In addition, the ψ transitions are more likely in the presence of water and may be a factor in the α relaxation processes occurring at the glass transition in amylose glasses. Two separate end to-end distance PMFs were calculated for maltohexaose to investigate the relative stabilities of helical conformations with the central ψ_3 angle in an *syn* orientation versus ribbon conformations with central ψ_3 angle in an *anti* orientation. These showed that the relative stabilities of the maltohexaose conformations alter depending on the extension of the maltohexaose molecule. At short extensions, the helical conformations are lower in energy than the ribbon conformations. However, as the molecule is extended, the situation reverses and ribbon conformations predominate. Then, at extreme extensions, helical conformations are favoured once more.

Transitions of the glycosidic linkages between two principal energy regions were found to be a reoccurring theme in the dynamic behaviour of $\alpha(1\rightarrow4)$–linked oligomers. The stretching simulation of the 18–unit malto–oligomer produced a force–extension curve similar to AFM experiment. However, the characteristic transition in this curve was identified as being caused predominately by a complex helix–ribbon–helix conformational transition of the strand. The 2D PMFs calculated for maltose as well as the end to end distance PMF for maltohexaose were invaluable in interpreting the energetic reasons for this series of conformational changes. Chair to–boat conformational changes of the pyranose rings only occurred anti cooperatively at very high stretching forces. Therefore, glycosidic angle transitions were found to predominately determine the dynamic response of $\alpha(1\rightarrow4)$-linked oligosaccharides to stretching forces and chair–to–boat transitions of the pyranose rings play a smaller role than has been recently proposed.

In conclusion, ψ transitions between *syn* and *anti*–parallel conformations of the dihedral angles have here been shown to be of prime importance for the preferred conforma-

University of Cape Town

tions, dynamics and relaxation behaviour of α(1→4)-linked glucans and are thus likely to be ultimately responsible for many of the unique physical properties of the amylose polysaccharide.

University of Cape Town

University of Cape Town

Chapter 7

The $\beta(1\rightarrow4)$ Glycosidic Linkage

7.1 Introduction

The previous two chapters discussed the $\alpha(1\rightarrow4)$–linkage. In this chapter, the investigation is extended to the conformation and dynamics of the Glc–$\beta(1\rightarrow4)$–Glc glycosidic linkage, which is the sole linkage found in the cellulose polymer. Comparison of the $\beta(1\rightarrow4)$ linkage with the $\alpha(1\rightarrow4)$ linkage should quantify the effects of a change in stereochemistry (α versus β) on the conformational space of the $(1\rightarrow4)$ glycosidic linkage.

Cellulose is the major structural component of higher plants and hence the most abundant biopolymer on earth. Cellulose forms rigid, inflexible, semi–crystalline fibres biologically and, in 1912, became the first polymer to be investigated by X–ray crystallography. However, it took a further 90 years for the structure of native cellulose to be solved satisfactorily[288] and the structure of some forms of cellulose remain to be determined. Pure cellulose exists in four distinct polymorphic forms, labelled I to IV. Values of the ϕ and ψ torsion angles for some of the cellulose allomorphs are listed in Table 7.2.[30] All have a 2 residue repeat unit, but differ in their molecular packing and hydrogen bond arrangements. Native cellulose (cellulose I) forms ribbon–like chains which pack into crystalline microfibrils. These microfibrils contain two crystalline allomorphs — cellulose Iα and cellulose Iβ — in which the chains are packed slightly differently in a parallel arrangement.[150,292] Though still under investigation, the currently accepted structure for cellulose Iβ has intramolecular hydrogen bonding between the O_5 and the O_3 hydroxyls of contiguous residues (a typical hydrogen bond occurring in cello oligomer structures) and all the primary alcohols are in the *tg* position.[288] The crystal structure for Iα remains to be solved. Cellulose I is a metastable form and undergoes a transition to stable cellulose II when it is either regenerated by spinning out of solution or treated with strong alkali (mercerization). Cellulose II has similar ϕ,ψ conformations, but the packing is antiparal-

University of Cape Town

Oligosaccharide	H_2O Solubility (%)
Cellobiose	∞
Cellotriose	25
Cellotetraose	12.5
Cellopentaose	5
Cellulose	Insoluble

Table 7.1: H_2O solubility of crystalline cello–oligomers.[287]

University of Cape Town

Cellulose	I_β[288]	II[289]	III_I[290]	IV_I[291]	IV_2[291]
ϕ	25.3	24.6	21.1	21.9	26.2
ψ	-25.0	-25.3	-27.2	-24.4	-28.1

Table 7.2: The ϕ, ψ torsion angles from the crystal structures for cellulose fibres, taken from the review by Zugenmaier.[30]

Cello oligomer	β cellobiose[23]	Methyl β-cellotrioside[25] (ave. of 8)	cellotetraose[26] (ave. of 8)
ϕ	44	26	26
ψ	-23	-27	-27
O_6 conformation	*gt*	*gt*	*gt*

Table 7.3: The ϕ, ψ torsion angles from the crystal structures of selected cello oligomer model molecules.

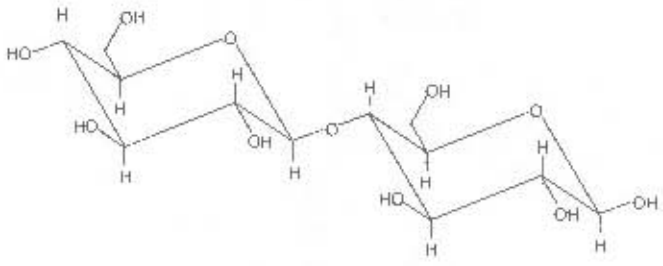


Figure 7.1: Line diagram of the β–cellobiose disaccharide (4 *O* β D glucopyranosyl–D–glucopyranose).

lel. A recent structure determination has shown all the primary alcohols to be in the *gt* conformation.[289]

Besides its crystalline forms, cellulose also occurs as amorphous glasses. Relaxation processes occurring in amorphous cellulose below its glass transition temperature have been explored by both mechanical and dielectric spectroscopy. Mechanical spectroscopy experiments exhibit two sub-T_g mechanical relaxation processes, termed β and γ.[59] The β–relaxation process has been found to be strongly dependent on water content and is undetectable by dielectric spectroscopy on amorphous dry cellulose, which shows only a broad γ–relaxation.[5] This γ relaxation decreases somewhat in strength with increasing water content. This activation energy, which corresponds to the free energy barrier that must be crossed in a transition process, has been calculated to be in the range of 32 to 36 kJ/mol (7.6 to 8.6 kcal/mol). The origin of this relaxation remains unclear, but it has been proposed to result from a combination of the hydroxyl and hydroxymethyl group relaxations. The β–relaxation activation energy is lowered from 85 to 60 kJ/mol (20.3 to 14.33 kcal/mol) in moving from a 0 to 6 w/w% glass. The molecular origins of the β–relaxation are thought to be localised motions of the main chain segments.

Although the crystal structure is known for most of the forms of cellulose,[30] as well as some of the short cello-oligomers,[24] detailed studies of their solution structure and dynamics are still lacking. An NMR study of β–1,4 linked cello–oligomers in DMSO and water solution by Sugiyama et al. predicted that the oligomers all have similar average dihedral angle values in both solvents, in the region of ϕ, ψ =45°,-38°.[36] These angles would result in the larger oligomers forming a loose helical coil in solution. The dynamics of cellulose in aqueous solution have been probed by AFM experiments, which provide further evidence for the stiffness and inflexibility of this molecule. The characteristic cellulose extension force curve measured by Marszalek et al.[122] (Figure 7.2), shows an extremely rapid initial elongation at very low force, with a subsequent rapid onset of a low elasticity region. Transitions of the type seen in the curve for amylose are absent.

University of Cape Town

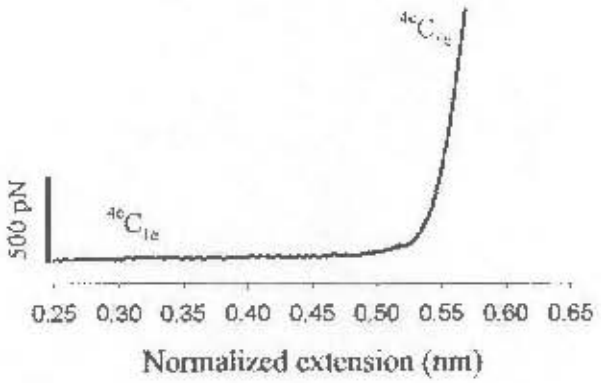


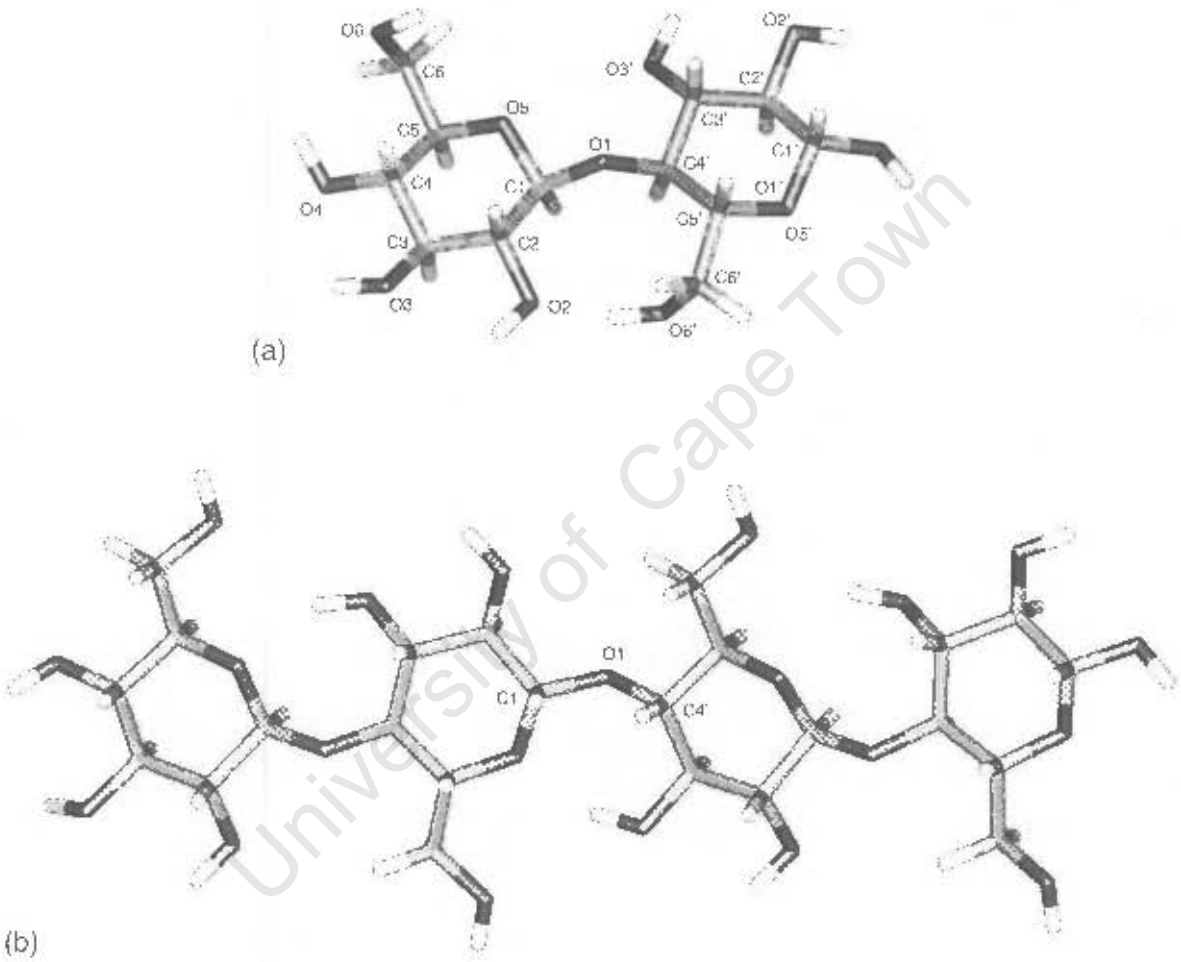
Figure 7.2: The force–extension curve for cellulose produced by the AFM experiments of Marszalek et al.[122] The curve was normalised using the equation: $X_{norm} = 0.542x/l^c$, where $0.542x$ corresponds to the length of the O_1O_4 vector in the β–D–glucopyranose monomer of cellulose and l^c is the contour length of cellulose determined from a freely jointed chain (FJC) model fit to the data.

This is thought to be because the glucose residues in cellulose do not undergo a chair–to–boat transition; ball–and–stick models show no increase in the O_1–O_4 vector for the boat conformation when O_1 is in a β orientation.[121]

As with amylose, attempts have been made to extrapolate the conformation and dynamics of cellulose from that of model di– and oligosaccharides. For example, as experimental X–ray and spectroscopic data on the various crystal forms of cellulose are not sufficient for a complete structure determination, cellulose structures have been refined using known crystal structures of short oligomer model compounds (dimers, trimers and tetramers).[25,26,293] Cellobiose (4–O–β–D–glucopyranosyl–D–glucopyranose, Figure 7.1) is the dimer residue of cellulose. Cellotriose and cellotetraose are the corresponding 3– and 4–unit models. The ϕ, ψ dihedral angle conformations found in the X–ray crystal structures of β–cellobiose, methyl β–cellotrioside and cellotetraose are listed in Table 7.1. Additional values for structures of related model compounds are listed by Peralta–Inga et al.[24]

The areas of conformational space available to the glycosidic linkage in cellobiose have been explored by calculation of relaxed ϕ,ψ adiabatic maps.[83,84,147,148,294] These studies indicate that the $\beta(1\rightarrow4)$ linkage is intrinsically more flexible than the $\alpha(1\rightarrow4)$ linkage: the equatorial conformation of the hemiacetal at C_1 leading to fewer steric clashes.[83,85,294] Moreover, optical rotation experiments conducted in the late 1980's on cellobiose in water showed a complex conformational equilibrium, likely to be produced by rapid transition of the molecular conformation between a number of minimum energy wells.[44] Thus, the contrasting physical properties of elastic, soluble amylose and the inelastic, insoluble cel-

University of Cape Town



University of Cape Town

Figure 7.3: Three-dimensional structures of cellobiose (a) and cellotetraose (b) in $\phi, \psi = 0,0$ conformations, with the atomic naming convention shown.

lulose cannot be simply explained by comparison of their constituent disaccharides. In addition to crystal packing forces, the presence of cooperative interresidue and intermolecular interactions have been proposed to fix the conformation of cellulose into a particular shape and thus cause the stiffness and strength of this polysaccharide. The hypothesis that increasing the length of cello-oligomers increases the strength of the cooperative interresidue hydrogen bonds is supported by the fact that the solubility of $\beta(1\rightarrow4)$–linked oligomers decreases dramatically with increasing residue number (Table 7.1).[287]

Recently, the molecular dynamics of polycrystalline cellobiose was investigated using solid state NMR.[63] This study suggested that the hydroxyl groups have at least two classes of dynamic behaviour (characterised by an activation energy of 8.74 kJ/mol= 2.09 kcal/mol), while rotation of the hydroxymethyl groups was found to be efficient only at high temperatures (activation energy 40 kJ/mol= 9.6 kcal/mol). Molecular dynamics simulations of the motion of cellobiose has not been nearly as numerous as maltose. The last published MD simulations of cellobiose in both vacuum[85] and solution[138] were performed ten years ago. These early investigations showed cellobiose to have similar ϕ, ψ conformations in solvent as in vacuum, with hydrogen bonding in solution persisting to a significant extent.

In this chapter, calculated ϕ, ψ PMF surfaces for β–cellobiose in vacuum and solution are analysed for their implications for the preferred conformations and dynamics of the $\beta(1\rightarrow4)$ linkage. A ϕ, ψ adiabatic map for β cellobiose was also calculated, but is not shown as it is very similar to the vacuum PMF. To explore the effect of additional monomers on the conformational freedom of the $\beta(1\rightarrow4)$–linkage, the cellobiose PMFs are compared to a calculated vacuum PMF for the central glycosidic linkage in cellotetraose. This is hoped to assist in explaining the decrease in solubility of the cello–oligomers with increasing residue number and thus shed light on the marked lack of conformational freedom in cellulose. As was demonstrated for maltose and dixylose in Chapter 5, comparisons of the τ dihedral (H_1–C_1–C_4–H_4) PMFs of cellulose and cellotetraose provide a 1D depiction of the shifts in the minimum energy valleys and the energy barrier heights of these two molecules. Finally, to investigate the dynamics of the cellulose polymer, a stretching simulation of a 12–unit cello–dextran, cellododecaose, was undertaken. The PMF for the early stages of the cellulose stretching simulation was calculated and is compared to that obtained for amylose.

University of Cape Town

7.2 Methods

7.2.1 2D ϕ,ψ PMF Calculations

The glycosidic linkage ϕ, ψ PMFs were calculated for β–cellobiose in vacuum and aqueous solution and β–cellotetraose in vacuum using the iterative adaptive umbrella sampling method described in Section 3.5. Both the cellobiose PMF calculations as well as the cellotetraose PMF calculation were begun from a $\phi, \psi = 0,0$ conformation (illustrated in Figure 7.3). The cellotetraose PMF was calculated by applying the umbrella potential to the central linkage only, while the other two linkages were allowed to relax freely.

In each case, the 360°×360° umbrella potential surface for the ϕ, ψ dihedral angles was represented as a two dimensional grid of points, with a grid separation of 2.5°. At each integration step of a simulation, the biasing potential energy to be applied to the current ϕ, ψ position was calculated from a cubic spline of the umbrella energy surface. After each simulation, a two-dimensional ϕ, ψ distribution histogram was calculated by summing the number of configurations in each 2.5°×2.5° bin over the production phase of the trajectory. The WHAM procedure was then applied, using a tolerance value of 0.001, to obtain an optimum combination of the all the simulations distributions and thus the next estimate of the PMF. This PMF surface was extrapolated into unexplored regions by setting unsampled bins to the maximum value of the sampled bins, treating the border regions as described in Section 3.4. The resultant PMF surface was smoothed three times using Equation 3.9. Subsequently, regions in the 2D ϕ, ψ PMF associated with high–energy steric clashes of the atoms were removed. This was done by setting any bin in the PMF with an energy value greater than a predetermined cut–off of 20 kcal/mol to the cutoff value. The umbrella potential for the next simulation was set to the negative of this PMF estimate.

Each umbrella simulation began with an equilibrium phase of 500 ps, followed by a production phase ranging from 2 to 20 ns in length. Only the production phase was used to calculate the simulation distribution. In the case of β–cellotetraose, the simulations had to run for long enough to enable the non-central glycosidic linkages to relax, requiring long simulations of 20 ns in length in order to converge completely.

Successive simulations in a series were gradually run for longer periods as the PMF estimates improved and more of the whole of conformational space was explored. The first simulation in each series was started from a minimised and equilibrated coordinate set and performed with a flat umbrella potential surface. Subsequent simulations were begun from the final conformation of the previous simulation, to facilitate gradual exploration of

University of Cape Town

phase space. After every 8 simulations, the next simulation was begun from the original equilibrated starting configuration. Convergence of the whole adaptive umbrella sampling procedure was defined to occur when, for a single simulation, every bin k in $n_{j,k}$ in the ϕ, ψ range of interest was occupied at least once. Up to 200 simulations were used to obtain the final PMF, but at any time not more than the latest 50 distributions were used to estimate the next PMF; earlier simulations were gradually discarded as new, more extensive distributions were produced. A total of approximately 500 ns of simulation time was required to produce each of the final PMF surfaces.

The 1D τ PMFs for cellobiose were calculated as described in Chapter 5. The τ population distribution was obtained by summing the contributing ϕ, ψ populations from the corresponding 2D ϕ, ψ PMFs. The τ PMF was then calculated from the τ population distribution by applying Equation 3.1.

7.2.2 Oligosaccharide Stretching Simulation

The 12-unit β cellododecaose strand was stretched by applying equal, but opposite, forces to the terminal O_1 and O_4 atoms at the reducing and non reducing ends of the chain respectively, using the method outlined in Chapter 3. Due to the long equilibration time required, the simulation was performed in vacuum. The starting structure for the simulation was produced by assigning all the glycosidic bond angle conformations in cellododecaose as $\phi, \psi = 50, 0$. The simulation began with an applied stretching force of 55.58 pN, to prevent the molecule from folding back on itself in the vacuum. During the course of the simulation, the stretching protocol involved ramping the force by 27.79 pN (0.4 kcal.mol $^{-1}$.Å $^{-1}$) every 4.5 ns, up to a maximum force of 3000 pN. This corresponds to a ramping rate of 6.18 pN/ns, which, though considerably faster than that used for amylose, was found to be adequate for reproduction of the experimental curve. Each 4.5 ns simulation step consisted of 500 ps of equilibration, followed by 4 ns of data collection. However, the first stage was run for longer than the others (5 ns) to ensure thorough equilibration and distribution of the applied force along the length of the molecule. A total of 110 ramping steps was applied during the course of the 500 ns simulation. A force–extension relationship for cellododecaose was obtained by plotting the average extension during each production step against the applied force. This curve was normalised by dividing the length of the molecule by the maximum extension achieved by the end of the simulation — 67 Å— and adjusted by subtracting the initial force applied to keep the molecule extended in vacuum. Using the WHAM procedure, an end–to–end PMF for stretching cellododecaose was calculated for the initial 30 stretching stages from the

University of Cape Town

time–series histogram generated at the production phase of each step.

7.2.3 Simulation Conditions

Molecular dynamics simulations were performed using the program CHARMM[222] (version 27b1), with modifications incorporated into the USERE routine in order to implement both the two dimensional adaptive umbrella sampling PMF calculations and the stretching simulation. The CSFF carbohydrate parameter set (Chapter 4) for the CHARMM force field was used to model carbohydrates in all the simulations. The TIP3P model was used to represent water.

For the solution simulations, the β–cellobiose disaccharide was placed in an equilibrated cube of 512 TIP3P waters. Solvent water molecules that overlapped with the solute molecule were removed and the system was equilibrated for 500ps. The solution simulation surrounded the cellobiose disaccharide with 489 TIP3P water molecules[248] in a cube of length 24.64 Å. Minimum image periodic boundary conditions were applied to the cube in order to eliminate edge effects.

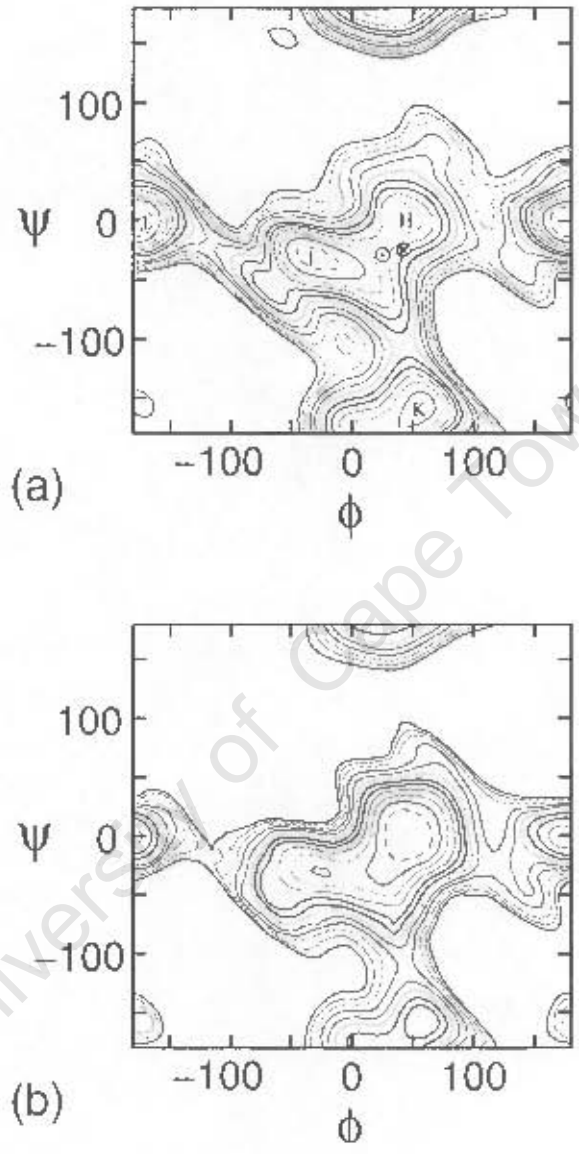
All simulations were performed under the canonical ensemble (constant n, V, T) and using stochastic Langevin dynamics with a frictional coefficient of 62.5 to maintain a constant temperature of 300K. Initial velocities for the atoms were selected at random from a Boltzmann distribution at 300 K. The equations of motion were integrated using a Leap-Frog Verlet integrator[256] with a step size of 1 fs. The SHAKE algorithm[258] was used to fix the length of bonds involving hydrogen atoms and the water molecule geometry throughout each simulation. Non-bonded interactions were truncated using a switching function applied on a neutral group basis between 10.0 and 12.0 Å. The groups corresponded to electrically neutral collections of atoms in the carbohydrate molecules and entire water molecules for the solvent.

University of Cape Town

7.3 Results and Discussion

7.3.1 Cellobiose

The ϕ, ψ PMF surfaces for β cellobiose in vacuum and solution (Figure 7.4) show the conformational space of the $\beta(1\rightarrow4)$ linkage to be less restricted than the corresponding $\alpha(1\rightarrow4)$ linkage. Specifically, the low–energy areas of the cellobiose PMF reach into the ϕ=180° region, which is not the case for maltose. These extended low–energy regions provide support for the hypothesis that the $\beta(1\rightarrow4)$ glycosidic linkage is intrinsically



University of Cape Town

Figure 7.4: Glycosidic linkage ϕ, ψ free energy surfaces for cellobiose. (a): PMF surface in vacuum. (b): PMF surface in TIP3P solution. Contours appear at 1 kcal intervals above the global energy minimum. The first contour at 1 kcal/mol is dashed. Contours higher than 12 kcal/mol above the global minimum are not shown. Crystal structure values of ϕ and ψ for cellobiose ($\odot$) and cellotetraose/cellulose I ($\otimes$) are indicated.

	Cellobiose		Cellotetraose
	Vacuum	Solution	Vacuum
H	2.4	0.0	0.0
I	0.6	1.9	1.0
J	0.0	3.5	1.3
K	1.4	4.3	2.4

Table 7.4: Approximate relative free energies of the four minimum energy regions in cellobiose and cellotetraose. Energies are rounded off to the nearest 0.1 kcal/mol.

flexible.[67]

Four minimum energy regions are apparent on the PMF surfaces, at ϕ, $\psi \approx$ 50,0 (marked H on Figure 7.4 (a)); -30,-30 (I); 180,0 (J) and 50,-175 (K). In contrast to the $\alpha(1\rightarrow4)$–linkage, conformations of the glycosidic linkage dihedral angles in the general ϕ,ψ =0,0 region of conformational space (wells H and I) correspond to *syn* conformations of the glucose residues, whereas conformations at either ϕ,ψ =0,180 (K) or ϕ,ψ =180,0 (J) correspond to *anti* conformations. The J and K well conformations are of interest because of their potential importance for providing stable turns in the cellulose polymer. Optical rotation experiments indicate that, while K–conformations are likely to be of small contribution for cellobiose in solution, J–conformations may be present to a significant extent.[44]

The vacuum surface, which is similar to the cellobiose adiabatic map calculated by Taylor et al. using the precursor HGFB CHARMM force field,[147] has the I– and J–minima close in energy (Table 7.4). The global minimum occurs in the J–well at ϕ,ψ= 180,-2.5. Previously calculated adiabatic maps all show minima at ϕ,ψ =0,180 and ϕ,ψ = 180,0, but do not agree on the exact location of the *anti* parallel minimum energy wells and the global energy minimum in cellobiose. Calculations using the MM3 force field tend to have one or two minima located in the $\phi \approx$ (0° to 60°), $\psi \approx$ (-60° to 60°) region.[83,84,148,294] However, a surface calculated using the HGFB CHARMM force field shows one minimum at $\phi,\psi \approx$ 45,0, another in the negative ϕ region at $\phi,\psi \approx$ -35,-40, with the global minimum appearing at ϕ,ψ =0,180.[147] In agreement with the last–mentioned study, a recent *ab initio* calculation of cellobiose conformations also found the global minimum energy structure to be a ϕ,ψ =180,0 conformation,[202] in contrast to the crystal conformation.[23]

The H minimum is the least favoured of the energy wells. The reported crystal structures for cellobiose, cellotetraose and the known structures of cellulose have ϕ,ψ angles

University of Cape Town

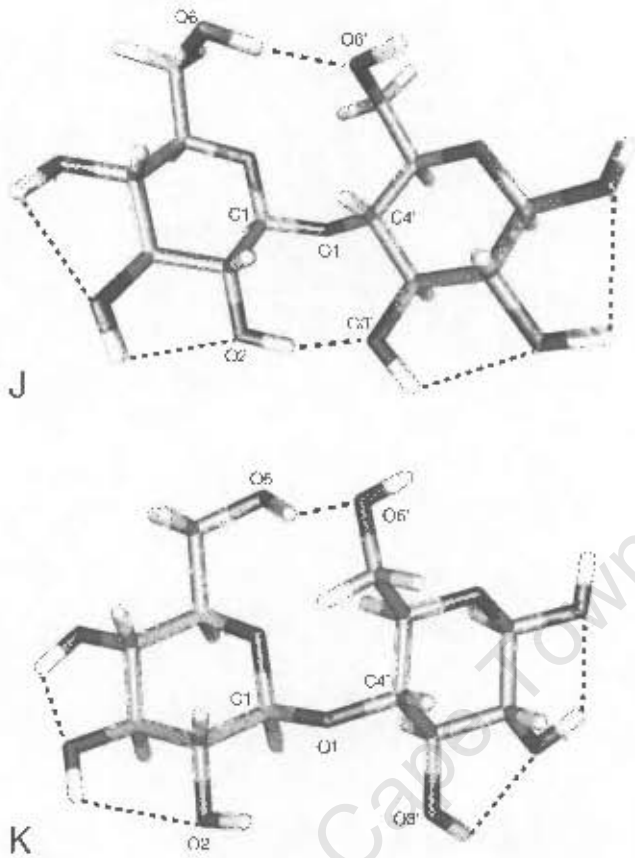


Figure 7.5: Examples of minimum energy configurations for cellobiose in vacuum from the J–well ($\phi, \psi = 180°,0°$) and the K–well ($\phi, \psi = 65°, -170°$). Stabilising hydrogen bonds are shown with dashed lines.

University of Cape Town

that fall chiefly on the saddle region between H and I and not within a minimum energy well, as shown in Figure 7.4 (a).[24,30] The crystal structures are approximately 4 kcal/mol above the global energy minimum in vacuum and between 1 and 3 kcal/mol above the minimum in solution. The discrepancy could be attributed to packing forces and intermolecular hydrogen bonds within the crystal stabilising the crystal conformations.

As was discussed in the maltose chapter, favourable conformations in vacuum are stabilised by internal hydrogen bonds. Interresidue hydrogen bonds are the most affected by the conformation of a glycosidic linkage and therefore have the greatest effect on which conformations are favoured. Conformations that promote strong bonds of this type are expected to be preferred over those which preclude them. In vacuum, conformations that have primary alcohols in a *syn* arrangement should have an energetic advantage over those in the *anti* conformation, as there is more flexibility in their orientation than with interresidue bonds involving secondary hydroxyls. Conformations from the J region are stabilised by two interresidue hydrogen bond interactions, shown in Figure 7.5 (a).

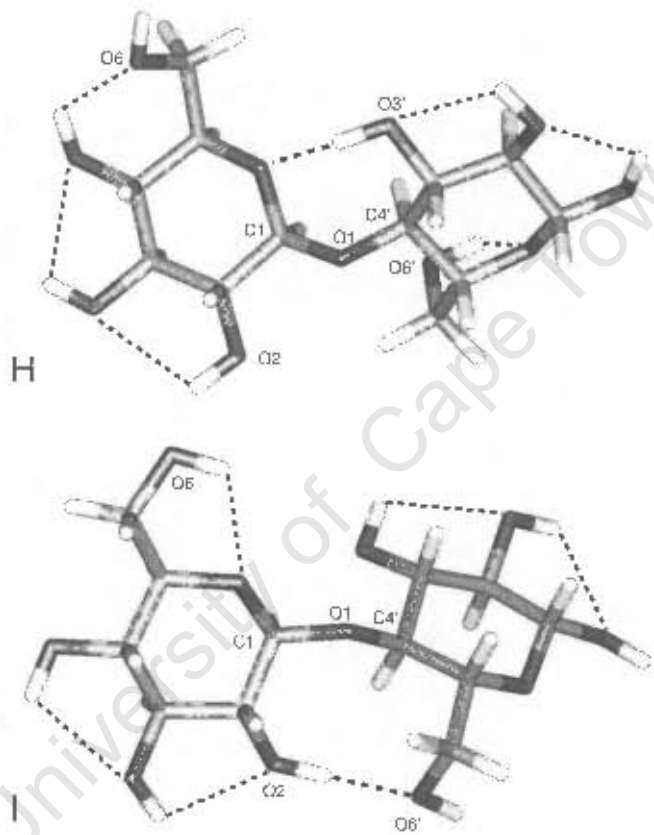


Figure 7.6: Examples of minimum energy configurations for cellobiose in vacuum from the H–well (ϕ, ψ = 40°,0°) and the I well (ϕ, ψ = -30 °, -30 °). Stabilising hydrogen bonds are shown with dashed lines.

University of Cape Town

The first is a strong hydrogen bond between the two primary alcohols, which are both in the *gg* conformation. The second is an O_2 O_3 hydrogen bond interaction. Both of these hydrogen bonds can alternate as to which are the donor and acceptor hydroxyls. K–region conformations are similar to the J well region, but in this region the formation of a hydrogen bond between the primary alcohols requires them both to be in the *gt* conformation. However, the O_2 and O_3' hydroxyl groups are too far apart in the K–well to allow for a second interresidue hydrogen bond. This results in the K–region being higher in energy than the J–region.

The H–well is the least favoured of the minimum energy wells in vacuum, even though the inclined ring orientation leads to few steric clashes. This is because the only possible stabilising interresidue hydrogen bond for these conformations is between the O_3' hydroxyl and the O_5 atom, as shown in Figure 7.6 (a hydrogen bond observed in all crystal structures of cellulose). In contrast, conformations in the favoured I–well allow for a strong hydrogen between O_2 and O_6' in the *gg* conformation. Though I–well conformations have not been identified in crystal structures, both optical rotation[44] and NMR[48] experiments on cellobiose suggest a small population of this region in solution.

Solvation has a much more dramatic effect on the free energy surface of the $\beta(1\rightarrow4)$–linkage as compared to the $\alpha(1\rightarrow4)$–linkage. The removal of the enthalpic advantage of intramolecular hydrogen bonds results in an increased population of entropically favourable conformations which allow for relatively unhindered rotation of the primary alcohol. Particular minimum regions, or energy valleys, can be resticted in vacuum either by steric clashes between atoms, or by a constraint on length of a stabilising hydrogen bond. If the former is the case, no broadening of a well would be expected in solution. However, if the latter holds, solution should increase the range of the minimum energy region by removing the need for the stabilising bond.

Solvation results in the lowering of the general H/I valley region with respect to the other two minima. In addition, there is a marked preference for the H–conformation over the I–conformation in solution. The solution global energy minimum has moved from the adiabatic and vacuum PMF value of ϕ,ψ = 180,0 to ϕ,ψ = 47.5, 2.5. The H–well conformations have few steric clashes and the hydroxymethyl group can undergo relatively unhindered rotation. Thus, removal of the dependence on intramolecular hydrogen bond formation results in considerable broadening of this well in solution. The 2 kcal/mol contour line in solution encompasses the region of the cellobiose and cellulose crystal structures and also extends to the postulated cello–oligomer solution structure of ϕ,ψ =45°,-38°.[36] In a sense, a crystal environment can be considered to be a "solution" which stabilises the H–conformation through hydrogen bonding interactions with

University of Cape Town

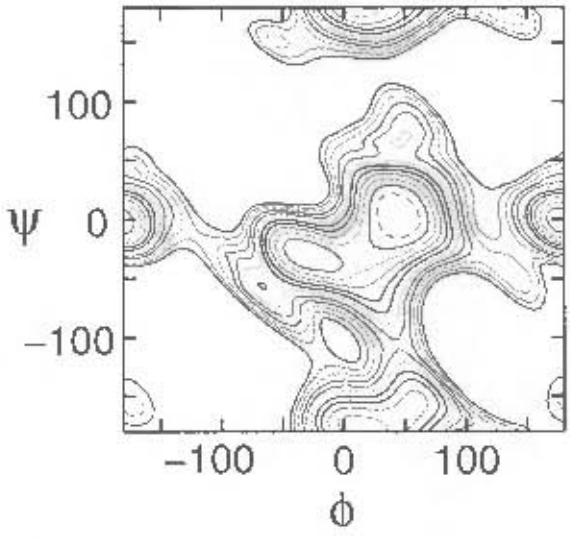


Figure 7.7: PMF ϕ, ψ energy surfaces for cellotetraose in vacuum. Contours appear at 1 kcal intervals above the global energy minimum. The first contour at 1 kcal/mol is dashed. Contours higher than 12 kcal/mol above the global minimum are not shown.

University of Cape Town

surrounding molecules.

The J–minimum has a similar morphology in solution as compared to vacuum, but is higher in energy. This leads to the conclusion that the extent of this narrow well is determined by steric clashes between the primary alcohols and the adjacent rings and not by the range of intra–molecular hydrogen bonds.

The τ PMFs (Figure 7.9) and their corresponding population distributions (Figure 7.10) calculated for cellobiose give a 1D depiction of the change in the free energy surface upon solvation. The relationship between the τ dihedral and the minima in the 2D PMF vacuum for cellobiose is illustrated in Figure 7.8. The reversal of the H and J–well energies and populations is clear from a comparison of the vacuum and solution τ PMF curves. In all cases, the population of the J–region is greater than the K–region, supporting Stevens and Sathyanarayana's assertion that the J–region is more likely as a bend conformation in cellulose.

A shift of the global energy minimum from the J–well to the H–well with concomitant broadening of the H–well is also apparent in the ϕ,ψ PMF for the central glycosidic linkage in β–cellotetraose. In addition, on this surface the K–well minimum has shifted from around ϕ,ψ =50, -160 to 15, -175 · a change which is obvious in a comparison of the τ PMFs and population distributions for cellobiose and cellotetraose shown in Figures 7.9 and 7.10 respectively. Glycosidic linkages in either exclusively J– or K–region conformations produce a semi–circular molecular configuration for cellotetraose, stabilised

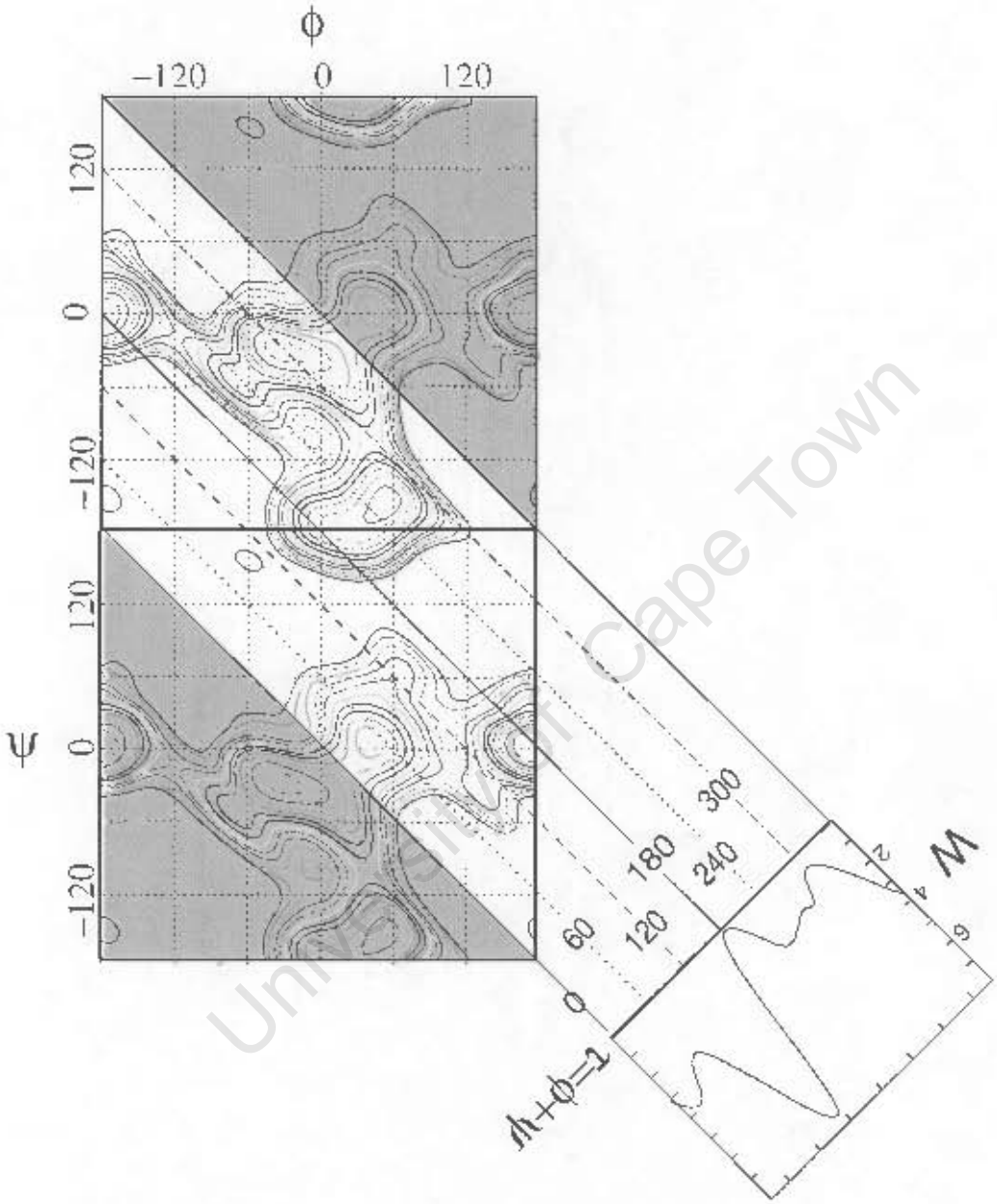


Figure 7.8: Diagram of the relationship between the β–cellobiose ϕ,ψ vacuum PMF and the τ PMF.

University of Cape Town

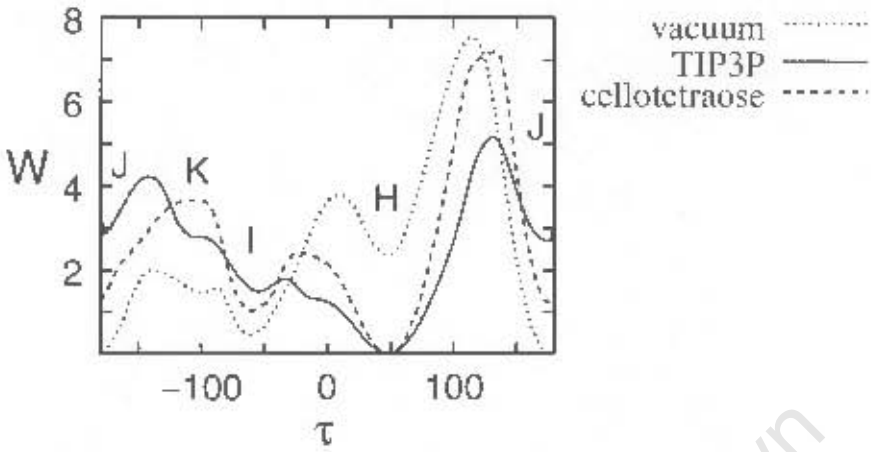


Figure 7.9: PMF surfaces for τ dihedral in β–cellobiose in vacuum and solution and cellotetraose in vacuum obtained from the ϕ,ψ populations. The graphs are shown relative to their global energy minima, with the energy in kcal/mol. Regions corresponding to the H, I, J and K minima on the ϕ,ψ maps are indicated.

University of Cape Town

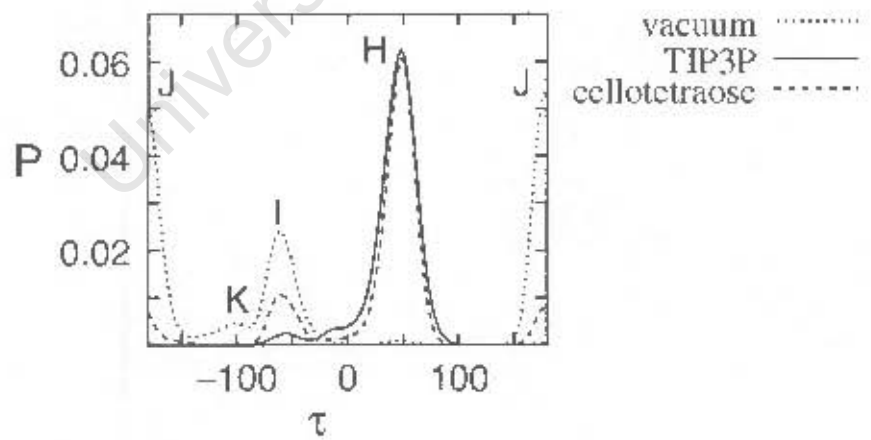


Figure 7.10: Normalised populations for τ dihedral in β–cellobiose in vacuum and solution and cellotetraose in vacuum obtained from the ϕ,ψ populations. Regions corresponding to the H, I, J and K minima on the ϕ,ψ maps are indicated.

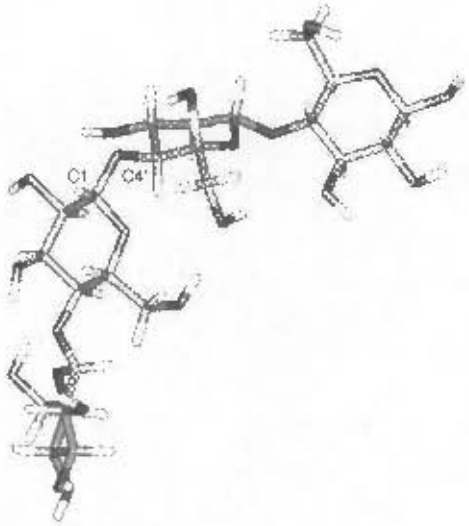


Figure 7.11: A "bent" conformation of cellotetraose, with the central linkage in a K–well conformation and the outside linkages in an H–well conformations.

University of Cape Town

by O_2–O_3 intraresidue hydrogen bonds. These are analogous to the circular $\alpha(1\rightarrow4)$–linked cyclodextrins. However, these conformations are not considered in the cellotetraose PMF. The calculation considered the central glycosidic linkage and allowed the adjacent linkages to relax from their original $\phi,\psi = 0,0$ conformations. As there was no applied umbrella potential to drive the outer linkages over an energy barrier into either the J or K– well, these remained within the general H/I region throughout the entire PMF calculation. Therefore, the J and K regions in the cellotetraose PMF correspond to "bent" conformations of the molecule, as shown in Figure 7.11. Repetition of H– and I–well conformations produces straight chains with a helical twist (Figure 7.12). The ϕ, ψ PMF in cellotetraose thus shows the relative energy of "bent" versus "straight" conformations. Staight conformations from the H–well are favoured in cellotetraose because of the absence of steric clashes and the presence of an O_3'–O_5 stabilising hydrogen bond. The range of the H–well in celloteraose accords well with a recent potential energy surface calculated for the region of a range of cello–oligomer crystal structures.[24] In addition, the energy barrier between the H– and I–wells is also lowered in cellotetraose. Judging from the cellobiose solution PMF, this effect is likely to be reinforced upon solvation.

The PMFs obtained for cellobiose and cellotetraose have implications for the structure of cellulose in vacuum and solution. If one considers only regularly repeating structures, the four minima on the cellobiose free energy surface result in four possible structures for a cellulose strand. However, although small oligomers may form circular J– or K–well conformations, steric clashes make these conformations highly improbable for longer

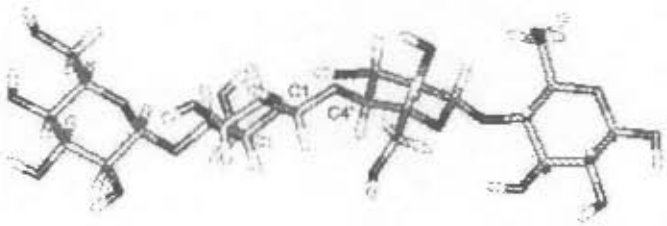


Figure 7.12: A "straight" conformation of cellotetraose, with linkage conformations taken from the H–well.

strands. In addition, all the cellobiose and cellulose crystal structures have glycosidic torsion angles in the H region, indicating no experimental evidence for J– and K–well bend conformations. However, these could potentially occur in solution.

Glycosidic linkage conformations taken exclusively from the H–well result in a ribbon structure with a more or less pronounced left–handed helical twist. The I–well results in a right handed helix. These have not been seen in the crystal structures available, but could potentially be of importance in amorphous glasses. Solvation and additional monomers favour the H well.

7.3.2 Stretching Simulation

The force–extension curve produced by the stretching simulation of β–cellododecaose appears in Figure 7.13. This is plotted on the same range as amylose to enable direct comparison of the curves. The cellulose simulation shows minimal elasticity in comparison with amylose, with a total increase in length over the course of the simulation of less than 10%. No chair to–boat transitions of pyranose rings occurred throughout the course of the simulation. The stretching curve is comparable to that produced by experiment (Figure 7.2). However, as the simulation was begun from a fairly extended structure corresponding to that in a crystal environment, the graph lacks the initial rapid extension seen in the experimental curve.

On closer examination, magnification of the curve allows for identification of a transition at just over 1000 pN (see inset in Figure 7.13). This transition is not seen in the experimental curve, either because it is so slight, or because it does not occur in solution. The transition is caused by rotation of the glycosidic linkage torsion angles from an H–well conformation (where the simulation was started) to an I well conformation, which produces a slight lengthening of the chain. Evidence for this transition is provided by the plots of the ϕ, ψ distribution during the simulation shown in Figure 7.15. Figure 7.15 (a) shows super imposed ϕ,ψ time series plots for all 11 glycosidic linkages throughout

University of Cape Town

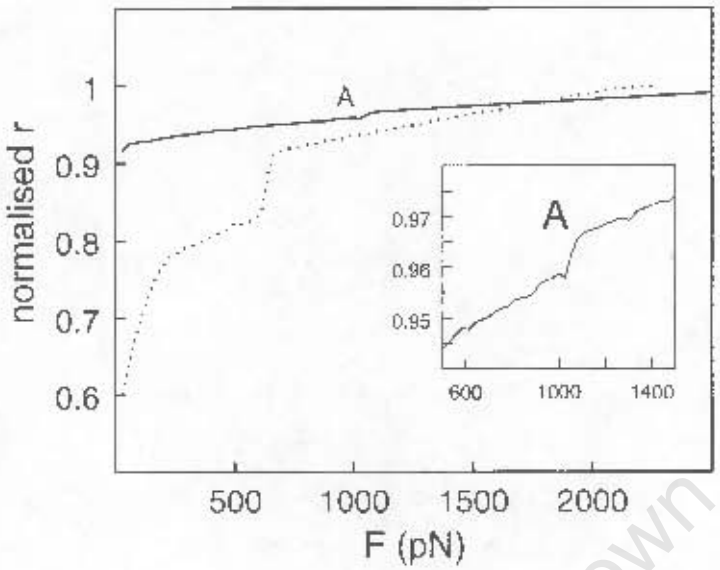


Figure 7.13: The force–extension curve produced by the cellododecaose stretching simulation (——). The end–to–end distance (r) is normalised by dividing by the final end–to–end distance at 2500 pN and is displayed on the same scale as the amylose force extension curve (shown as a dotted line). The inset shows a magnification of the cellododecaose curve at point A.

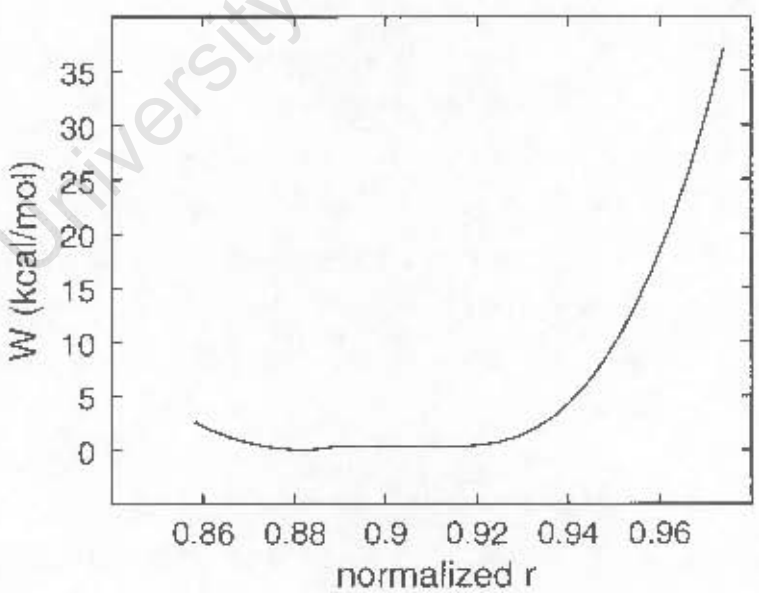


Figure 7.14: PMF calculated from the distributions produced by the first 36 steps (up to an applied stretching force of 1000 pN) of the cellododecaose stretching simulation. The end–to–end distance (r) is normalised as in Figure 7.13.

University of Cape Town

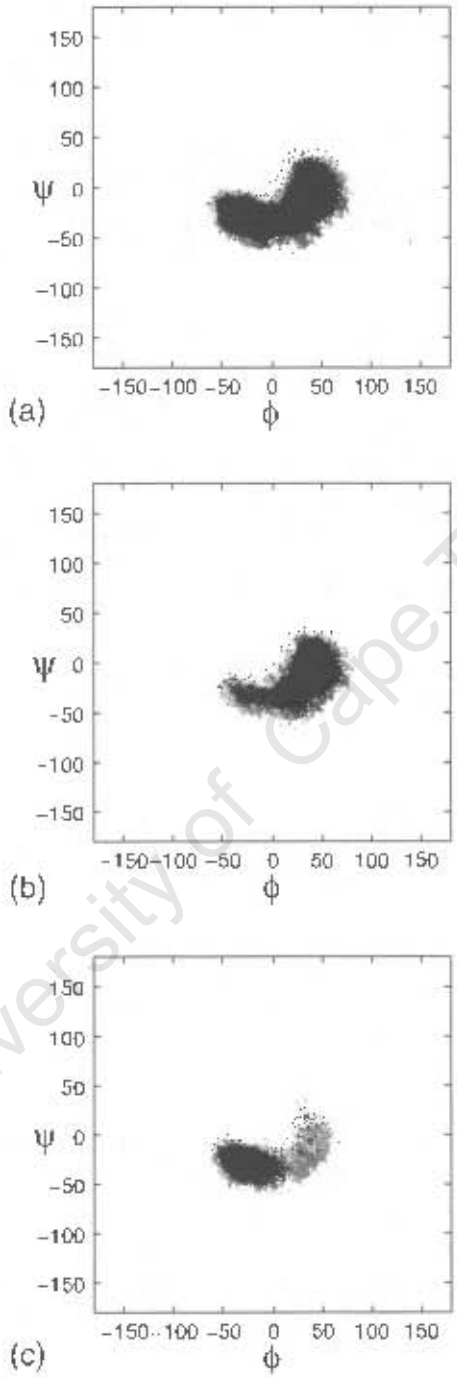


Figure 7.15: Superimposed time series of all the ϕ, ψ angles in the stretching simulation of cellododecaose for: (a) the entire simulation (b) the simulation below 850 pN (c) the simulation above 1300 pN.

University of Cape Town

the course of the run. The time series for the first part of the simulation, before the transition [Figure 7.15 (b)] shows motion limited to the H well, while after the transition it is chiefly in the I well [Figure 7.15 (c)]. This transition mechanism may have relevance to the possible molecular origin of the β relaxation processes occurring in dry, amorphous cellulose, which have been suggested to arise from localised motions of the main chain segments.[3]

It is clear from the time series distribution in Figure 7.15 that motion of the glycosidic linkages is limited to the H– and I–wells and that the J– and K–wells are not visited at all. If these wells were at all energetically favourable, it is expected that the additional energy provided to the cellulose molecule by the stretching force would result in them being explored occasionally during the course of the 500 ns simulation. This is not the case. Therefore, although the J–well and K–well may be present to a significant extent in the $\beta(1\rightarrow4)$-linked cellobiose, they seem to be highly disfavoured in the cellulose polymer itself. However, compression of the polymer is likely to make the more compact "bend" conformations more favourable.

7.4 Conclusions

The PMF calculations for β–cellobiose and β–cellotetraose showed that, while the $\phi, \psi = 180,0$ conformation is the lowest in energy for the disaccharide, the addition of either water as a solvent or adjacent residues to the chain favours the region of the cello–oligomer crystal conformations. In addition, although the ϕ, ψ PMFs in vacuum and solution clearly show that the $\beta(1\rightarrow4)$-linkage in cellobiose is more flexible than the $\alpha(1\rightarrow4)$–linkage, the range of motion is diminished for the central linkage of cellotetraose and the early stages of stretching of cellododecaose showed only the H–well to be significantly occupied. Thus, the cumulative effect of additional units is to considerably restrict the low–energy regions of the $\beta(1\rightarrow4)$ linkage, resulting in the inelasticity of the cellulose polymer. It is interesting that this effect is seen in a single isolated strand of cellulose without any of the packing effects that have been postulated to cause the strength, rigidity and immobility of cellulose.

The force–extension curve calculated from the stretching simulation of cellododecaose is similar to that produced by AFM experiments on cellulose. The simulated experiment showed a previously unidentified slight transition in the cellulose stretching curve at high forces. This feature was produced by a shift of the glycosidic linkage conformations, from the H– to the I–well, and may have implications for relaxation processes occurring in cellulose.

University of Cape Town

Chapter 8

The $\alpha(1\rightarrow1)\alpha$ and the $\alpha(1\rightarrow6)$ Glycosidic Linkages

University of Cape Town

8.1 Introduction

The methods developed in this thesis for determining the conformational free energy and physical properties of the glycosidic linkages have thus far been applied exclusively to the α and β forms of the (1→4) linkage. This penultimate chapter extends the investigation to two other glycosidic linkages: the $\alpha(1\rightarrow1)\alpha$ and the $\alpha(1\rightarrow6)$ linkages. These two linkages were chosen as they represent both the least flexible ($\alpha(1\rightarrow1)\alpha$) and most flexible ($\alpha(1\rightarrow6)$) of the glucan glycosidic linkages.

The non-reducing disaccharide α,α–trehalose (α–D Glc (1 1) α D–Glc) is produced by joining two α–D–glucose rings with a (1→1) linkage (Figure 8.1). This molecule is of interest principally because it is the most potently active cryoprotective and antidessicant disaccharide known. It is the major blood sugar in most insects and is found in large concentrations in a variety of organisms adapted to periods of cold and drought, such as Baker's yeast. The α,α–trehalose disaccharide is naturally far more abundant than its α,β–trehalose and β,β trehalose isomers and will be referred to here simply as trehalose.

There has been considerable interest in the structure of trehalose and the molecular reasons for its cryoprotective properties. This disaccharide is known to be relatively rigid, as the combination of two axial bonds together with the close proximity of the hydroxymethyl groups drastically reduces the conformational space available to its glycosidic linkage. The glycosidic linkage conformation in trehalose is described by the ϕ and ψ dihedral angles, though in this case the two angles are symmetrically equivalent. Crystal structures[20,131] and solution experiments[132,176,178] have shown trehalose to adopt

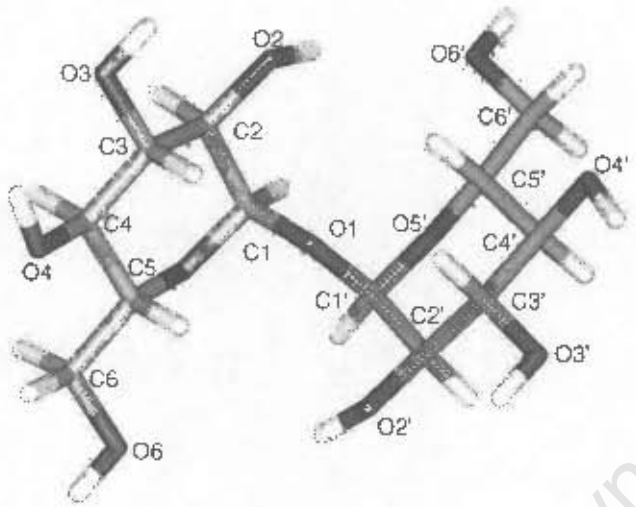


Figure 8.1: A three-dimensional structure of the α,α trehalose disaccharide, showing the atomic naming scheme.

a single conformation. This is supported by several computational studies, particularly MM adiabatic map calculations[81,143,177,295] and simulations of trehalose in aqueous solution,[78,81,144–146,177] which also indicate that this carbohydrate adopts a single conformation in both vacuum and solution.

Most simulations of trehalose in solution were performed chiefly to investigate the effects that this molecule has on the water solvent, with the aim of explaining its antidessicant properties. Analysis of trehalose simulations in water by Liu et. al found that it imposes structure on the surrounding water molecules, which extends as far as the third hydration shell.[177] Another study found that the hydration states of trehalose in the solid and liquid state were similar.[81]

The $\alpha(1\rightarrow6)$ linkage is present in the disaccharide isomaltose (α–D–Glc–(1–6)–α–D–Glc). This linkage forms the backbone of the polyglucan dextran and the branch points that distinguish amylopectin from the purely $\alpha(1\rightarrow4)$ linked amylose polysaccharide. The $\alpha(1\rightarrow6)$–linkage comprises three bonds, the additional bond being the source of its flexibility. Rotation about the three bonds is described by the ϕ_6, ψ_6 and ω_6 dihedral angles (Section 1.1.2):

$$\phi_6 = O_5 - C_1 - O_1 - C_6'$$

$$\psi_6 = C_1 - O_1 - C_6' - Ct_5$$

University of Cape Town

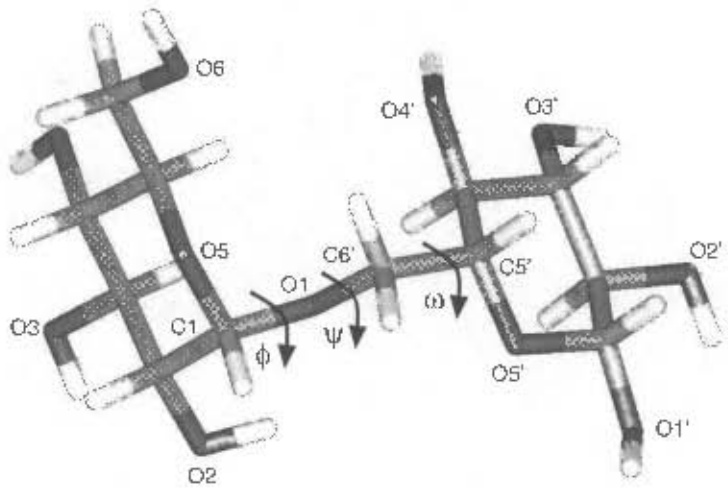


Figure 8.2: The isomaltose disaccharide, showing the atomic naming scheme and the ϕ_6, ψ_6 and ω_6 dihedral angles.

$$\omega_6 = O_1 - C_6' - C_5' - O_5'$$

Note that ϕ_6 and ψ_6 are defined slightly differently to the other ϕ and ψ dihedral angles in this thesis, in keeping with prior definitions.[57] The definition of ω^6 is essentially the same as the definition of ω^* in Equation 1.2 and this bond has the same *gg*, *gt* and *tg* minimum energy conformations found for the hydroxymethyl group in glucose. Because of the extra dimension that a third dihedral angle brings to an adiabatic map, only a few studies have attempted to chart the conformational space available to the α(1→6) linkage.[296,297]

Dielectric experiments on the α(1→6)–linked dextran polysaccharide show two sub–T_g mechanical relaxations, a γ–relaxation with an activation energy of 32 kJ/mol (7.4 kcal/mol) and a β–relaxation with activation energy 82 kJ/mol (20 kcal/mol).[3] In dextran, which has no hydroxymethyl group, the β– and γ–relaxations are attributed to motions of the main chain segments and to hydroxyl group rotations respectively.[3] The strength of the β–relaxation of dextran is affected by water, but to a lesser extent than in cellulose.

AFM stretching experiments on dextran in solution show a characteristic discontinuity near 1000 pN, as shown in Figure 8.3.[51,53,121] Simulations of the stretching of a five–unit segment of dextran in water found the main contribution to elasticity at low forces to stem from a twist of the C_5 C_6 bond that results in elongation of the filaments, with the discontinuity caused by an (unspecified) flip of the C_5–C_6 bond.[51] The elasticity of the dextran filaments after the transition was proposed to be dominated by the bending of bond angles. However, Marszalek et al. recently suggested that transition of the C_5–C_6

University of Cape Town

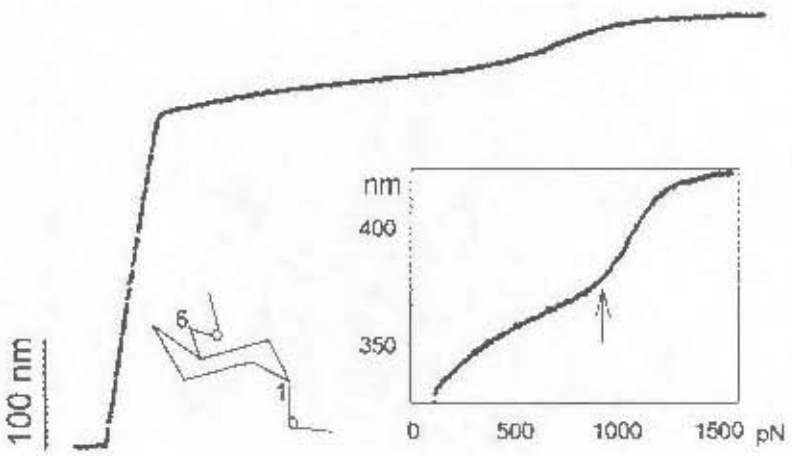


Figure 8.3: The force–extension curve for dextran, reproduced from Marszalek et al.[53] The insets show the α–D–glycopyranose ring and the transition region at high resolution.

University of Cape Town

bond (from either a *gg* or *gt* orientation to *tg*) occurs earlier in the curve, in the range of 200 to 400 pN, and that the transition observed at 1000 pN is due to a chair to boat transformation of the glucose residues.[53] They proposed the final high energy region to be dominated by the bending of bond angles.

In this chapter, the free energy surfaces for the glycosidic linkages in the trehalose and isomaltose disaccharides are explored. Two–dimensional ϕ, ψ PMFs for trehalose were calculated in vacuum and solution in order to establish the minimum energy conformation of this molecule and the range of low–energy conformations available to it. As in the case of maltose and cellobiose, 1D PMFs of the τ pseudo dihedral angle in vacuum and solution were calculated from the 2D ϕ,ψ PMFs. The conformational freedom of the isomaltose disaccharide was determined using a novel 2D PMF calculated for the $\alpha(1\rightarrow6)$–linkage in vacuum. Though valuable, three–dimensional maps remain difficult to visualise and the free energy surface is much easier to picture if collapsed into two dimensions. This is possible if the ϕ_6 dihedral angle is ignored and the PMF calculated simply as a function of ψ_6 and ω_6. This measure can be justified by the fact that the ϕ_6 is known to be inflexible due to the *exo* anomeric effect: of the seven energy minima previously identified, the lowest six had the same ϕ_6 value of 80° (see Table 8.1). This PMF calculation for isomaltose was performed in vacuum only because of the long simulations times required to average over all the low–energy ϕ_6 conformations. Finally, simulations of the stretching of dextran were performed in order to clarify which molecular events cause the transition in the measured force–extension curve.

8.2 Methods

8.3 2D ϕ, ψ PMF Calculations

The 2D ϕ, ψ PMF for trehalose in vacuum and solution and the 2D ψ_6, ω_6 PMF for isomaltose in vacuum were calculated using the iterative adaptive umbrella sampling method described in Section 3.5.

In each case, the 360°×360° umbrella potential surface for the two glycosidic linkage dihedral angles was represented as a two-dimensional grid of points, with a grid separation of 2.5°. At each integration step of a simulation, the biasing potential energy to be applied to the current ϕ, ψ position was calculated from a cubic spline of the umbrella energy surface. After each simulation, a two-dimensional ϕ, ψ distribution histogram was calculated by summing the number of configurations in each 2.5°×2.5° bin over the production phase of the trajectory. The WHAM procedure was then applied, using a tolerance value of 0.001, to obtain an optimum combination of the all the simulations distributions and thus the next estimate of the PMF. This PMF surface was extrapolated into unexplored regions by setting unsampled bins to the maximum value of the sampled bins, treating the border regions as described in Section 3.4. The resultant PMF surface was smoothed three times using Equation 3.9. Subsequently, regions in the 2D ϕ, ψ PMF associated with high-energy steric clashes of the atoms were removed. This was done by setting any bin in the PMF with an energy value greater than a predetermined cut-off of 20 kcal/mol to the cutoff value. The umbrella potential for the next simulation was set to the negative of this PMF estimate.

Each umbrella simulation began with an equilibrium phase of 500 ps, followed by a production phase ranging from 2 to 20 ns in length. Only the production phase was used to calculate the simulation distribution. Successive simulations in a series were gradually run for longer periods as the PMF estimates improved and more of the phase space was explored. The first simulation in each series was started from a minimised and equilibrated coordinate set and performed with a flat umbrella potential surface. Subsequent simulations were begun from the final conformation of the previous simulation, to facilitate gradual exploration of the whole of conformational space. After every 8 simulations, the next simulation was begun from the original equilibrated starting configuration. Convergence of the whole adaptive umbrella sampling procedure was defined to occur when, for a single simulation, every bin k in $n_{j,k}$ in the ϕ, ψ range of interest was occupied at least once. Up to 200 simulations were used to obtain the final PMF, but at any time not more than the latest 50 distributions were used to estimate the next PMF; earlier

University of Cape Town

simulations were gradually discarded as new, more extensive distributions were produced. A total of approximately 500 ns of simulation time was required to produce each of the final PMF surfaces.

The 1D τ PMF for trehalose and the ω_6 PMF for isomaltose were calculated as described in Chapter 5 by summing the populations obtained from the corresponding 2D ϕ, ψ PMFs and applying Equation 3.1.

8.4 Oligosaccharide Stretching Simulations

A 12–unit strand was found to be of sufficient length for the dextran stretching simulations. Six conformations of a β dextradodecaose strand were stretched by applying equal, but opposite, forces to the terminal O_1 and O_6 atoms at the reducing and non–reducing ends of the chain respectively. The force was applied using the method outlined in Chapter 3. The simulations were performed in vacuum.

To investigate the effect of the starting structure on the final force–extension curve, six stretching simulations were performed in vacuum with different starting structures. Six initial starting structures were built, each using dihedral angle conformations taken exclusively from one of the six minimum energy wells previously identified.[57] This produced six different regularly repeating conformations, which are described in Table 8.1.

The simulations were all begun with a non–zero force of 55.58 pN, to prevent the molecules from collapsing in the vacuum. During the course of the simulations, the stretching protocol involved ramping the force by 27.79 pN (0.4 kcal.mol^{-1}.Å^{-1}) every 5 ns, up to a maximum force of 3000 pN. This corresponds to a ramping rate of 5.58 pN/ns. AS was the case with the cellulose oligosaccharide, the dextran fragments were found to require much less equilibration time than was necessary for the 18-unit amylose strand to achieve curves comparable with experiment. Each 5 ns simulation step consisted of 1 ns of equilibration, followed by 5 ns of data collection. A total of 100 ramping steps were applied during the course of each 500 ns stretching simulation. Force–extension relationships for the various simulations were obtained by plotting the average extension during each production step against the applied force. The curves were normalised by dividing the length of the molecule by the maximum extension achieved by the end of the simulation (76 Å) and the resultant stretching curves compared to each other.

University of Cape Town

8.4.1 Simulation Conditions

Molecular dynamics simulations were performed using the program CHARMM[222] (version 27b1), with modifications incorporated into the USERE module in order to implement both the two-dimensional adaptive umbrella sampling PMF calculations and the stretching simulations. The CSFF carbohydrate parameter set (Chapter 4) for the CHARMM force field was used to represent the carbohydrate molecules in all the simulations. The TIP3P model was used to represent water.

For the solution simulations, the trehalose disaccharide was placed in a previously equilibrated cube of 512 TIP3P waters. Solvent water molecules that overlapped with the solute molecule were removed and the system was equilibrated for 500 ps. The solution simulation surrounded the trehalose disaccharide with 489 TIP3P water molecules[248] in a cube of length 24.64 Å. This gives a density of 1.013 g/cm^3. The cube was subjected to minimum image periodic boundary conditions to eliminate edge effects.

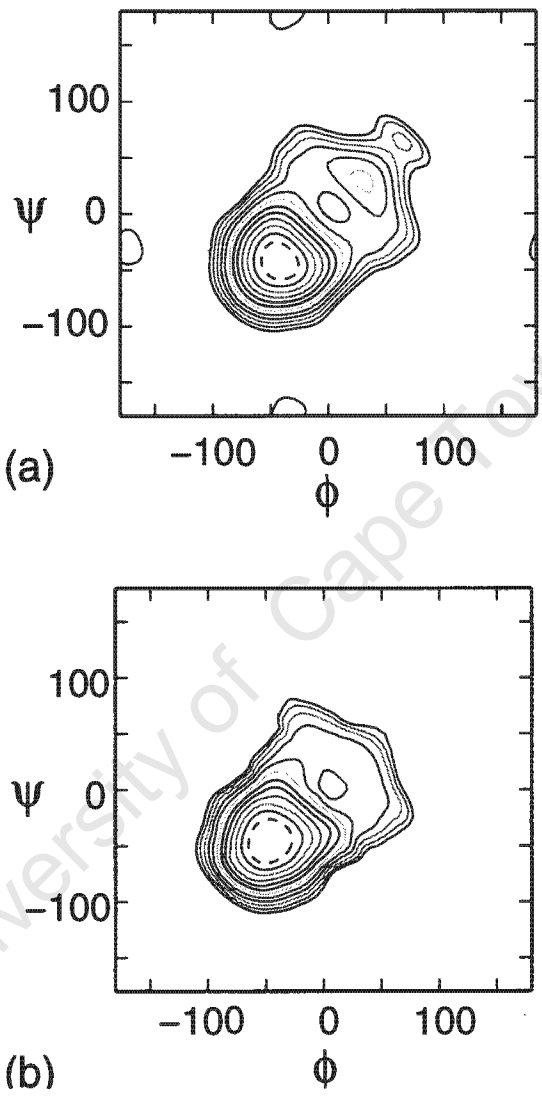
Initial velocities for the atoms were selected at random from a Boltzmann distribution at 300 K. All simulations were performed in the canonical ensemble (constant n, V, T), using stochastic Langevin dynamics with a frictional coefficient of 62.5 to maintain a constant temperature of 300 K. The equations of motion were integrated using a Leap-Frog Verlet integrator[256] with a step size of 1 fs. The SHAKE algorithm[258] was used to fix the length of bonds involving hydrogen atoms and the water molecule geometry throughout each simulation. Non-bonded interactions were truncated using a switching function applied on a neutral group basis between 10.0 and 12.0 Å. The groups corresponded to electrically neutral collections of atoms in the carbohydrate molecules and entire water molecules for the solvent.

University of Cape Town

8.5 Results

8.5.1 The α(1→1)α linkage

The vacuum PMF for trehalose, shown in Figure 8.4 (a), has a symmetrical shape similar to published adiabatic maps, the diagonal symmetry of the map being a consequence of the molecular symmetry. The global energy minimum is situated at ϕ, ψ = -40.0°,-42.5° on the vacuum map and ϕ, ψ = -45.0°,-45.0° on the solution map. Crystal structures of trehalose show an asymmetric extended conformation with ϕ, ψ = -58°,-45°; which is stabilised by an O_6–O'_2 hydrogen bond with the primary alcohols in *gt* conformations.[20,131] Experiments in solution have revealed similar conformations: optical rotation experiments show ϕ, ψ = -60°,-60° ±5°[132,176] and NMR proton–carbon coupling measurements have



University of Cape Town

Figure 8.4: Contoured ϕ,ψ free energy surfaces for trehalose (a) in vacuum and (b) TIP3P water solution. In each case, contours appear at 1 kcal/mol intervals above the global energy minimum, to a maximum of 12 kcal/mol. Contours higher than this are not shown. The 1 kcal/mol contour is shown as a dashed line.

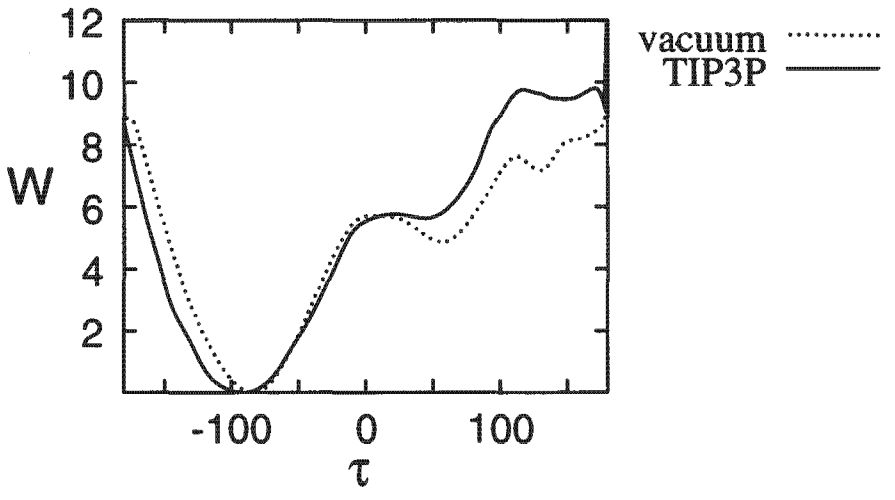


Figure 8.5: PMF surfaces for the α,α–trehalose τ dihedral in vacuum and solution obtained from the ϕ,ψ populations. The graphs are shown relative to their global energy minima, with the energy in kcal/mol.

ϕ, ψ = -41°,-41° ±5°.[178] Therefore, the global minima for both the vacuum and solution PMFs are in the region of the crystal and solution structures. Adiabatic maps calculated for trehalose using the MM3[143] and MM2CARB[295] force fields have also identified this region as the global energy minimum. However, both of the previously published adiabatic map calculations using CHARMM force fields[81,177] contradict experimental data in placing the global minimum energy configuration near ϕ, ψ = 60°,60°. In particular, the minimum energy structure identified using the HGFB force field had the primary alcohols in the *tg* conformation.[177] As Chapter 4 showed the HGFB force field to favour the *tg* conformation, this suggests that accurate parameterisation of the primary alcohol is necessary for identification of the correct global energy minimum in trehalose.

The solution and vacuum PMFs for trehalose are remarkably similar. In contrast to the (1→4)-linked disaccharides discussed in previous chapters, the glycosidic linkage conformation of trehalose is largely unaffected by the surrounding water molecules. The similarities between the solution and vacuum free energy surfaces for trehalose can be easily seen in a comparison of the 1D τ PMFs for solution and vacuum shown in Figure 8.5. Solvation increases the preference for the central energy region by slightly raising the energy of the high valleys at ϕ, ψ = 30°, 30° and ϕ, ψ = 65°, 65°. These more compact conformations are sterically strained and their stabilising interresidue hydrogen bonds are disrupted by the aqueous solution. Trehalose is thus quite rigid in water. This results in the location of the oxygen atoms being essentially fixed in relation to one another, which may result in the structure that MD simulations have shown trehalose to impose on the surrounding water solution.

University of Cape Town

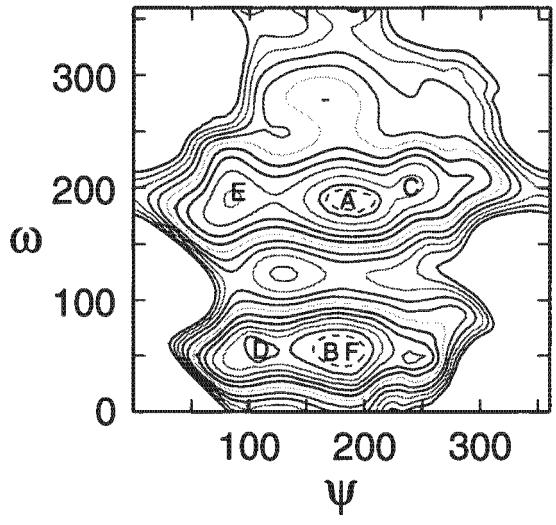


Figure 8.6: Contoured 2D ψ_6, ω_6 free energy surface for isomaltose in vacuum. Contours appear at 1 kcal/mol intervals above the global energy minimum, to a maximum of 12 kcal/mol. Contours higher than this are not shown. The location of the six previously identified minimum energy conformations[57] are marked A to F on the map.

University of Cape Town

8.5.2 The $\alpha(1\rightarrow6)$ linkage

The 2D ψ_6, ω_6 PMF for isomaltose is shown in Figure 8.6. This figure has been shifted from the usual range of -180° to 180° to a 0° to 360° range, so that the minimum regions fall in the centre of the map. A recent four–dimensional adiabatic map conformation analysis of the $\alpha(1\rightarrow6)$ linkage identified six low–energy conformations for isomaltose.[57] These are listed in Table 8.1 and are marked on the PMF surface. The energy surface has two minimum energy regions, which are fairly flexible in the ψ_6 dimension, separated by an 8 kcal/mol energy barrier. The relative energies of the three ω_6 minima and the energy barriers between them are more easily seen in the 1D ω_6 PMF, shown in Figure 8.7. One region is at ω_6 = 55° (corresponding to a *gg* conformation) and the other at ω_6 = 180° (corresponding to a *gt* conformation). The global minimum occurs at ψ_6, ω_6 = 180°, 55°, though this is only very slightly higher than the *gt* minimum. The barrier heights between the minima in the 1D ω_6 PMF are between 7 and 10 kcal/mol. This is close to the activation energy of the γ mechanical relaxation measured in dextran (7.4 kcal/mol).[3] Therefore, this relaxation could be caused by rotations of the ω_6 dihedral and not the suggested hydroxyl group rotations. The stronger β–relaxation (20 kcal/mol) could possibly be related to rotation of the less flexible ϕ_6 or ψ_6 dihedral angles.

When repeated, the six minimum energy glycosidic linkage conformations identified for the isomaltose $\alpha(1\rightarrow6)$ linkage produce quite different structures for the 12–unit dex-

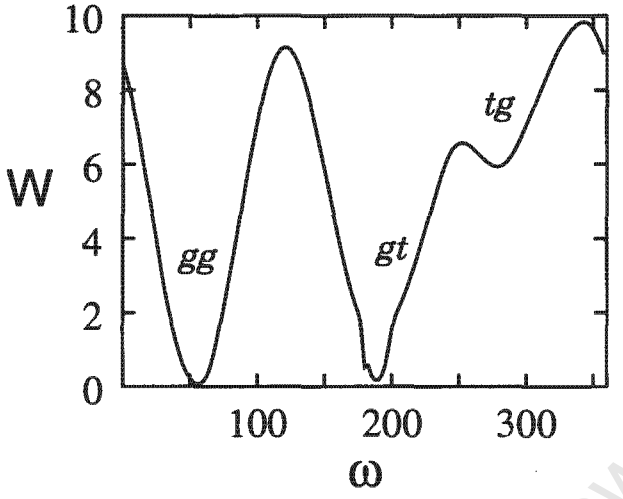


Figure 8.7: A 1D vacuum PMF surface for the ω_6 dihedral angle in isomaltose, obtained from the ψ_6, ω_6 populations. The graph is shown relative to the global energy minimum, with the energy in kcal/mol.

University of Cape Town

Dihedral angles			d (Å)	r (Å)	Description	Designation
ϕ_6	ψ_6	ω_6				
80	180	180 (*gt*)	6.13	42	flattened, broad helix	A
80	180	60 (*gg*)	8.09	52	kinked strand	B
80	-120	200 (*gt*)	2.80	25	compressed, angular helix	C
80	100	40 (*gg*)	8.26	32	broad helix	D
80	80	200 (*gt*)	9.13	50	twisted ribbon	E
140	180	60 (*gg*)	8.51	57	stretched rope	F

Table 8.1: The six minimum energy conformations previously identified for isomaltose (A–F) with their corresponding $O_1 \rightarrow O_6$ distances in each residue (d). Measured $O_1 \rightarrow O_6$ end-to-end distances (r) for 12–unit dextradodecaose molecules built with all linkages in each conformation are included together with a description of the resultant molecular conformation.

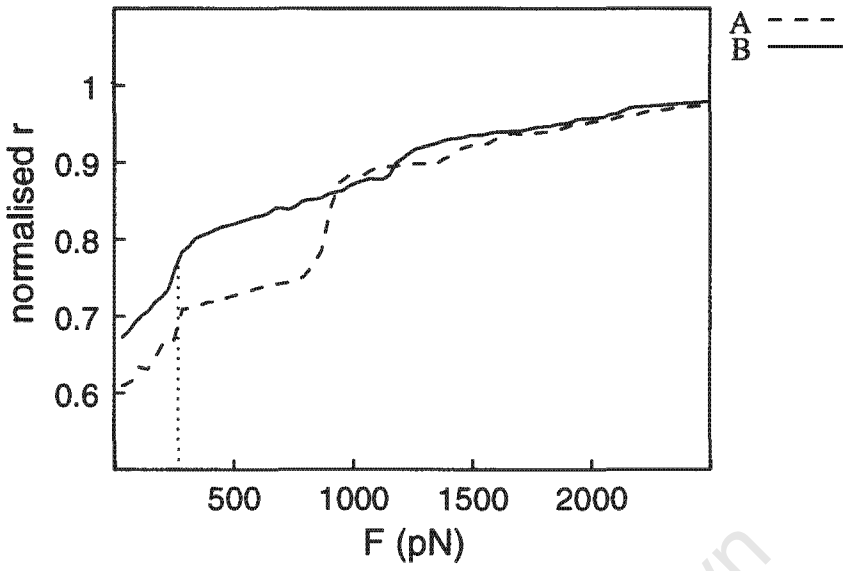
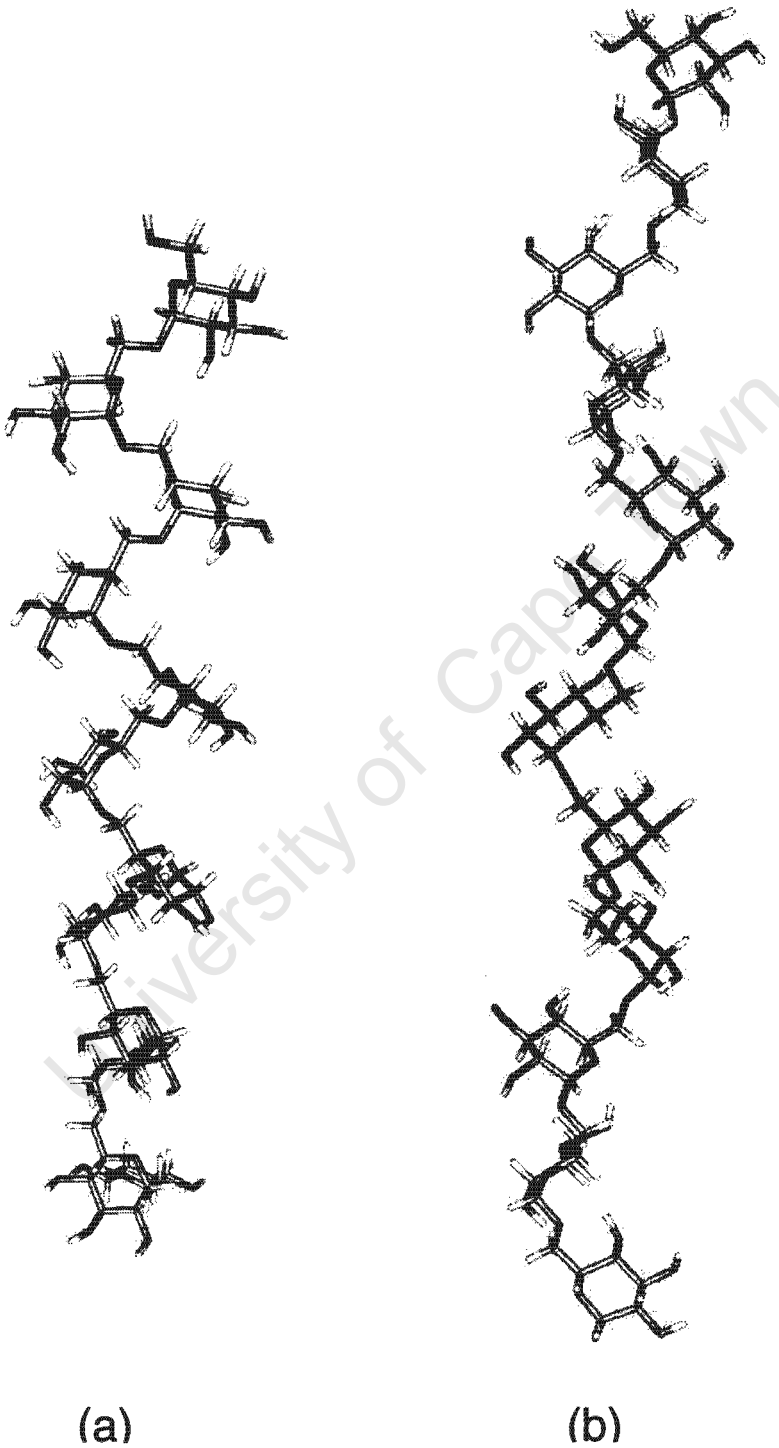


Figure 8.8: Calculated force–extension curves for dextran conformations from the A (– – –) and B (——) wells. Curves are normalised by dividing by the maximum extension achieved by the end of the simulation (=76 Å). The dotted line shows the estimated zero point for a simulation in aqueous solution.

University of Cape Town

tradodecaose strand. These structures are described and their approximate end–to–end distances listed in Table 8.1. Stretching of the six different starting structures resulted in two general types of curve — one with three identifiable domains and the other with four. Examples of each are shown in Figure 8.8. The curve marked 'A' was generated from an A–well starting conformation (as shown in Figure 8.9 (a)). However, similar curves were produced from all starting structures in which the ω dihedral was in the *gt* conformation (i.e. the A, C and E conformations). Similarly, the curve marked 'B', though produced from a B–well starting structure (Figure 8.9 (b)), corresponds to starting structures with a *gg* conformation (i.e. B, D and F conformations). The B curve has a small transition at approximately 1200 pN, whereas the A curve shows a dramatic transition at about 900 pN and another smaller one at 1400 pN. The force required to bring about a transition in the B–curve is 200 pN higher than in the measured experimental curve, which is about the same discrepancy as found for amylose in Chapter 6. As discussed in Chapter 6, this can be attributed to the faster ramping rate used in the simulation as compared to experiment, the fact that the simulation is performed in vacuum and the possiblility that the curves are not converged due to sampling limitations. Nevertheless, the B–curve matches quite closely with the experimental profile shown in Figure 8.3. This implies that the structure of dextran in solution falls within the B, D and F structures. This is in agreement with the results for the hydroxymethyl group discussed in Chapter 4, where



University of Cape Town

Figure 8.9: Dextradodecaose in an (a) A–well conformation and (b) B–well conformation.

solvation was shown to favour the *gg* conformation. In addition, the B–structure itself is the most extended of the minimum energy structures and Chapter 6 showed solution to favour more extended conformations of amylose. If, assuming that the dodecadextran strand would elongate rapidly at low force in solution, the initial force required to extend the molecule is subtracted from the B–curve ($\approx$ 250 pN, as shown by the dotted line in Figure 8.8), the force at transition corresponds quite closely to experiment.

The discontinuity in the experimental curve at 1000 pN has recently been identified as being caused by chair–to–boat transitions of the pyranose rings in dextran.[53] However, analysis of the dynamics simulations contradicts this finding: the 1200 pN transition in the B–curve for dextran was instead found to be associated with rotation of the C_5–C_6 dihedral angles from the *gg* to the *tg* conformation. In the case of the A–curve, the 900 pN discontinuity was associated with an ω_6 rotation from the *gt* to the *gg* conformation and that at 1200 pN with another rotation to the final *tg* conformation. In each case, chair–to–boat transitions began only *after* the linkages were in their maximum extension *tg* orientation and continued in an anti–cooperative fashion until completion at 3000 pN.

Rotation of the ω_6 dihedral to the *tg* conformation occurs because this is the most extended conformation for the $\alpha(1\rightarrow6)$–linkage and thus temporarily relieves strain on the molecular bonds and angles. As *tg* is the most disfavoured of the three ω_6 minima and has a large energy barrier separating it from either of the other two low–energy conformations, this transition only occurs under force. In the case of the A–curve, the *gt* to *gg* rotation occurs for similar reasons; the *gg* orientation is more extended than the *gt* conformation, but force is required to overcome the energy barrier between them. The force required to bring about the *gg* to *gg* rotation could be lowered in solution: previous studies have shown isomaltose to explore a broader region of conformational space in solution than in vacuum.[57]

University of Cape Town

8.6 Conclusions

The global energy minimum found for the trehalose disaccharide has the same location in vacuum as in solution. Encouraging support for the correct parameterisation of the hydroxymethyl dihedral angles is given by the fact that this is the first CHARMM force field to produce a minimum energy conformation in agreement with crystal structures and NMR solution experiments. Comparison of the solution and vacuum PMFs showed that the conformation of trehalose is largely unaffected by water, in contrast to the $(1\rightarrow4)$ disaccharides investigated in this thesis. Trehalose also has a very limited range of motion about the minimum energy conformation in both vacuum and solution. This rigidity in

solution may provide a clue as to the origin of trehalose's marked cryoprotective properties. An interesting point is that the flexibility of the three disaccharides investigated in solution in this work increases (trehalose < maltose < cellobiose) as their cryoprotective abilities decrease (trehalose > maltose > cellobiose).

The 2D PMF for the isomaltose disaccharide gives a view of the energy barriers between the three ω_6 low energy conformations. On the basis of height of these barriers, it is suggested that the γ–relaxation measured for dextran could be caused by rotations of the ω_6 dihedral and not the proposed hydroxyl group rotations. In addition, the isomaltose PMF surface was found to be valuable for interpreting the glycosidic linkage motions that occurred in the stretching simulations of a dextradodecaose strand. Force–extension curves showing a profile similar to the experimental curve were produced from starting structures with the ω_6 dihedral in the *gg* conformation, but not from conformations with ω_6 dihedral in the *gt* conformation. This implies that the *gg* orientation predominates in solution. The transition in the dextran curve was found to be caused by a rotation of the C_5-C_6 bond, in agreement with the original study of Rief et al. Moreover, the C_5-C_6 rotation was determined to be specifically associated with transition of the ω_6 torsion angles from *gg* to *tg* conformations. Though energetically unfavourable under normal conditions, the *tg* orientation of this bond leads to the most extended overall conformation of the dextradodecaose molecule. Anticooperative chair-to-boat transitions of the pyranose rings only occurred at very high stretching forces. Therefore the force extension profile of dextran is determined chiefly by the rotational properties of the α(1→6) glycosidic linkage. This study thus further demonstrates the power of free energy simulations used in conjunction with stretching simulations for identifying the molecular events that determine macromolecular behaviour and hence physical properties. In addition, the reproduction of the experimental transition at approximately the experimental force provides further support for the correct parameterisation of the primary alcohol dihedral angle.

University of Cape Town

University of Cape Town

Chapter 9

Final Conclusions and Future Work

University of Cape Town

In this thesis, free energy calculations using an adaptive umbrella sampling method were developed and employed to establish the conformational freedom of selected carbohydrates. Two principal types of potential of mean force calculation were used: free energy calculations for rotations about dihedral angles (both one– and two–dimensional) and end–to–end distance free energy calculations for the extension and compression of oligosaccharide strands. In addition, novel one-dimensional representations of two–bond glycosidic linkages (and two-dimensional representations of three–bond glycosidic linkages) and a new method for simulating the AFM stretching of single molecules were developed.

One-dimensional PMFs for rotation about the hydroxymethyl group in β-glucose and β-galactose were used to validate the reparameterisation of the hydroxymethyl group. These PMFs confirm that the adjusted force field exhibits equilibrium primary alcohol populations as well as rotational frequencies for glucose and galactose that are consistent with experimental evidence. Normal mode calculations on α-D-glucose also showed the force field to be satisfactory. Simulations of atomic force microscopy stretching experiments on amylose, dextran and cellulose produced force-extension curves remarkably similar to those obtained from AFM, further indicating the reliability of the force field. This is particularly encouraging in the case of the dextran polymer, which has a $\alpha(1\rightarrow6)$-linkage through the hydroxymethyl group. Further corroboration was obtained from the trehalose PMF calculation: this is the first CHARMM force field to correctly identify the minimum energy structure, which requires the hydroxymethyl group to be in the *gg* conformation.

The potential of mean force calculations provided a valuable characterisation of the conformational space available to the $\alpha(1\rightarrow4)$-, $\beta(1\rightarrow4)$- and $\alpha(1\rightarrow1)$ glycosidic linkages in both vacuum and solution and the $\alpha(1\rightarrow6)$ linkage (in vacuum). These PMF calculations

are the first complete ϕ,ψ PMF calculations to be performed on disaccharides in either solution or vacuum. This achievement was made possible by several factors. Firstly, the modification of the carbohydrate force field resulted in faster equilibration of the primary and secondary alcohols. Secondly, the increased speed of parallel computer cluster used in this study allowed for considerably longer simulation times than were previously possible. Together, these two factors enabled the simulations to achieve the thorough sampling of the primary and secondary hydroxyls that is necessary for generating representative conformational distributions for the PMF calculations. Lastly, the development of a successful adaptive umbrella sampling protocol, including procedures for extrapolation, smoothing and identification of possible pitfalls, resulted in faster convergence of the PMF calculations.

Though time-consuming to calculate, two–dimensional glycosidic linkage PMFs in solution are particularly useful in that they can be compared with their corresponding vacuum PMFs in order to quantify the effect of solvent on the conformational preferences and the rate of conformational transitions of the linkage. Using this method, the free energy surfaces (and thus the conformation and dynamics) of the $\alpha(1\rightarrow4)$– and $\beta(1\rightarrow4)$–linkages were both found to be profoundly affected by solvent. However, this was not the case for the $\alpha(1\rightarrow1)\alpha$–linkage in the disaccharide trehalose. The fact that this molecule is remarkably unaffected by an aqueous solvent may provide a clue as to its cryoprotectant properties.

The primary focus of this study was the $\alpha(1\rightarrow4)$-linkage. Comparison of the PMFs for the glycosidic linkage in maltose and dixylose showed that the hydroxymethyl group of the pyranose ring has a dramatic effect on the preferred conformations and dynamics of the $\alpha(1\rightarrow4)$-linkage. In addition, calculation of the maltose PMFs in vacuum and solution allowed for extrapolation to the effects of solvent on the conformation and stretching dynamics of the amylose polysaccharide. The proposed amylose conformation is in agreement with previous suggestions that amylose has a solution conformation composed of longish helical segments interspersed with non-helical segments. Amylose is expected to display greater flexibility in solution and also an increased probability of ψ dihedral angle transitions. The dynamics behaviour of amylose was further investigated by calculation of the the end–to–end distance PMF for a maltohexaose oligosaccharide. This indeed showed ψ transitions to have a marked effect on the dynamics and preferred conformation of the strand. Transition of the ψ dihedral angles from a *syn* to an *anti* conformation alters the maltohexaose chain conformation from a helix to a ribbon. In addition, an interesting change occurs in the relative stabilities of the maltohexaose conformations, depending on the extension of the maltohexaose molecule. At short extensions, the helical confor-

University of Cape Town

mations are lower in energy than the ribbon conformations. However, as the molecule is extended, the situation reverses and ribbon conformations predominate. Then, at extreme extensions, helical conformations are favoured once more. This mechanism was further illustrated by the simulation of the stretching of a 18–unit maltose oligomer. A similar helix–ribbon–helix transition occurred during the stretching simulation and was identified as being primarily responsible for the elastic response of the amylose polysaccharide to a stretching force. The key finding of this study of the $\alpha(1\rightarrow4)$-linkage is that ψ transitions are more important to both the conformation and the dynamics of the amylose polymer than is currently thought.

The $\beta(1\rightarrow4)$–linkage in cellobiose was found to be more flexible than the $\alpha(1\rightarrow4)$–linkage in maltose. However, novel two-dimensional PMFs calculated in vacuum for the central linkage in a tetramer showed the conformational freedom of the $\alpha(1\rightarrow4)$–linkage to be considerably less affected by the presence of adjacent glucose units that the $\beta(1\rightarrow4)$–linkage. The maltohexaose and maltose PMFs in vacuum were similar, leading to the conclusion that the maltose molecule is a good model for the $\alpha(1\rightarrow4)$-linkage in oligo- and polysaccharides. However, for the $\beta(1\rightarrow4)$–linkage, lengthening of the oligosaccharide chain reduces the flexibility of the linkage and its preferred conformation. Therefore, unlike maltose, the cellobiose disaccharide was found to be a poor model for the $\beta(1\rightarrow4)$–linkage in an oligo– or polysaccharide.

The physical behaviour of three oligosaccharides (fragments of maltose, cellulose and dextran) when subjected to a stretching force was investigated by simulations that reproduce the AFM experimental curves. These simulations provided insight into the molecular behaviour of the polysaccharides subjected to tensile stress in atomistic detail unobtainable with current experimental techniques. Moreover, stretching simulations of oligosaccharides employed in conjunction with free energy calculations of the component glycosidic linkages were shown to be a powerful method for evaluating the preferred conformations, dynamics and physical response of polysaccharide chains. For all three of the oligomers studied, the free energy surface of the glucan linkage dimers was found to be essential for rationalising the dynamic response of the oligosaccharide chains.

Overall, chair-to-boat transitions of the glucose rings were shown to play a smaller role in determining polysaccharide elastic responses than was previously thought: the chain dynamics of the three polysaccharide investigated - amylose, cellulose and dextran - were are all dominated by glycosidic linkage transitions at forces below 1000pN. The characteristic transition in the amylose and dextran curves, instead of being caused by the postulated chair-to-boat transitions, were in both cases produced by glycosidic linkage transitions between minimum energy wells. Chair-to-boat transitions did occur in amylose

University of Cape Town

and dextran, but at higher forces than has been suggested. Therefore, the elasticity of the polysaccharides was found to be chiefly determined by the glycosidic linkages.

Because of the broad scope of carbohydrate research, there are a multitude of avenues available for future applications of the methodology developed in this thesis. One obvious extension is the characterisation of other glycosidic linkages, such as the biologically-important (1→3) and (1→2) linkages, and their associated polymers. The differences between the axial and equatorial versions of other linkages can also be studied. For example, a comparison could be made of the $\alpha(1\rightarrow6)$-linkage properties of isomaltose with the β–linkage in the disaccharide gentiobiose. Further stretching experiments could be performed on the corresponding oligosaccharides to analyse their response to an applied force. An example of a molecule to be investigate in this way is the $\alpha(1\rightarrow4)$-linked galactan, pectin, which shows two transitions in its force–extension curve. In addition, the dynamic behaviour of heteropolysaccharides containing more than one type of linkage could also be investigated.

The effect of solution on oligosaccharides could be explored more thoroughly by solution simulations. However, free energy calculations remain computationally expensive and the application of the PMF calculations developed here to larger solvated systems is dependent on access to sufficiently large parallel systems. Alternatively, another possibility is to experiment with free energy calculations using a generalised solvent models to see how effectively they reproduce the solution PMF.

One aspect of this study is that knowledge of the general conformational space for a glycosidic linkage should assist in producing reduced models of carbohydrates, analogous to those developed for peptides to investigate protein folding. The individual atoms and bonds of a glucose ring could be replaced with rigid beads joined by flexible linkages, rotation about which would be described by the calculated PMF surfaces for the particular glycosidic linkage. A suitable solution PMF would allow studies of this kind to approximate the effect of solvent in a computationally inexpensive manner.

This thesis has shown that conformation and dynamic motion of the constituent glycosidic linkages to a large extent determines the conformation, flexibility and dynamic response of a polysaccharide. Therefore, knowledge of the free energy surface of the glycosidic linkages and oligosaccharide strands is the key to understanding the micro– and macroscopic properties of polysaccharides. Free energy calculations provide valuable insight into the molecular details of dynamics motion and polymer properties and may

University of Cape Town

ultimately allow for the identification of the molecular changes required to produce a "designer" molecule with particular desirable properties for the food, biomedical, textile or paper industries.

University of Cape Town

University of Cape Town

Bibliography

[1] H. Bizot, P. Le Bail, B. Leroux, J. Davy, P. Roger, and A. Buleon. Calorimetric evaluation of the glass transition in hydrated, linear and branched polyanhydroglucose compounds. *Carbohydr. Polym.*, 32:33–50, 1997.

[2] T. R. Noel, R. Parker, S. M. Ring, and S. G. Ring. A calorimetric study of structural relaxation in a maltose glass. *Carbohydr. Res.*, 319:166–171, 1999.

[3] H. Montès and J. Y. Cavaillé. Secondary relaxations in dried amorphous cellulose and dextran. *Polymer*, 40:2649–2657, 1999.

[4] T. R. Noel, R. Parker, and S. G. Ring. Effect of molecular structure and water content on the dielectric relaxation behaviour of amorphous low molecular weight carbohydrates above and below their glass transition. *Carbohydr. Res.*, 329:839–845, 2000.

[5] H. Charley and C. Weaver. *Foods: a scientific approach.* Prentice-Hall, Inc., 1998.

[6] W. L. Doane. USDA research on starch-based biodegradable plastics. *Starch/Stärke*, 44(8):293–295, 1992.

[7] A. Benbrahim. Degrade and deliver. *Chemistry in Britain*, pages 40–43, 2002.

[8] C. L. Swanson, R. L. Shogren, G. F. Fanta, and S. H. Imam. Starch-plastic materials - preparation, physical properties, and biodegradability (a review of recent USDA research). *Journal of Environmental Polymer Degradation*, 1(2):155–165, 1993.

[9] R. L. Shogren and B. K. Jasberg. Aging properties of extruded high-amylose starch. *Journal of Environmental Polymer Degradation*, 2(2):99–109, 1994.

[10] D. Lourdin, P. Colonna, G. J. Brownsey, T. R. Noel, and S. G. Ring. Structural relaxation and physical aging of starchy materials. *Carbohydr. Res.*, 337:827–833, 2002.

[11] R. L. Shongren. Effect of moisture content on the melting and subsequent physical aging of cornstarch. *Carbohydr. Polym.*, 19:83–90, 1992.

[12] S. H. D. Hulleman, F. H. P. Janssen, and H. Feil. The role of water during plasticization of native starches. *Polymer*, 39(10):2043–2048, 1998.

University of Cape Town

[13] Q. Liu and D. B. Thompson. Effects of the moisture content and different gelatinization heating temperatures on retrogradation of waxy-type maize starches. *Carbohydr. Res.*, 314:221–235, 1998.

[14] A. Stading, Å. Rindlav-Westling, and P. Gatenholm. Humidity-induced structural transitions in amylose and amylopectin films. *Carbohydr. Polym.*, 45:209–217, 2001.

[15] G. Rappenecker and P. Zugenmaier. Detailed refinement of the crystal structure of V_h-amylose. *Carbohydr. Res.*, 89:11–19, 1981.

[16] A. M. Stephen and H. F. Zobel. Starch: Structure, analysis and application. In A. M. Stephen, editor, *Food Polysaccharides and Their Applications*, Food sience and technology. Marcel Dekker Inc., New York, 1995.

[17] K. Gessler, I. Usón, T. Takaha, N. Krauss, S. M. Smith, S. Okada, G. M. Sheldrick, and W. Saenger. V-amylose at atomic resolution: X-ray structure of a cycloamylose with 26 glucose residues (cyclomaltohexaicosaose). *Proc. Natl. Acad. Sci. USA*, 96:4246–4251, 1999.

[18] A. P. Gunning, T. P. Giardina, C. B. Faulds, N. Juge, S. G. Ring, G. Williamson, and V. J. Morris. Surfactant-mediated solubilisation of amylose and visualization by atomic force microscopy. *Carbohydr. Polym.*, 51:177–182, 2003.

[19] G. J. Quigley, A. Sarko, and R. H. Marchessault. Crystal and molecular structure of maltose monohydrate. *J. Am. Chem. Soc.*, 92(20):5834–5839, 1970.

[20] G. M. Brown, D. C. Rohrer, B. Berking, C. A. Beevers, R. O. Gould, and R. Simpson. The crystal structure of α,α-trehalose dihydrate from three independent X-ray determinations. *Acta Cryst.*, pages 3145–3158, 1972.

[21] M. E. Gress and G. A. Jeffrey. A neutron diffraction refinement of the structure of β-maltose monohydrate. *Acta. Crystallogr.*, B33:2490–2495, 1977.

[22] S. S. C. Chu and G. A. Jefferey. The crystal structure of methyl-β-maltopyranoside. *Acta Crystallogr.*, 23:1038–1049, 1967.

[23] S. S. C. Chu and G. A. Jefferey. The refinement of the crystal structures of β-D-glucose and cellobiose. *Acta Crystallogr. Sect B*, 24:830–838, 1968.

[24] Z. Peralta-Inga, G. P. Johnson, M. K. Dowd, J. A. Rendleman, E. D. Stevens, and A. D. French. The crystal structure of the α-cellobiose $\cdot 2NaI \cdot H_2O$ complex in the context of related structures and conformational analysis. *Carbohydr. Res.*, 337:851–861, 2002.

University of Cape Town

[25] S. Raymond, B. Henrissat, D. T. Qui, Å.Kvivk, and H. Chancy. The crystal structure of methyl β-cellotrioside monohydrate 0.25 ethanolate and its relationship to cellulose II. *Carbohydr. Res.*, 277:209–229, 1995.

[26] K. Gessler, N. Krauss, T. Steiner, C. Betzel, A. Sarko, and W. Saenger. β-D-cellotetraose hemihydrate as a structural model for cellulose II. An X-ray diffraction study. *J. Am. Chem. Soc.*, 117(46):11397–11406, 1995.

[27] B. Klar, B. Hingerty, and W. Saenger. Topography of cyclodextrin inclusion complexes. XII. Hydrogen bonding in the crystal structure of α-cyclodextrin hexahydrate: the use of a multicounter detector in neutron diffraction. *Acta Crys.,Sect.B*, 36:1154–1164, 1980.

[28] K. Linder and W. Saenger. Crystal and molecular structure of cyclohepta-amylose dodecahydrate. *Carbohydr. Res.*, 99:103–115, 1982.

[29] J. Jacob, K. Geßler, D. Hoffman, H. Sanbe, K. Koizumi, S. M. Smith, T. Takaha, and W. Saenger. Strain-induced band-flips in cyclodecaamylose and higher homologues. *Angew. Chem. Int. Ed.*, 37(5):606–609, 1998.

[30] P. Zugenmaier. Conformation and packing of various crystalline cellulose fibers. *Prog. Polym. Sci.*, 26:1341–1417, 2001.

[31] A. S. Shashkov, G. M. Lipkind, and N. K. Kochetkov. Nuclear Overhauser effects for methyl β-maltoside and the conformational states of maltose in aqueous solution. *Carbohydr. Res.*, 147:175–182, 1986.

[32] P. Dais. Carbon-13 magnetic relaxation and local chain motion of amylose in dimethyl sulfoxide. *Carbohydr. Res.*, 160:73–93, 1987.

[33] J. Kowalewski, L. Mäler, and G. Widmalm. NMR relaxation studies of oligosaccharides in solution: reorientational dynamics and internal motion. *J. Mol. Liq.*, 78:255–261, 1998.

[34] D. A. Brant, H.-S. Liu, and Z. S. Zhu. The dependence of glucan conformational dynamics on linkage position and stereochemistry. *Carbohydr. Res.*, 278:11–26, 1995.

[35] D. A. Brant. Shapes and motions of polysaccharide chains. *Pure & Appl. Chem.*, 69(9):1885–1892, 1997.

[36] H. Sugiyama, K. Hisamichi, T. Usui, K. Sakai, and J.-I. Ishiyama. A study of the conformation of β-1,4-linked glucose oligomers, cellobiose to cellohexaose, in solution. *J. Mol. Struct.*, 556:173–177, 2000.

[37] I. J. van den Dries, D. van Dusschoten, and M. A. Hemminga. Mobility in maltose-water glasses studied with ^{1}H NMR. *J. Phys. Chem. B*, 102:10483–10489, 1998.

University of Cape Town

[38] I. J. van den Dries, D. van Dusschoten, M. A. Hemminga, and E. van der Linden. Effects of water content and molecular weight on spin probe and water mobility in malto-oligomer glasses. *J. Phys. Chem. B*, 104:10126–10132, 2000.

[39] M. Linnerborg, A. Weintraub, and G. Widmalm. Structural studies utilizing ^{13}C-enrichment of the *O*-antigen polysaccharide from the enterotoxigenic *Escherichia coli* O159 cross-reacting with *Shigella dysenteriae* type 4. *Eur. J. Biochem*, 266:246–251, 1999.

[40] E. Vinogravdov, B. O. Petersen, and J. O. Duus. Isolation and characterization of non-labeled and ^{13}C-labeled mannans from *Pichia pastoris* yeast. *Carbohydr. Res.*, 325:216–221, 2000.

[41] C. Zwahlen and S. J. F. Vincent. Determination of ^{1}H homonuclear scalar couplings in unlabeled carbohydrates. *J. Am. Chem. Soc.*, 124:7235–7239, 2002.

[42] D. A. Rees. Conformational analysis of polysaccharides. Part V. The characterisation of linkage conformations (chain conformations) by optical rotation at a single wavelength. *J. Chem. Soc. B*, pages 877–887, 1970.

[43] D. A. Rees and D. Thorn. Polysaccharide conformation. Part 10. Solvent and temperature effects on the optical rotation and conformation of model carbohydrates. *J. C. S. Perkin II*, pages 191–201, 1977.

[44] E. S. Stevens and B. K. Sathyanarayana. Potential energy surfaces of cellobiose and maltose in aqueous solution: a new treatment of disaccharide optical rotation. *J. Am. Chem. Soc.*, 111(12):4149–4154, 1989.

[45] C.-W. von der Lieth and T. Kozár. A (critical) survey of modelling protocols used to explore the conformational space of oligosaccharides. *J. Molec. Struct. (Theochem)*, 395-396:225–244, 1997.

[46] C. Landersjö, R. Stenutz, and G. Widmalm. Conformational flexibility of carbohydrates: A folded conformer at the ϕ dihedral angle of a glycosidic linkage. *J. Am. Chem. Soc.*, 119:8695–8698, 1997.

[47] G. M. Lipkind, V. E. Verovsky, and N. K. Kochetkov. Conformational states of maltose and cellobiose in solution: a comparison of calculated and experimental data. *Carbohydr. Res.*, 133:1–13, 1984.

[48] G. M. Lipkind, A. S. Shashkov, and N. K. Kochetkov. Nuclear Overhauser effect and conformational states of cellobiose in solution. *Carbohydr. Res.*, 141:191–197, 1985.

[49] D. A. Cumming and J. P. Carver. Virtual and solution conformations of oligosaccharides. *Biochemistry*, 26:6664–6676, 1987.

University of Cape Town

[50] D. A. Brant. Novel approaches to the analysis of polysaccharide structures. *Curr. Opin. Struc. Biol.*, 9:556–562, 1999.

[51] M. Rief, F. Oesterhelt, B. Heymann, and H. E. Gaub. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science*, 275:1295–1297, 1997.

[52] H. Li, M. Rief, F. Oesterhelt, H. E. Gaub, X. Zhang, and J. Shen. Single-molecule force spectroscopy on polysaccharides by AFM. *Chem. Phys. Lett.*, 305:197–201, 1999.

[53] P. E. Marszalek, H. Li, A. F. Oberhauser, and J. M. Fernandez. Chair-boat transitions in single polysaccharide molecules observed with force-ramp AFM. *Proc. Natl. Acad. Sci. USA*, 99(7):4278–4283, 2002.

[54] K. Lycknert, A. Maliniak, and G. Widmalm. Analysis of oligosaccharide conformation by NMR spectroscopy utilizing $^{1}H,^{1}H$ and $^{1}H,^{13}C$ residual dipolar couplings in a dilute crystalline phase. *J. Phys. Chem. A*, 105:5119, 2001.

[55] B. Stevensson, C. Landersjo, G. Widmalm, and A. Maliniak. Conformational distribution function of a dissacharide in a liquid crystalline phase determined using NMR spectroscopy. *J. Am. Chem. Soc.*, 124:5946–5947, 2002.

[56] M. Kadkhodaei, H. Wu, and D. A. Brant. Comparison of the conformational dynamics of the $(1 \rightarrow 4)$ and $(1 \rightarrow 6)$-linked α-D-glucans using ^{13}C-NMR relaxation. *Biopolymers*, 31:1581–1592, 1991.

[57] R. B. Best, G. E. Jackson, and K. J. Naidoo. Molecular dynamics and NMR study of the $\alpha(1\rightarrow4)$ and $\alpha(1\rightarrow6)$ glycosidic linkages: maltose and isomaltose. *J. Phys. Chem. B*, 105:4742–4751, 2001.

[58] K. I. Shingel. Determination of structural peculiarities of dextran, pullulan and γ-irradiated pullulan by Fourier-transform IR spectroscopy. *Carbohydr. Res.*, 337:1445–1451, 2002.

[59] H. Montès, K. Mazeau, and J. Y. Cavaillé. Secondary mechanical relaxations in amorphous cellulose. *Macromolecules*, 30:6977–6984, 1997.

[60] T. R. Noel, R. Parker, and S. G. Ring. A comparative study of the dielectric relaxation behaviour of glucose, maltose, and their mixtures with water in the liquid and glassy states. *Carbohydr. Res.*, 282:193–206, 1996.

[61] J. Stenger, M. Cowman, F. Eggers, E. M. Eyring, U. Kaatze, and S. Petrucci. Molecular dynamics and kinetics of monosaccharides in solution. A broadband ultrasonic relaxation study. *J. Phys. Chem. B*, 104:4782–4790, 2000.

University of Cape Town

[62] R. Polacek, J. Stenger, and U. Kaatze. Chair-chair conformational flexibility, pseudorotation and exocyclic group isomerization of monosaccharides in water. *J. Chem. Phys.*, 116(7):2973–2982, 2002.

[63] H. R. Tang and P. S. Belton. Molecular dynamics of polycrystalline cellobiose studied by solid-state NMR. *Solid State Nuclear Magnetic Resonance*, 21:117–133, 2002.

[64] M. R. Wormald, A. J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliot, and R. A. Dwek. Conformational studies of oligosaccharides and glycopeptides: Complementarity of NMR, X-ray crystallography and molecular modelling. *Chem. Rev.*, 102:371–386, 2002.

[65] E. P. Stroyan and E. S. Stevens. An improved model for calculating the optical rotation of simple saccharides. *Carbohydr. Res.*, 327:447–453, 2000.

[66] R. Bukowski, L. M. Morris, R. J. Woods, and T. Weimar. Synthesis and conformational analysis of the T-antigen disaccharide (β-D-Gal-(1$\rightarrow$ 3)-α-D-GalNAc-OMe). *Eur. J. Org. Chem.*, pages 2697–2705, 2001.

[67] B. J. Hardy, A. Gutierrez, K. Lesiak, E. Seidl, and G. Widmalm. Structural anyalysis of the solution conformation of methyl 4-*O*-β-D-glucopyranosyl-α-D-glucopyranoside by Molecular Mechanics and *ab initio* calculation, stochastic dynamics simulation and NMR spectroscopy. *J. Phys. Chem.*, 100:9187–9192, 1996.

[68] K. Ott and B. Meyer. Molecular dynamics simulations of maltose in water. *Carbohydr. Res.*, 281:11–34, 1996.

[69] C. X. Wang, W. Z. Chen, V. Tran, and R. Douillard. Analysis of interfacial water structure and dynamics in α-maltose solution by molecular dynamics simulation. *Chem. Phys. Lett.*, 251:268–274, 1996.

[70] A. Vishnyakov, G. Widmalm, J. Kowalewski, and A. Laaksonen. Molecular daynamics simulation of the α-D-Manρ-(1 $\rightarrow$ 3)-β-D-Glcρ-OMe disaccharide in water and water/DMSO solution. *J. Am. Chem. Soc.*, 121:5403–5412, 1999.

[71] A. Vishnyakov, A. Laaksonen, and G. Widmalm. Molecular dynamics simulation of the α-D-manρ-(1 $\rightarrow$ 3)-β-D-glcρ-OMe in methanol and in dimethyl sulfoxide solutions. *J. Mol. Graphics Modell.*, 19(3/4):338–342, 2001.

[72] K. J. Naidoo and M. Kuttel. Water structure about the dimer and hexamer repeat units of amylose from molecular dynamics computer simulations. *J. Comput. Chem.*, 22(4):445–456, 2001.

[73] K. Lycknert, C. Höög, and G. Widmalm. Conformational studies of the two anomeric methyl glycosides of α-D-Manρ-(1$\rightarrow$2)-D-Glcρ by molecular simulations and NMR ^{1}H, ^{1}H T-ROESY experiments. *J. Chem. Soc., Perkin Trans. 2*, pages 416–421, 2002.

University of Cape Town

[74] J. W. Brady. Molecular dynamics simulations of α-D-glucose in aqueous solution. *J. Am. Chem. Soc.*, 111:5155–5165, 1989.

[75] J. W. Brady and R. K. Schmidt. The role of hydrogen bonding in carbohydrates: Molecular dynamics simulations of maltose in aqueous solution. *J. Phys. Chem.*, 97:958–966, 1993.

[76] S. A. Galema, E. Howard, J. B. F. N. Engberts, and J. P. Grigeria. The effect of stereochemistry upon carbohydrate hydration. A Molecular Dynamics simulation of β-D-galactopyranose and (α,β)-D-talopyranose. *Carbohydr. Res.*, 265:215–225, 1994.

[77] Q. Liu and J. W. Brady. Model dependence of the aniostropic structuring of solvent water around sugars in molecular dynamics simulations. *J. Phys. Chem. B*, 101(8):1817–1312, 1997.

[78] P. B. Conrad and J. J. Pablo. Computer simulation of the cryoprotectant dissacharide α,α-trehalose in aqueous solution. *J. Phys. Chem. A.*, 103:4049–4055, 1999.

[79] T. Astley, G. G. Birch, M. G. B. Drew, and P. M. Rodger. Lifetime of a hydrogen bond in aqueous solutions of carbohydrates. *J. Phys. Chem.*, 103:5080–5090, 1999.

[80] A. Almond, A. Brass, and J. K. Sheehan. Oligosaccharides as model systems for understanding water-biopolymer interaction: hydrated dynamics of a hyaluronan decamer. *J. Phys. Chem. B*, 104:5634–5640, 2000.

[81] S. B. Engelsen and S. Pérez. Unique similarity of the asymmetric trehalose solid-state hydration and the diluted aqueous-solution hydration. *J. Phys. Chem. B*, 104:9301–9311, 2000.

[82] S. N. Ha, L. J. Madsen, and J. W. Brady. Conformational analysis and molecular dynamics simulations of maltose. *Biopolymers*, 27:1927–1952, 1988.

[83] A. D. French. Comparisons of rigid and relaxed conformational maps of maltose and cellobiose. *Carbohydr. Res.*, 188:206–211, 1989.

[84] M. K. Dowd, A. D. French, and P. J. Reilly. Conformational analysis of the anomeric forms of sophorose, laminarabiose and cellobiose using MM3. *Carbohydr. Res.*, 233:15–34, 1992.

[85] B.J. Hardy and A. Sarko. Conformational analysis and molecular dynamics simulation of cellobiose and larger cellooligomers. *J. Comput. Chem.*, 14(7):831–847, 1993.

[86] T. M. Glennon, Y. Zheng, S. M. le Grand, B. A. Shutzberg, and K. M. Jnr. Merz. A force field for monosaccharides and (1 $\rightarrow$ 4) linked polysaccharides. *J. Comput. Chem.*, 15(9):1019–1040, 1994.

University of Cape Town

[87] K. J. Naidoo and J. W. Brady. The application of simulated annealing to the conformational analysis of disaccharides. *Chem. Phys.*, 224:261–273, 1997.

[88] R. K. Schmidt, B. Teo, and J. W. Brady. Use of umbrella sampling in the calculation of the potential of mean force for maltose in vacuum from molecular dynamics simulations. *J. Phys. Chem*, 99:11339–11343, 1995.

[89] K. J. Naidoo and J. W. Brady. Calculation of the Ramachandran potential of mean force for a disaccharide in aqueous solution. *J. Am. Chem. Soc.*, 121(10):2244–2252, 1999.

[90] S. Immel and F. W. Lichtenthaler. The conformation of sucrose in water: a molecular dynamics approach. *Liebigs Ann.*, pages 1925–1937, 1995.

[91] A. E. Mark, S. P. van Helden, P. E. Smith, L. H. M. Jassen, and W. F. van Gunsteren. Convergence properties of free energy calulations: α-cyclodextrin complexes as a case study. *J. Am. Chem. Soc.*, 116:6293–6302, 1994.

[92] H. Senderowitz and W. C. Still. Anomeric free energy of D-mannose in water: a comparison of free energy pertubation, potential of mean force and MC(JBW)/SD simulations. *J. Phys. Chem. B.*, 101:1409–1412, 1997.

[93] P. Höchtl, S. Boresch, and O. Steinhauser. Dielectric properties of glucose and maltose solutions. *J. Chem. Phys.*, 112(22):9810–9821, 2000.

[94] S. J. Angyal. Conformational analysis in carbohydrate chemistry. I: Conformational free energies. The conformations and a:b ratios of aldopyranoses in aqueous solution. *Aust. J. Chem.*, 21:2737–2746, 1968.

[95] S. J. Angyal. The composition and conformation of sugars in solution. *Angwe. Chem. Int. Ed. Engl.*, 8:157–226, 1969.

[96] J. T. Edward. Stability of glycosides to acid hydrolysis. *Chem. Ind (London)*, pages 1102–1104, 1955.

[97] E. A. C. Lucken. Chemical applications of nuclear quadropole resonance spectroscopy. Part III. The inductive effect of substituents belonging to the first row of the periodic table. *J. Chem. Soc.*, pages 2954–2960, 1959.

[98] S. Wolfe, M. H. Whangbo, and D. J. Mitchell. On the magnitudes and origins of the "anomeric effects", "exo-anomeric effects", "reverse anomeric effects", and C—X and C—Y bond-lengths in XCH2YH molecules. *Carbohydr. Res.*, 69:1–10, 1979.

[99] I. Tvaroška and T. Bleha. Anomeric and exo–anomeric effect in carbohydrate chemistry. *Adv. Carbohydr. Chem. Biochem,*, 47:45–124, 1989.

University of Cape Town

[100] M. Jaseja, A. S. Perlin, and P. Dais. Two-dimensional NMR spectral study of the tautomeric equilibria of D-fructose and related compounds. *Magn. Reson. Chem.*, 28:283–289, 1990.

[101] J.-P. Praly and R. U. Lemieux. Influence of solvent on the magnitude of the anomeric effect. *Can. J. Chem*, 65:213–220, 1987.

[102] C. J. Cramer and D. G. Truhlar. Quantum chemical conformational analysis of glucose in aqueous solution. *J. Am. Chem. Soc.*, 115:5745–5753, 1993.

[103] P. L. Polavarapu and C. S. Ewig. *Ab ibitio* and computed molecular structures and energies of the conformers of glucose. *J. Comput. Chem.*, 13(10):1255–1261, 1992.

[104] U. Salzner and P. V. Schleyer. *Ab Initio* examination of anomeric effects in tetrahydropyrans, 1,3-dioxanes, and glucose. *J. Org. Chem.*, 59(8):2138–2155, 1994.

[105] J. W. Brown and B. D. Wladkowski. *Ab initio* studies of the exocyclic hydroxymethyl rotational surface in α-D-glucopyranose. *J. Am. Chem. Soc.*, 118:1190–1193, 1996.

[106] S. E. Barrows, J. W. Storer, C. J. Cramer, A. D. French, and D. G. Truhlar. Factors controlling relative stability of anomers and hydroxymethyl conformers in glucopyranose. *J. Comput. Chem.*, 19(10):1111–1129, 1998.

[107] C. Höög and G. Widmalm. Free energy simulations of D-xylose in water and methyl D-xylopyranoside in methanol. *J. Phys. Chem. B*, 105(27):6375–6379, 2001.

[108] S. Ha, J. Gao, B. Tidor, J. W. Brady, and M. Karplus. Solvent effect on the anomeric equilibrium in D-glucose: A free energy simulation analysis. *J. Am. Chem. Soc.*, 113:1553–1557, 1991.

[109] B. P. van Eijck, R. W. W. Hooft, and J. Kroon. Molecular dynamics study of conformational and anomeric equilibria in aqueous D-glucose. *J. Phys. Chem.*, 97:12093–12099, 1993.

[110] C. Simmerling, T. Fox, and P. A. Kollman. Use of locally enhanced sampling in free energy calculations: testing and application to the α-β anomerization of glucose. *J. Am. Chem. Soc.*, 120:5771–5782, 1998.

[111] R. K. Schmidt, M. Karplus, and J. W. Brady. The anomeric equilibrium in D–xylose: Free energy and the role of solvent structuring. *J. Am. Chem. Soc.*, 118:541–546, 1996.

[112] I. Tvaroška and J. P. Carver. *Ab Initio* molecular orbital calculation of carbohydrate model compounds. 2. Conformational analysis of axial and equatorial 2-methoxytetrahydropyrans. *J. Phys. Chem.*, 98:9477–9485, 1994.

University of Cape Town

[113] I. Tvaroška and J. P. Carver. *Ab initio* molecular orbital calculation of carbohydrate model compounds. 3. Effect of the electric field on conformations about the glycosidic linkage. *J. Phys. Chem.*, 99:6234–6241, 1995.

[114] D. Cremer and J. A. Pople. A general definition of ring puckering coordinates. *J. Am. Chem. Soc.*, 97(6):1354–1358, 1975.

[115] J. W. Brady. Molecular dynamics simulations of α-D-glucose. *J. Am. Chem. Soc.*, 108:8153–8160, 1986.

[116] J. W. Brady. Molecular dynamics simulations of β-D-glucopyranose. *Carbohydr. Res.*, 165:306–312, 1987.

[117] G. M. Brown and H. A. Levey. α-D-glucose: Precise determination of crystal and molecular structure by neutron-diffraction analysis. *Science*, 147:1038–1039, 1965.

[118] M. K. Dowd, A. D. French, and P. J. Reilly. Modeling of aldopyranosyl ring puckering with MM3(92). *Carbohydr. Res.*, 264:1–19, 1994.

[119] P. O'Donoghue and Z. A. Luthey-Schulten. Barriers to forced transitions in polysaccharides. *J. Phys. Chem. B.*, 104:10398–10405, 2000.

[120] B. Heymann and H. Grubmüller. 'Chair-boat' transitions and side groups affect the stiffness of polysaccharides. *Chem. Phys. Lett.*, 305:202–208, 1999.

[121] P. E. Marszalek, A. F. Oberhauser, Y.-P. Pang, and J. M. Fernandez. Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring. *Nature*, 396:661–664, 1998.

[122] P. E. Marszalek, Y.-P. Pang, H. Li, Y. E. Yazal, A. F. Oberhauser, and J. M. Fernandez. Atomic levers control pyranose ring conformations. *Proc. Natl. Acad. Sci. USA*, 96:7894–7898, 1999.

[123] K. Bock and J. Duus. A conformational study of hydroxymethyl groups in carbohydrates investigated by H1–NMR spectroscopy. *J. Carbohydr. Chem.*, 13(4):513–543, 1994.

[124] Y. Nishida, H. Ohrui, and H. Meguro. ^{1}H-NMR studies of (6R)- and (6S)-deuterated D-hexoses: Assignment of the preferred rotamers about C6-C6 bond of D-glucose and D-galactose derivatives in solutions. *Tetrahedron Lett.*, 25(15):1575–1578, 1984.

[125] H. Ohrui Y. Nishida and H. Meguro. 1H-NMR analyses of rotameric distributions of C5-C6 bonds of D-glucopyranoses in solution. *J. Carbohydr. Chem.*, 7:239–250, 1988.

[126] G. D. Rockwell and T. B. Grindley. Effect of solvation on the rotation of hydroxymethyl groups in carbohydrates. *J. Am. Chem. Soc.*, 120:10953–10963, 1998.

University of Cape Town

[127] K. De Vries and H. M. Buck. Different rotamer populations around the C5—C6 bond for α- and β-D-galactopyranosides through the combined interaction of the gauche and anomeric effects: A 300-MHz ^{1}H-NMR and MNDO study. *Carbohydr. Res.*, 165:1–16, 1987.

[128] I. Tvaroška and J. P. Carver. *Ab Initio* molecular orbital calculation of carbohydrate model compounds. 6. The gauche effect and conformations of the hydroxymethyl and methoxymethyl groups. *J. Phys. Chem. B*, 101:2992–2999, 1997.

[129] K. N. Kirschner and R. J. Woods. Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U.S.A.*, 98:10541, 2001.

[130] C. Höög, C. Landersjö, and G. Widmalm. Oligosaccharides display both rigidity and high flexibility in water as determined by ^{13}C NMR relaxation and ^{1}H, ^{1}H NOE spectroscopy: Evidence of anti-phi and anti-psi torsions in the same glycosidic linkage. *Chem. Eur. J.*, 7:3069–3071, 2001.

[131] G. A. Jeffrey and R. Nanni. The crystal structure of anhydrous α,α-trehalose at -150°. *Carbohydr. Res.*, 137:21–30, 1985.

[132] C. A. Duda and E. S. Stephens. Trehalose conformation in aqueous solution from optical rotation. *J. Am. Chem. Soc.*, 112:7406, 1990.

[133] A. F. Bell, L. Hecht, and L. D. Barron. Disaccharide solution stereochemistry from vibrational Raman optical activity. *J. Am. Chem. Soc.*, 116:5155–5161, 1994.

[134] B. A. Burton and D. A. Brant. Comparative flexibility, extension and conformation of simple polysaccharide chains. *Biopolymers*, 22:1769–1792, 1983.

[135] C. Fringant, I. Tvaroška, K. Mazeau, M. Rinaudo, and J. Desbrieres. Hydration of α-maltose and amylose: molecular modelling and thermodynamics study. *Carbohydr. Res.*, 278:27–41, 1995.

[136] F. A. Momany and J. L. Willet. Computational studies on carbohydrates: I. density functional *ab initio* geometry optimization on maltose configurations. *J. Comput. Chem.*, 21(13):1204–1219, 2000.

[137] S. Melberg and K. Rasmussen. Conformations of disaccharides by empirical force-field calculations: part ii, β-cellobiose. *Carbohydr. Res.*, 71:25–34, 1979.

[138] B.J. Hardy and A. Sarko. Molecular dynamics simulation of cellobiose in water. *J. Comput. Chem.*, 14(7):848–857, 1993.

[139] V. H. Tran and J. W. Brady. Disaccharide conformational flexibility. I. An adiabatic potential energy map for sucrose. *Biopolymers*, 29(6-7):961–976, 1990.

University of Cape Town

[140] V. H. Tran and J. W. Brady. Disaccharide conformational flexibility. II. Molecular dynamics simulations of sucrose. *Biopolymers*, 29(6-7):977–997, 1990.

[141] P. J. Hadjuk, D. A. Horita, and L. E. Lerner. Picosecond dynamics of simple monosaccharides as probed by NMR and molecular dynamics simulations. *J. Am. Chem. Soc.*, 115:9196–9201, 1993.

[142] G. Longhi, M. Malandrino, and S. Abbate. Investigations on the mobility of the glycosidic linkage in sucrose by study of the phase space structure of a two-degrees of freedom model. *J. Mol. Graphics Modell.*, 18:153–162, 2000.

[143] M. K. Dowd, P. J. Reilly, and A. D. French. Conformational analysis of trehalose disaccharides and analogues using MM3. *J. Comput. Chem.*, 13(1):102–114, 1992.

[144] M. C. Donnamaria, E. I. Howard, and J. R. Grigera. Interaction of water with α,α-trehalose in solutions: Molecular dynamics approach. *J. Chem. Soc. Faraday Trans.*, 90(18):2731–2735, 1994.

[145] M. Sakurai, M. Murata, Y. Inoue, A. Hino, and S. Kobayashi. Molecular Dynamics study of an aqueous solution of trehalose and maltose: Implication for the biological function of trehalose. *Bull. Chem. Soc. Jpn.*, 70(4):847–858, 1997.

[146] G. Bonanno, R. Noto, and S. L. Fornili. Water interaction with alpha,alpha-trehalose: molecular dynamics simulation. *J. Chem. Soc. Faraday Trans.*, 94(18):2755–2762, 1998.

[147] J. S. Taylor, B. Teo, D. B. Wilson, and J. W. Brady. Conformational modeling of substrate binding to endocellulase E2 from *Thermomonospora fusca*. *Protein Engineering*, 8(11):1145–1152, 1995.

[148] E. Mikros, G. Labrinidus, and S. Pérez. Conformational analysis of C-disaccharides using molecular mechanics calculations. *J. Carbohydr. Chem.*, 19(9):1319–1349, 2000.

[149] D. A. Rees and W. E. Scott. Polysaccharide conformation. Part IV. Computer model-building for linear and branched pyranoglycans. Correlations with biological function. Preliminary assessment of interresidue forces in aqueous solution. Further interpretation of optical rotation in terms of chain conformation. *J. Chem. Soc. B*, pages 469–479, 1971.

[150] J. Sugiyama, R. Vuong, and H. Chanzy. Electron diffraction study on the two crystalline phases occuring in native cellulose from an algal cell wall. *Macromolecules*, 24:4168–4175, 1991.

[151] W. Hindrichs, G. Büttner, M. Steifa, C. Betzel, V. Zabel, B. Pfannemüler, and W. Saenger. An amylose antiparallel double helix at atomic resolution. *Science*, 238:205–208, 1987.

University of Cape Town

[152] W. Hindrichs and W. Saenger. Crystal and molecular structure of the hexasaccharide complex (ρ-nitrophenyl α-maltohexaoside)$_2$ · Ba$(I_3)_2$ · 27H_2O. *J. Am. Chem. Soc.*, 112:2789–2796, 1990.

[153] W. Banks and C. T. Greenwood. The hydrodynamic behaviour of amylose in good solvents. *Carbohydr. Res.*, 7:414–420, 1968.

[154] W. Banks and C. T. Greenwood. Amylose: a non-helical polymer in aqueous solution. *Polymer*, 12:141–145, 1971.

[155] W. Banks and C. T. Greenwood. The conformation of amylose in alkaline salt solution. *Carbohydr. Res.*, 21:229–234, 1972.

[156] F. R. Dintzis and R. Tobin. Optical rotation of some α-1,4-linked glucopyranosides in the system H_2O-DMSO and solution conformation of amylose. *Biopolymers*, 7:581–593, 1969.

[157] F. R. Dintzis and R. Tobin. Light-scattering and specific refractive increment behaviour of amylose and dextran in dimethyl sulfoxide-water. *Carbohydr. Res.*, 66:71–83, 1978.

[158] N. W. H. Cheetham and L. Tao. Amylose conformational transitions in binary dmso/water mixtures. *Carbohydr. Polym.*, 35:287–295, 1998.

[159] J. Shimada, H. Kaneko, T. Takada, S. Kitamura, and K. Kajiwara. Conformation of amylose in aqueous solution: small angle x-ray scattering measurements and simulations. *J. Phys. Chem. B*, 104:2136–2147, 2000.

[160] G. S. Buliga and D. A. Brant. Theoretical interpretation of the unperturbed aqueous-solution conformation of pullulan. *Int. J. Biol. Macromol.*, 9:77–86, 1987.

[161] A. J. Benesi and D. A. Brant. Trends in molecular motion for a series of glucose oligomers and the corresponding polymer pullulan as measured by carbon-13 NMR relaxation. *Macromolecules*, 18(6):1109–1116, 1985.

[162] P. J. Flory. *Statistical Mechanics of Chain Molecules.* Wiley-Interscience, New York, 1969.

[163] J.H.-Y. Liu, K. A. Brameld, D. A. Brant, and W. A. Goddard III. Conformational analysis of aqueous pullulan oligomers: an effective computational approach. *Polymer*, 43:509–516, 2002.

[164] D. P. Miller and J. J. de Paulo. Calorimetric solution properies of simple saccharides and their significance for the stabilization of biological structure and function. *J. Phys. Chem. B*, 104:8876–8883, 2000.

[165] A. L. Ollett, R. Parker, and A. C. Smith. Deformation and fracture–behavior of wheat–starch plasticized with glucose and water. *J. Mater. Sci.*, 26(5):1351–1356, 1991.

University of Cape Town

[166] T. R. Noel, S. G. Ring, and M. A. Whittam. Dielectric relaxations of small carbohydrate molecules in the liquid and glassy states. *J. Phys. Chem.*, 96:5562–5567, 1992.

[167] R. K. Chan, K. Pathmanathan, and G. P. Johari. Dielectric relaxations in the liquid and glassy states of glucose and its water mixtures. *J. Phys. Chem.*, 90:6358–6362, 1986.

[168] Gangasharan and S. S. N. Murthy. Study of α-,β-, and γ-relaxation processes in some supercooled liquids and supercooled plastic crystals. *J. Chem. Phys.*, 99:9865–9873, 1993.

[169] U. Trommsdorff and I. Tomka. Structure of amorphous starch. 2. Molecular interactions with water. *Macromolecules*, 28:6138–6150, 1995.

[170] E. R. Caffarena and J. R. Grigera. Hydration of glucose in the rubbery and glassy states studied by molecular dynamics simulation. *Carbohydr. Res.*, 315:63–69, 1999.

[171] A. Benczédi, I. Tomka, and F. Escher. Thermodynamics of amorphous starch-water systems. 1. Concentration fluctuations. *Macromolecules*, 31:3062–3074, 1998.

[172] J. F. Carpenter and J. H. Crowe. The mechanism of cryoprotection of proteins by solutes. *Cryobiology*, 25:459–470, 1988.

[173] J. H. Crowe, J. F. Carpenter, L. M. Crowe, and T. J. Anchordoguy. Are freezing and dehydration similar stress vectors? a comparison of modes of interaction stabilizing solutes within biomolecules. *Cryobiology*, 27:219–231, 1990.

[174] T. Arakawa, Y. Kita, and J. F. Carpenter. Protein-solvent interactions in pharmaceutical formulations. *Pharm. Res.*, 8(3):285–291, 1991.

[175] J. L. Green and C. A. Agnell. Phase relations and vitrification in saccharide-water solutions and the trehalose anomaly. *J. Phys. Chem.*, 93:2880–2882, 1989.

[176] C. A. Duda and E. S. Stephens. Solution conformations of β-β-trehalose and its C-disaccharide analog from optical rotation. *J. Am. Chem. Soc.*, 115:8487, 1993.

[177] Q Liu, R. K. Schmidt, B. Teo, P. A. Karplus, and J. W. Brady. Molecular dynamics studies of the hydration of α,α-trehalose. *J. Am. Chem. Soc.*, 119(33):7851–7862, 1997.

[178] G. Batta, K. E. Kövér, J. Gervay, M. Horyák, and G. M. Roberts. Temperature dependence of molecular conformation, dynamics, and chemical shift anisotropy of α,α-trehalose in D_2O by NMR relaxation. *J. Am. Chem. Soc.*, 119:1336–1345, 1997.

[179] J. H. Crowe. Anhydrobiosis: An unsolved problem. *Am. Nat.*, 105(946):563–573, 1971.

[180] J. Szejtli. *Cyclodextrin Technology.* Topics in Inclusion Science. Kluwer Academic Publishers, 1988.

University of Cape Town

[181] S. N. Ha, A. Giammona, M. Field, and J. W. Brady. A revised potential-energy surface for molecular studies of carbohydrates. *Carbohydr. Res.*, 180:207–221, 1988.

[182] D. J. Giesen, G. D. Hawkins, D. A. Liotard, C. J. Cramer, and D. G. Truhlar. A universal model for the quantum mechanical calculation of free energies of solvation in non-aqueous solvents. *Theor. Chem. Acc.*, 98:85–108, 1997.

[183] S. A. Galema, J. B. F. N. Engberts, H. Höiland, and G. M. Forland. Informative thermodynamic properties of the effect of stereochemistry on carbohydrate hydration. *J. Phys. Chem*, 97:6885–6889, 1993.

[184] M. A. Kabayama, D. Patterson, and L. Piche. The thermodynamics of mutarotation of some sugars. I. Measurement of the heat of mutarotation by microcalorimetry. *Can. J. Chem.*, 36:557–562, 1958.

[185] M. A. Kabayama and D. Patterson. The thermodynamics of mutarotation of some sugars. II. Theorectical considerations. *Can. J. Chem.*, 36:563–573, 1958.

[186] F. Franks. Physical chemistry of small carbohydrates - equilibrium solution properties. *Pure & Appl. Chem.*, 59:1189–1202, 1987.

[187] F. Franks. Solute–water interactions: Do polyhydroxy compounds alter the properties of water? *Cryobiology*, 20:335–345, 1983.

[188] R. Behrends, M. K. Cowman, F.Eggers, E. M. Eyring, U. Kaatze, J. Majewski, S. Petrucci, K.-H. Richmann, and M. Reich. Ultrasonic relaxation and fast chemical kinetics of some carbohydrate aqueous solutions. *J. Am. Chem. Soc.*, 119:2182–2186, 1997.

[189] E. Tylianakis, P. Dais, I. Andre, and F. F. Taravel. Rotational dynamics of linear polysaccharides in solution. ^{13}C relaxation study on amylose and inulin. *Macromolecules*, 28:7962–7966, 1995.

[190] L. Mäler, G. Widmalm, and J. Kowalewski. Dynamical behavior of carbohydrates as studied by carbon-13 and proton nuclear spin relaxation. *J. Phys. Chem.*, 100:17103–17110, 1996.

[191] T. Rundlöf, R. M. Venable, R. W. Pastor, J. Kowalewski, and G. Widmalm. Distinguishing anisotropy and flexibility of the pentasaccharide LNF-1 in solution by carbon-13 NMR relaxation and hydrodynamic modeling. *J. Am. Chem. Soc.*, 121(50):11847–11854, 1999.

[192] A. Perico, M. Mormino, R. Urbani, A. Cesàro, E. Tylianakis, P. Dais, and D. A. Brant. Local dynamics of carbohydrates. 1. Dynamics of simple glycans with different chain linkages. *J. Phys. Chem. B*, 103:8162–8171, 1999.

University of Cape Town

[193] D. A. Brant and W. L. Dimpfl. A theoretical interpretation of the aqueous solution properties of amylose and its derivatives. *Macromolecules*, 3(5):655–665, 1970.

[194] H. W. Starkweather. Aspects of simple non–coperative relaxations. *Polymer*, 32(13):2443–2448, 1991.

[195] H. W. Starkweather. Noncooperative relaxations. *Macromolecules*, 21:1798, 1988.

[196] J. P. M. Lommerse, L. M. J. Kroon-Batenburg, J. P. Kamerling, and J. F. G. Vliegenthart. Conformational analysis of the xylose-containing N-glycan of pineapple stem bromelain as part of the intact glycoprotein. *Biochemistry*, 34(25):8196–8206, 1995.

[197] D. A. Brant and K. D. Goebel. A general treatment of the configurational statistics of polysaccharides. *Macromolecules*, 8(4):522–530, 1975.

[198] R. C. Jordan, D. A. Brant, and A. Cesàro. A Monte Carlo study of the amylosic chains conformations. *Biopolymers*, 17:2617–2632, 1978.

[199] M. A. Murcko and R. A. DiPaola. *Ab Initio* molecular conformational analysis of prototypical organic systems. 1. Ethylene glycol and 1,2-dimethoxyethane. *J. Am. Chem. Soc.*, 114:10010–10018, 1992.

[200] C. Molteni and M. Parrinello. Glucose in aqueous solution by first principles molecular dynamics. *J. Am. Chem. Soc.*, 120:2168–2171, 1998.

[201] A. Chung-Phillips and Y. Y. Chen. An *ab initio* study of fructose in the gas phase. *J. Phys. Chem. A*, 103(7):953–964, 1999.

[202] G. L. Strati, J. L. Willet, and F. A. Momany. *Ab initio* computational study of β-cellobiose conformers using B3LYP/6-311++GG**. *Carbohydr. Res.*, 337:1833–1849, 2003.

[203] A. D. French, A. Kelterer, G. P. Johnson, M. K. Dowd, and C. J. Cramer. HF/6-31G* energy surfaces for disaccharide analogs. *J. Comput. Chem.*, 22(1):65–78, 2001.

[204] R. J. Woods. Carbohydrate force fields. In Paul van Ragné Schleyer, editor, *Encyclopedia of Computational Chemistry*, volume 1, pages 220–232. John Wiley and Sons Ltd., Baffins Lane, Chichester, 1998.

[205] S. Pèrez, A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jimenez-Berbero, A. Povenda, J.-F. Espinosa, B. P. van Eyck, G. Johnson, A. D. French, M. L. C. E. Kouwijzer, P. D. J. Grootenuis, A. Bernadi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten, and K. Rasmussen. A comparison and chemometric analysis of several molecular mechanics force fields and parameter sets applied to carbohydrates. *Carbohydr. Res.*, 314:141–155, 1998.

University of Cape Town

[206] K. Mazeau and S. Pérez. The preferred conformations of the four oligomeric fragments of Rhamnogalacturonan II. *Carbohydr. Res.*, 311(4):203–217, 1998.

[207] R. A. Klein, S. O. Jónsdóttir, and H. Egge. Aspects of conformational mobility in charged oligosaccharides. *J. Molec. Struct. (Theochem)*, 395-396:313–332, 1997.

[208] P. D. J Grootenhuis and C. A. G. Haasnoot. A CHARMM–based force–field for carbohydrates using the CHEAT approach — carbohydrate hydroxyl groups represented by extended atoms. *Mol. Simul.*, 10:75–95, 1993.

[209] M. L. C. E. Kouwijzer and P. D. J. Grootenhuis. Parameterization and application of CHEAT95, an extended atom force field for hydrated (oligo)saccharides. *J. Phys. Chem*, 99:13426–13436, 1995.

[210] S. W. Homans. A molecular mechanics force field for the conformational analysis of oligosaccharides: Comparison of theoretical and crystal structures of Manα1-3Manβ1-4GlcNAc. *Biochemistry*, 29:9110–9118, 1990.

[211] Q. Liu and J. W. Brady. Anisotropic solvent structuring in aqueous sugar solutions. *J. Am. Chem. Soc.*, 118:12276–12286, 1996.

[212] K. Ott and B. Meyer. Parameterization of GROMOS force field for oligosaccharides and assessment of efficiency of molecular dynamics simulations. *J. Comput. Chem.*, 17(8):1068–1084, 1996.

[213] K. Ueda and J. W. Brady. Molecular dynamics simulations of carrabiose. *Biopolymers*, 41:323–330, 1997.

[214] R. Eklund and G. Widmalm. Molecular dynamics of an oligosaccharide using a force field for carbohydrates. *Carbohydr. Res.*, 338:393–398, 2003.

[215] A. R. Leach. *Molecular Modelling. Principles and Applications.* Addison Wesley Longman Limited, 1996.

[216] P. Kollman. Free energy calculations: Applications to chemical and biological phenomena. *Chem. Rev.*, 93:2395–2417, 1993.

[217] R. Palma, M. E. Himmel, and J. W. Brady. Calculation of the potential of mean force for the binding of glucose to benzene in solution. *J. Phys. Chem. B*, 104:7228–7234, 2000.

[218] R. Palma, M. E. Himmel, G. Liang, and J. W. Brady. Molecular mechanics studies of cellulases. In M. E. Himmel, editor, *Glycosyl Hydrolases in Biomass Conversion: ACS Symposium Series.* ACS, Washington, 2001.

University of Cape Town

[219] M. Kuttel, J. W. Brady, and K. J. Naidoo. Carbohydrate solution simulations: Producing a force field with experimentally consistent hydroxyl rotational frequencies. *J. Comput. Chem.*, 23(13):1236–1243, 2002.

[220] C. J. Cramer and D. G. Truhlar. An SCF solvation model for the hydrophobic affect and absolute free energies of aqueous solvation. *Science*, 256, 1992.

[221] C. Molteni and M. Parrinello. Condensed matter effects on the structure of crystalline glucose. *Chem. Phys. Lett.*, 275:409–413, 1997.

[222] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.

[223] P. K. Weiner and P. A. Kollman. Amber: Assisted model building program with energy refinement. A general program for modeling molecules and their interactions. *J. Comput. Chem.*, 2:287–303, 1981.

[224] W. van Gunsteren and H. Berendsen. GROMOS: GROningen MOlecular Simulation software. Technical report, Laboratory of Physical Chemistry, University of Groningen, 1988.

[225] W. Smith and T. R. Forester. DL_POLY 2.0 A general purpose parallel molecular dynamics package. *J. Molecular Graphics*, 14:136–151, 1996.

[226] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Oxford University Press Inc., 1987.

[227] W. F. van Gunsteren and H. J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, 29:992–1023, 1990.

[228] M. Karplus and G. A. Petsko. Molecular dynamics simulations in biology. *Nature*, 347:631–639, 1990.

[229] C. L. Brooks III and D. A. Case. Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.*, 93(7):2487–2502, 1993.

[230] N. Follope and Jr. A. D. MacKerell. All-atom empirical force field for nucleic acids: I. parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.*, 21(2):86–104, 2000.

[231] N. L. Allinger. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J. Am. Chem. Soc.*, 99(25):8127–8134, 1977.

University of Cape Town

[232] N. L. Allinger and J. H. Lii. Molecular mechanics. the MM3 force field for hydrocarbons. 2. Vibrational frequencies and thermodynamics. *J. Am. Chem. Soc.*, 111(23):8566–8575, 1989.

[233] R. U. Lemieux and S. Koto. The conformational properties of glycosidic linkages. *Tetrahedron*, 30:1933–1944, 1974.

[234] K. Bock. The preferred conformation of oligosaccharides in solution inferred from high resolution NMR data and hard–sphere exo–anomeric calculations. *Pure & Appl. Chem.*, 55:605–622, 1983.

[235] S. Reiling, M. Schlenkrich, and J. Brickman. Force field parameters for carbohydrates. *J. Comput. Chem.*, 17(4):450–468, 1996.

[236] R. J. Woods, R. A. Dwek, and C. J. Edge. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J. Phys. Chem*, 99:3832–3846, 1995.

[237] H. Senderowitz, C. Parish, and W. C. Still. Carbohydrates: United atom AMBER parameterization of pyranoses and simulations yielding anomeric free energies. *J. Am. Chem. Soc.*, 118:2078–2086, 1996.

[238] H. Senderowitz and W. C. Still. A quantum mechanically derived all-atom force field for pyranose oligosaccharides. amber parameters and free energy simulations. *J. Org. Chem.*, 62:1427–1438, 1997.

[239] S. K. Gregurick, J. H.-Y. Liu, D. A. Brant, and R. B. Gerber. Anharmonic vibrational self-consistent field calculations as an approach to improving force fields for monosaccharides. *J. Phys. Chem. B*, 103:3476–3488, 1999.

[240] F. A. Momany and J. L. Willet. Computational studies on carbohydrates: in vacuo studies using a revised amber force field, AMB99C, designed for α–(1→4) linkages. *Carbohydr. Res.*, 326:194–209, 2000.

[241] W. Damm, A. Frontera, J. Tirado-Rives, and W. L. Jorgensen. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.*, 18(16):1955–1970, 1997.

[242] D. Kony, W. Damm, S. Stoll, and W. F. van Gunsteren. An improved OPLS-AA force field for carbohydrates. *J. Comput. Chem.*, 23(15):1416–1429, 2002.

[243] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera,

University of Cape Town

D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586, 1998.

[244] A. Rahman and F. H. Stillinger. Molecular Dynamics study of liquid water. *J. Chem. Phys.*, 55:3336–3359, 1971.

[245] F. H. Stillinger and A. Rahman. Improved simulation of liquid water by Molecular Dynamics. *J. Chem. Phys.*, 60:1545–1557, 1974.

[246] F. H. Stillinger and A. Rahman. Molecular dynamics study of liquid water under high compression. *J. Chem. Phys.*, 61:4973–4980, 1975.

[247] W. L. Jorgensen. Transferable intermolecular potential functions for water, alcohols and ethers. Application to liquid water. *J. Am. Chem. Soc.*, 103:335–340, 1981.

[248] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulations of liquid water. *J. Chem. Phys*, 79(2):926–935, 1983.

[249] W. L. Jorgensen and J. Corky. Temperature dependence of TIP3P, SPC and TIP4P water from NPT Monte Carlo simulations: seeking temperature of maximum density. *J. Comput. Chem.*, 19(10):1179–1186, 1998.

[250] M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112(20):8910–8920, 2000.

[251] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91:6269–6271, 1987.

[252] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. Interaction models for water in relation to protein hydration. In B. Pullman, editor, *Intermolecular Forces*, pages 331–342. Reidel, Dordrecht, 1981.

[253] M. W. Mahoney and W. L. Jorgensen. Diffusion constant of the TIP5P model of liquid water. *J. Chem. Phys.*, 114(1):363–366, 2001.

[254] A. K. Soper. On the determination of the pair correlation function from liquid structure factor measurements. *Chem. Phys.*, 107:61–67, 1986.

[255] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159(1):98–103, 1967.

[256] R. W. Hockney. The potential calculation and some applications. *Methods in Computational Physics*, 9:136–211, 1970.

University of Cape Town

[257] K. Tasaki, S. McDonald, and J. W. Brady. Observations concerning the treatment of long-range interactions in molecular dynamics simulations. *J. Comput. Chem.*, 14(3):278–284, 1993.

[258] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the Cartesian equations of motion of a system with numerical constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[259] E. K. Wilson. Supercomputing strategies. *C&EN*, pages 46–50, 2001.

[260] B. R. Brooks and M. Hodoscek. Parallelization of CHARMM for MIMD machines. *Chemical Design Automation News*, 7(12):16–22, 1992.

[261] L. Kalé, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comp. Phys.*, 151:283–312, 1999.

[262] M. Mezei. Adaptive umbrella sampling: Self-consistent determination of the non-Boltzmann bias. *J. Comp. Phys.*, 68:237–248, 1987.

[263] R. W. W. Hooft, B. P. van Eijck, and J. Kroon. An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol. *J. Chem. Phys*, 97(9):6690–6694, 1992.

[264] E. M. Boczko and C. L. Brooks. Contant-temperature free energy surfaces for physical and chemical processes. *J. Phys. Chem.*, 97:4509–4513, 1993.

[265] C. Bartels and M. Karplus. Multidimensional adaptive umbrella sampling: Applications to main-chain and side-chain peptide conformations. *J. Comput. Chem.*, 18(12):1450–1462, 1997.

[266] C. Bartels and M. Karplus. Probability distributions for complex systems: adaptive umbrella sampling of the potential energy. *J. Phys. Chem.*, 102:865–880, 1998.

[267] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13(8):1011–1021, 1992.

[268] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, 16(11):1339–1350, 1995.

[269] S. Kumar, P. W. Payne, and M. Vãsquez. Method for free-energy calculations using iterative techniques. *J. Comput. Chem.*, 17(10):1269–1275, 1996.

University of Cape Town

[270] H. Senderowitz, F. Guarnieri, and W. C. Still. A smart Monte Carlo technique for free energy simulations of multiconformational molecules. Direct calculations of conformational populations of organic molecules. *J. Am. Chem. Soc.*, 117:8211–8219, 1995.

[271] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C: the art of scientific computing.* Cambridge University Press, 2nd edition, 1992.

[272] A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 61(23):2635–2638, 1988.

[273] A. M. Ferrenberg and R. H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63(12):1195–1198, 1989.

[274] P. Klewinghaus, B. P. van Eijck, M. L. C. E. Kouwijzer, and J. Kroon. Molecular dynamics study of conformational equilibria in aqueous D-glucose and D-galactose. *J. Molec. Struct. (Theochem)*, 395:289–295, 1997.

[275] M. M. Kuttel. Developing analytical tools for saccharides in condensed phases. Master's thesis, University of Cape Town, 1999.

[276] M. Dauchez, P. Derreumaux, and G. Vergoten. Vibrational molecular force field of model compounds with biologic interest. II. Harmonic dynamics of both anomers of glucose in the crystalline state. *J. Comput. Chem.*, 14(3):263–277, 1992.

[277] J. W. Brady. Molecular Dynamics simulations of carbohydrate molecules. *Adv. Biophys. Chem.*, 1:155–202, 1990.

[278] D. Lourdin, S. G. Ring, and P. Colonna. Study of plasticizer-oligomer and plasticizer-polymer interactions by dielectric analysis: maltose-glycerol and amylose-glycerol-water systems. *Carbohydr. Res.*, 306:551–558, 1998.

[279] A. R. Kirby, S. A. Clark, R. Parker, and A. C. Smith. The deformation and failure behaviour of wheat starch plasticized with water and polyols. *J. Mater. Sci.*, 28:5937–5942, 1993.

[280] C. A. Stortz. Dissacharide conformational maps: how adiabatic is an adiabatic map? *Carbohydr. Res.*, 322:77–86, 1999.

[281] T. A. Waigh, I. Hopkinson, A. M. Donald, M. F. Butler, F. Heidelbach, and C. Riekel. Analysis of the native structure of starch granules with X-ray microfocus diffraction. *Macromolecules*, 30:3813–3820, 1997.

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

University of Cape Town

[283] H.-C. H. Wu and A. Sarko. The double-helical molecular structure of crystalline b-amylose. *Carbohydr. Res.*, 61(7-25):27–40, 1978.

[284] A. Imberty, H. Chanzy, S. Pérez, A. Buléon, and V. Tran. The double-helical nature of the crystalline part of A-starch. *J. Mol. Biol.*, 201:365–378, 1988.

[285] M. St-Jacques, P. R. Sundararajan, K. J. Taylor, and R. H. Marchessault. Nuclear Magnetic Resonance and conformational studies on amylose and model compounds in dimethyl sulphoxide solution. *J. Am. Chem. Soc.*, 98(15):4386–4391, 1976.

[286] E. Evans and K. Richie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72:1541–1555, 1997.

[287] R. H. Marchessault and P. R. Sundararajan. Cellulose. In G. O. Aspinall, editor, *The Polysaccharides*, volume 2 of *Molecular Biology*. Academic Press, Orlando, Florida 32887, 1983.

[288] V. L. Finkenstadt and R. P. Millane. Crystal structure of *valonia* cellulose Iβ. *Macromolecules*, 31:7776–7783, 1998.

[289] P. Langan, Y. Nishiyama, and H. Chanzy. X-ray structure of mercerized cellulose II at 1 Å resolution. *Biomacromolecules*, 2:410–416, 2001.

[290] A. Sarko, J. Southwick, and J. Hayashi. Packing analysis of carbohydrates and polysaccharides. 7. Crystal structure of cellulose and III_1 and its relationship to other cellulose polymorphs. *Macromolecules*, 9:857–867, 1976.

[291] E. S. Gardiner and A. Sarko. Packing analysis of carbohydrates and polysaccharides. 16. The crystal structures of cellulose IV_1 and IV_2. *Can. J. Chem.*, 63:173–180, 1985.

[292] M. Koyama, W. Helbert, T. Imai, J. Sugiyama, and B. Henrissat. Parallel-up structure evidences the molecular directionality during biosynthesis of bacterial cellulose. *Proc. Natl. Acad. Sci. USA*, 94:9091–9095, 1997.

[293] S. Raymond, A. Heyraud, D. Tran Qui, Å. Kvick, and H. Chanzy. Crystal and molecular structure of β-D cellotetraose hemihydrate as a model of cellulose II. *Macromolecules*, 28:2096–2100, 1995.

[294] C. A. Stortz and A. S. Cerezo. Depicting the MM3 potential energy surfaces of trisaccharides by single contour maps: application to β-cellotriose and α-maltotriose. *Carbohydr. Res.*, 338:95–107, 2003.

[295] I. Tvaroška and L. Vaclavik. Stereochemistry of nonreducing disaccharides in solution. *Carbohydr. Res.*, 160:137–149, 1987.

University of Cape Town

[296] M. K. Dowd, P. J. Reilly, and A. D. French. Relaxed-residue conformational mapping of the three linkage bonds of isomaltose and gentiobiose with mm3 (92). *Biopolymers*, 34:625–638, 1994.

[297] R. B. Best, S. B. Fowler, J. L. Toca-Herrera, and J. Clarke. A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc. Natl. Acad. Sci. U.S.A.*, 99(19):12143–12148, 2002.

University of Cape Town

Appendix A

Computed α-D-glucose vibrational frequencies

Observed (Raman and infrared) frequencies (from Dauchez et al.[276]), computed vibrational frequencies (cm^{-1}) for the CSFF and PHLB force fields and an approximate description of each mode of vibration for α-D-glucose.

Observed frequencies		Calculated frequencies		
Raman	IR	CSFF	PHLB	Approximate potential energy distribution
41				
49	48			
59	60			
70	70			
76	77	74		τC_5C_6 +CCO +CCC +CC
85	88		87	$\tau C_5C_6 + \tau CC_{cyc}$ + CCC + OCC
92				
101	98	103	117	τC_5C_6 + CCO + CCC + CC
112	112			
	126			
134	136	126	139	CCC + CC + CCO + COC + τCC_{cyc} + τC_5C_6
142				
	152			

University of Cape Town

Table II. (continued)

Observed frequencies		Calculated frequencies		
Raman	IR	CSFF	PHLB	Approximate potential energy distribution
155				
184	188	186	186	$C_4C_5C_6 + C_1O_5C_5 + C_5C_6O_6 + CCO + CC_{cyc}$
234	237	241	286	$\tau C_6O + O_5C_5C_6$
	255			
277	270			
292	290	294	309	$\tau C_3O_3 + \tau C_5C_6 + O_5C_5C_6 + C_4C_5C_6 + COH + CO + CC$
312		309	326	$OCC + \tau C_3O_3 + \tau C_5C_6$
		318	327	$\tau CO + CCO$
		335	337	$CCO + \tau CO + \tau CC$
	350	355	357	$CCO + CC + CO + \tau CO$
369		370	379	$\tau CO + CCC + OCC + CC$
400	399	395	407	$\tau CO + OCO_5$
411	412	408	413	$CCO + CC + CO + C_6C_5O$
426	430	436	445	$CCO + CC + CO + \tau CO$
443	442	444	449	$CCO + OC_1O_5 + CC + \tau CO + CO$
		469	472	$CCC + CO + CCC + CCO + \tau CO$
497		513	519	$CCO + CCC + CC + CO + \tau CO$
542	549	552	555	$CCC + CO + CC + CO + \tau CO\ (C_1)$
559				
580	572	574	572	$CCO + CO + CC + CCC$
615	618	610	615	$O_1C_1O_5 + O_5CC + CO_5C + CCO + CCC + CO + CC$
657	645	682	683	$O_1C_1O_5 + O_5CC + CO_5C + CCO + CCC + CO + CC$
773	775	734	736	$OCC + CC$
		802	804	$O_5CC + CO + CCC + CCO$
842	840	845	857	$C_5C_6H + HC_6O + CO$
916	917			
		930	935	$CO + CC + CCC + COH$
1004	995			
1023		1033	1033	$CO\ (C_6) + C_6OH + HC_6O + HC_6H$
1055	1045	1063	1065	$CC + CO + HCC + HCO$
1070		1072	1073	$CO + CC + HCC + CCO$
1078	1078	1077	1081	$CO + HCO$ (c3)

University of Cape Town

Table II. (continued)

Observed frequencies		Calculated frequencies		
Raman	**IR**	**CSFF**	**PHLB**	**Approximate potential energy distribution**
1105	1105	1100	1101	CCH + HCO + CO + CC
1113	1113	1109	1111	HCO (c1) + HCC + CO + CC + CCO
1123				
1133	1145	1142	1146	HCC + CC
1153		1156	1163	CCO + CCC + CC + CO + HCO + COH + HCC
1204	1202	1186	1188	HCC + OCC + HCO + CO
1226	1222	1217	1218	HCO + C_1O_5 + C_5C_6 + CO + CCH + CCC
		1248	1246	C_1O_5 +c5O_5 + HC_6O +c5c6H + CCH +HCO
		1264	1265	CCH +COH
1279	1275	1279	1282	COH +O_5C_1H +CCH + CCO +C_1O +C_1O_5 +C_5C_6H
		1291	1293	COH +O_5C_1H +CCH + CCO +C_1O +C_1O_5 +C_5C_6H
1296	1295	1297	1301	COH +O_5C_1H +CCH + CCO +C_1O
		1324	1324	COH + C_1OH +O_5C_1H +CCH + CCO
1332	1340	1331	1333	COH +O_5C_1H +CCH + CCO
		1338	1341	COH + CCH + HCO
1347		1351	1353	C_6OH +HC_6H
1372	1375			
1407	1407	1389	1391	COH + HCO + CO
1440	1440	1422	1427	C_5C_6H + HC_6H + C_6OH
1460	1460	1495	1497	HC_6H + HC_6O + C_6O_6 + C_5C_6
		1515	1515	CO + CC + HCO + COH
		1539	1539	CC + HCO + CO + CCH
		1592	1592	CC + CO + HCO + HCC + CCC
		1631	1628	CO + CC + HCO + OCC + CCC
		1641	1641	CC + CO + HCO + HCC + OCC
		1739	1738	CO + CC + CCC + HCO + CCH
		1768	1766	CCC + CO + CC + HCO + OCC
2877	2885	2899	2899	C_6-H
2890	2898	2926	2926	C-H
2914	2920	2931	2931	C-H
		2934	2934	C-H
		2938	2939	C-H

University of Cape Town

Table II. (continued)

Observed frequencies		Calculated frequencies		
Raman	IR	CSFF	PHLB	Approximate potential energy distribution
2946	2947	2945	2945	C-H
2959		2958	2959	C-H
	3100	3383	3685	O-H
3385	broad	3385	3688	O-H
		3387	3689	O-H
3405	band	3389	3691	O-H
3415	3500	3393	3694	O-H

University of Cape Town

Appendix B

CSFF CHARMM Topology File

University of Cape Town

```
************************************************************************
* CHARMM27 Carbohydrate topology file (CSFF_top.inp)
*    for pyranose simulations
* Michelle Kuttel, J. W. Brady and Kevin J. Naidoo
* Aug/Sep 2001
* Modified from PHLB (Palma-Himmel-Liang-Brady) force field parameters:
*     added CPS atom type and altered primary alcohol
*     dihedral force constants from CTS values
*     to lower energy the barriers to primary alcohol rotation
* Comments to Kevin J. Naidoo email:knaidoo,science.uct.ac.za
************************************************************************
*
27    1
!
!references
!
!PHLB force field precursor
!
!R. Palma and M. E. Himmel and G. Liang and J. W. Brady. "Molecular
!Mechanics Studies of Cellulases" in "Glycosyl Hydrolases in Biomass
!Conversion: ACS Symposium Series", published by ACS, 2001, editor
!M. E. Himmel
!
!HBFB force field precursor
!
!S. N. Ha and A. Giammona and M. Field and J. W. Brady, "A revised
```

```
!potential-energy surface for molecular studies of carbohydrates",
!Carbohydr. Res., 1988,180,207-221

MASS     4 HT     1.00800 ! TIP3P water hydrogen
MASS    56 OT    15.99940 ! TIP3P water oxygen
MASS    60 OSPC  15.99940 ! SPC water oxygen
MASS    61 HSPC   1.00800 ! SPC water hydrogen
MASS    73 HAS    1.00800 ! sugar aliphatic hydrogen
MASS    74 HOS    1.00800 ! sugar hydroxyl hydrogen
MASS    75 CTS   12.01100 ! sugar aliphatic carbon
MASS    76 OHS   15.99940 ! sugar hydroxy oxygen
MASS    78 CBS   12.01100 ! C1 in beta sugars
MASS    79 CPS   12.01100 ! primary alcohol carbon (CT3)
MASS    77 OES   15.99940 ! sugar ring oxygen

AUTOGENERATE angles dihedrals
! DEFAULTS for patching FIRST residue (with NTER) and LAST (with CTER)
DEFA FIRS NTER LAST CTER

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!
!                          CARBOHYDRATES
!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

!_____________________________ Residues _______________________________
RESI AGLC    0.00000     ! 4C1 alpha-D-glucopyranose monomer
GROU                     !
ATOM C1   CTS     0.200  !          H61
ATOM H1   HAS     0.090  !           |
ATOM O1   OHS    -0.660  !      H62-C6-O6-HO6
ATOM HO1  HOS     0.430  !           |
ATOM C5   CTS     0.250  !      H4  C5--O5   O1-HO1
ATOM H5   HAS     0.090  !       \ / \   \  /
ATOM O5   OES    -0.400  ! HO4-O4-C4  H5  C1-H1
GROU                     !         \     /
ATOM C2   CTS     0.140  !       H3-C3--C2-O2-HO2
ATOM H2   HAS     0.090  !           |   |
```

University of Cape Town

```
ATOM O2   OHS    -0.660  !          O3  H2
ATOM HO2  HOS     0.430  !          |
GROU                     !          HO3
ATOM C3   CTS     0.140  !
ATOM H3   HAS     0.090  !
ATOM O3   OHS    -0.660  !
ATOM HO3  HOS     0.430  !
GROU                     !
ATOM C4   CTS     0.140  !
ATOM H4   HAS     0.090  !
ATOM O4   OHS    -0.660  !
ATOM HO4  HOS     0.430  !
GROU                     !
ATOM C6   CPS     0.050  !
ATOM H61  HAS     0.090  !
ATOM H62  HAS     0.090  !
ATOM O6   OHS    -0.660  !
ATOM HO6  HOS     0.430  !
BOND C1   O1      C1  H1      O1  HO1     C1  O5      C1  C2
BOND C2   H2      C2  O2      O2  HO2     C2  C3      C3  H3
BOND C3   O3      O3  HO3     C3  C4      C4  H4      C4  O4
BOND O4   HO4     C4  C5      C5  H5      C5  C6      C6  H61
BOND C6   H62     C6  O6      O6  HO6     C5  O5
DONO HO1 O1
DONO HO2 O2
DONO HO3 O3
DONO HO4 O4
DONO HO6 O6
ACCE O1
ACCE O2
ACCE O3
ACCE O4
ACCE O5
ACCE O6
!    I    J    K    L      R(IK)   T(IKJ)    PHI    T(JKL)   R(KL)
IC   O1   C2  *C1   H1     1.3889  109.35 -122.69  108.98   1.0950
IC   O1   O5  *C1   C2     1.3889  111.55 -121.57  110.06   1.5340
IC   O2   C3  *C2   H2     1.4154  112.27 -118.21  108.23   1.0919
```

University of Cape Town

```
IC   O2   C1   *C2   C3     1.4154  110.87 -125.56  111.08   1.5253
IC   O3   C4   *C3   H3     1.4157  110.61  120.65  108.81   1.1068
IC   O3   C2   *C3   C4     1.4157  108.09  120.77  109.86   1.5177
IC   O4   C5   *C4   H4     1.4252  110.90 -120.61  108.35   1.1024
IC   O4   C3   *C4   C5     1.4252  108.31 -122.08  111.17   1.5287
IC   C6   O5   *C5   H5     1.5099  108.10  118.69  109.65   1.1042
IC   C6   C4   *C5   O5     1.5099  111.57  119.10  108.69   1.4274
IC   O6   H62  *C6   H61    1.4132  110.47 -120.32  107.85   1.0945
IC   O6   C5   *C6   H62    1.4132  110.45 -121.53  108.99   1.0959
IC   O5   C1    C2   C3     1.4254  110.06   54.09  111.08   1.5253
IC   C1   C2    C3   C4     1.5340  111.08  -51.23  109.86   1.5177
IC   C2   C3    C4   C5     1.5253  109.86   53.25  111.17   1.5288
IC   C3   C4    C5   O5     1.5177  111.17  -57.46  108.69   1.4274
IC   C4   C5    O5   C1     1.5288  108.69   62.25  113.77   1.4254
IC   C5   O5    C1   C2     1.4274  113.77  -60.97  110.06   1.5340
IC   C4   C5    C6   O6     1.5287  111.57 -170.28  110.45   1.4132
IC   O5   C1    O1   HO1    1.4254  111.55   74.87  107.83   0.9684
IC   C1   C2    O2   HO2    1.5340  110.87 -100.51  112.13   0.9638
IC   C2   C3    O3   HO3    1.5253  108.09 -165.88  112.08   0.9730
IC   C3   C4    O4   HO4    1.5177  108.31  134.18  106.97   0.9713
IC   C5   C6    O6   HO6    1.5099  110.44 -143.88  107.72   0.9641
PATC FIRS NONE LAST NONE


RESI GLDO    0.00000     ! 4C1 6-deoxy alpha-D-glucopyranose monomer
GROU                     !
ATOM C1   CTS     0.200  !          H61
ATOM H1   HAS     0.090  !           |
ATOM O1   OHS    -0.660  !      H62-C6-H63
ATOM HO1  HOS     0.430  !           |
ATOM C5   CTS     0.250  !      H4  C5--O5   O1-HO1
ATOM H5   HAS     0.090  !       \ / \   \  /
ATOM O5   OES    -0.400  ! HO4-O4-C4  H5  C1-H1
GROU                     !         \     /
ATOM C2   CTS     0.140  !       H3-C3--C2-O2-HO2
ATOM H2   HAS     0.090  !           |   |
ATOM O2   OHS    -0.660  !          O3  H2
ATOM HO2  HOS     0.430  !           |
```

University of Cape Town

```
GROU                        !          HO3
ATOM C3   CTS      0.140
ATOM H3   HAS      0.090
ATOM O3   OHS     -0.660
ATOM HO3  HOS      0.430
GROU
ATOM C4   CTS      0.140
ATOM H4   HAS      0.090
ATOM O4   OHS     -0.660
ATOM HO4  HOS      0.430
GROU
ATOM C6   CPS     -0.270
ATOM H61  HAS      0.090
ATOM H62  HAS      0.090
ATOM H63  HAS      0.090
BOND C1   O1       C1   H1       O1   HO1      C1   O5       C1   C2
BOND C2   H2       C2   O2       O2   HO2      C2   C3       C3   H3
BOND C3   O3       O3   HO3      C3   C4       C4   H4       C4   O4
BOND O4   HO4      C4   C5       C5   H5       C5   C6       C6   H61
BOND C6   H62      C6   H63      C5   O5
DONO HO1 O1
DONO HO2 O2
DONO HO3 O3
DONO HO4 O4
ACCE O1
ACCE O2
ACCE O3
ACCE O4
ACCE O5
!    I    J    K    L      R(IK)   T(IKJ)    PHI    T(JKL)   R(KL)
IC   O1   C2   *C1  H1     1.3889  109.35 -122.69  108.98   1.0950
IC   O1   O5   *C1  C2     1.3889  111.55 -121.57  110.06   1.5340
IC   O2   C3   *C2  H2     1.4154  112.27 -118.21  108.23   1.0919
IC   O2   C1   *C2  C3     1.4154  110.87 -125.56  111.08   1.5253
IC   O3   C4   *C3  H3     1.4157  110.61  120.65  108.81   1.1068
IC   O3   C2   *C3  C4     1.4157  108.09  120.77  109.86   1.5177
IC   O4   C5   *C4  H4     1.4252  110.90 -120.61  108.35   1.1024
IC   O4   C3   *C4  C5     1.4252  108.31 -122.08  111.17   1.5287
```

University of Cape Town

```
IC   C6   O5   *C5   H5     1.5099  108.10  118.69  109.65   1.1042
IC   C6   C4   *C5   O5     1.5099  111.57  119.10  108.69   1.4274
IC   H63  H62  *C6   H61    1.4132  110.47 -120.32  107.85   1.0945
IC   H63  C5   *C6   H62    1.4132  110.45 -121.53  108.99   1.0959
IC   O5   C1    C2   C3     1.4254  110.06   54.09  111.08   1.5253
IC   C1   C2    C3   C4     1.5340  111.08  -51.23  109.86   1.5177
IC   C2   C3    C4   C5     1.5253  109.86   53.25  111.17   1.5288
IC   C3   C4    C5   O5     1.5177  111.17  -57.46  108.69   1.4274
IC   C4   C5    O5   C1     1.5288  108.69   62.25  113.77   1.4254
IC   C5   O5    C1   C2     1.4274  113.77  -60.97  110.06   1.5340
IC   C4   C5    C6   H63    1.5287  111.57 -170.28  110.45   1.4132
IC   O5   C1    O1   HO1    1.4254  111.55   74.87  107.83   0.9684
IC   C1   C2    O2   HO2    1.5340  110.87 -100.51  112.13   0.9638
IC   C2   C3    O3   HO3    1.5253  108.09 -165.88  112.08   0.9730
IC   C3   C4    O4   HO4    1.5177  108.31  134.18  106.97   0.9713
PATC FIRS NONE LAST NONE

RESI AXYL   0.00000     ! 4C1 alpha-D-xylopyranose
GROU
ATOM C1   CTS     0.200  !
ATOM H1   HAS     0.090  !
ATOM O1   OHS    -0.660  !
ATOM HO1  HOS     0.430  !
ATOM C5   CTS     0.160  !
ATOM H51  HAS     0.090  !      H51   H52
ATOM H52  HAS     0.090  !        \ /
ATOM O5   OES    -0.400  !      H4  C5--O5   O1-HO1
GROU                     !       \ /     \  /
ATOM C2   CTS     0.140  ! HO4-O4-C4       C1-H1
ATOM H2   HAS     0.090  !         \     /
ATOM O2   OHS    -0.660  !       H3-C3--C2-O2-HO2
ATOM HO2  HOS     0.430  !          |   |
GROU                     !          O3  H2
ATOM C3   CTS     0.140  !          |
ATOM H3   HAS     0.090  !          HO3
ATOM O3   OHS    -0.660  !
ATOM HO3  HOS     0.430  !
GROU                     !
```

University of Cape Town

```
ATOM C4   CTS     0.140  !
ATOM H4   HAS     0.090  !
ATOM O4   OHS    -0.660  !
ATOM HO4  HOS     0.430  !
BOND C1   O1      C1   H1      O1   HO1     C1   O5      C1   C2
BOND C2   H2      C2   O2      O2   HO2     C2   C3      C3   H3
BOND C3   O3      O3   HO3     C3   C4      C4   H4      C4   O4
BOND O4   HO4     C4   C5      C5   H51     C5   H52       C5   O5
DONO BLNK HO1
DONO BLNK HO2
DONO BLNK HO3
DONO BLNK HO4
ACCE O1
ACCE O2
ACCE O3
ACCE O4
ACCE O5
!    I    J    K    L      R(IK)   T(IKJ)    PHI   T(JKL)   R(KL)
IC   O1   C2  *C1   H1     1.3889  109.35 -122.69  108.98   1.0950
IC   O1   O5  *C1   C2     1.3889  111.55 -121.57  110.06   1.5340
IC   O2   C3  *C2   H2     1.4154  112.27 -118.21  108.23   1.0919
IC   O2   C1  *C2   C3     1.4154  110.87 -125.56  111.08   1.5253
IC   O3   C4  *C3   H3     1.4157  110.61  120.65  108.81   1.1068
IC   O3   C2  *C3   C4     1.4157  108.09  120.77  109.86   1.5177
IC   O4   C5  *C4   H4     1.4252  110.90 -120.61  108.35   1.1024
IC   O4   C3  *C4   C5     1.4252  108.31 -122.08  111.17   1.5287
IC   H52  O5  *C5   H51    1.0813  108.10  118.69  109.65   1.1042
IC   H52  C4  *C5   O5     1.0813  111.57  119.10  108.69   1.4274
IC   O5   C1   C2   C3     1.4254  110.06   54.09  111.08   1.5253
IC   C1   C2   C3   C4     1.5340  111.08  -51.23  109.86   1.5177
IC   C2   C3   C4   C5     1.5253  109.86   53.25  111.17   1.5288
IC   C3   C4   C5   O5     1.5177  111.17  -57.46  108.69   1.4274
IC   C4   C5   O5   C1     1.5288  108.69   62.25  113.77   1.4254
IC   C5   O5   C1   C2     1.4274  113.77  -60.97  110.06   1.5340
IC   O5   C1   O1   HO1    1.4254  111.55   74.87  107.83   0.9684
IC   C1   C2   O2   HO2    1.5340  110.87 -100.51  112.13   0.9638
IC   C2   C3   O3   HO3    1.5253  108.09 -165.88  112.08   0.9730
IC   C3   C4   O4   HO4    1.5177  108.31  134.18  106.97   0.9713
```

University of Cape Town

```
PATC FIRS NONE LAST NONE

!------------------------ Patches ---------------------------------

PRES BETA         0.0    ! patch to make the C1 group equitorial (beta)
GROU                     !
ATOM C1   CBS     0.200  !   <--- this is the changed ATOM
IC   O1   C2  *C1   H1     1.3890  105.75  114.54  108.17   1.0950
IC   O1   O5  *C1   C2     1.3890  111.55  117.06  110.06   1.5340

PRES C2          0.0   ! patch to make the C2 group axial
IC   O2   C3  *C2   H2     1.4215  111.64  128.12  105.85   1.0010
IC   O2   C1  *C2   C3     1.4215  106.16  120.87  109.90   1.5186

PRES C3          0.0   ! patch to make the C3 group axial
IC   O3   C4  *C3   H3     1.4157  110.61 -120.65  108.81   1.1068
IC   O3   C2  *C3   C4     1.4157  108.09 -120.77  109.86   1.5177

PRES C4          0.0   ! patch to make the C4 group axial
IC   O4   C5  *C4   H4     1.4215  107.18  119.92  110.37   1.1004
IC   O4   C3  *C4   C5     1.4215  110.06  118.59  111.17   1.5289

PRES C5          0.0   ! patch to make C5 group axial, as in an L-sugar
IC   C6   O5  *C5   H5     1.5099  108.10 -118.69  109.65   1.1042
IC   C6   C4  *C5   O5     1.5099  111.57 -119.10  108.69   1.4274


PRES GL14         0.0    ! 1-4 linkage when the reducing ring
GROUP                    ! is an alpha-sugar
ATOM 1c5 CTS     0.1     !
ATOM 1h5 HAS     0.1     !
ATOM 1o5 OES    -0.4     !
ATOM 1H1 HAS     0.1     !
ATOM 1C1 CTS     0.30    !
ATOM 1O1 OES    -0.4     !
ATOM 2C4 CTS     0.10    !
ATOM 2H4 HAS     0.10    !
DELETE ATOM 1HO1
```

University of Cape Town

```
DELETE ATOM 2O4
DELETE ATOM 2HO4
BOND 1O1 2C4
ANGL  1C1 1O1 2C4   1O1 2C4 2C3
ANGL  1O1 2C4 2C5   1O1 2C4 2H4
DIHE  1O5 1C1 1O1 2C4       1C1 1O1 2C4 2c5
DIHE  1C2 1C1 1O1 2C4       1C1 1O1 2C4 2c3
DIHE  1h1 1C1 1O1 2C4       1C1 1O1 2C4 2h4
DIHE  1O1 2C4 2C5 2C6       1O1 2C4 2C5 2H5
DIHE  1O1 2C4 2C5 2O5       1O1 2C4 2C3 2C2
DIHE  1O1 2C4 2C3 2H3       1O1 2C4 2C3 2O3
IC    1H1  1C1  1O1  2C4    1.0882  110.64    4.85  117.87   1.4195
IC    1C1  1O1  2C4  2H4    1.4138  117.87   13.26  110.59   1.0977
IC    2C3  2C5 *2C4  1O1    1.5265  109.15 -118.85  110.91   1.4195
IC    2C5  1O1 *2C4  2H4    1.5340  110.91  121.00  110.59   1.0977
IC    1O1  2C4  2C5  2O5    1.4195  110.91 -178.47  109.01   1.4229
IC    1O1  2C4  2C3  2C2    1.4195  107.98  174.59  110.07   1.5280

PRES GB14        0.0    ! 1-4 linkage when the reducing ring
GROUP                   ! is a beta-sugar
ATOM 1C5 CTS    0.1     !
ATOM 1H5 HAS    0.1     !
ATOM 1O5 OES   -0.4     !
ATOM 1H1 HAS    0.1     !
ATOM 1C1 CBS    0.30    !
ATOM 1O1 OES   -0.4     !
ATOM 2C4 CTS    0.10    !
ATOM 2H4 HAS    0.10    !
DELETE ATOM 1HO1
DELETE ATOM 2O4
DELETE ATOM 2HO4
BOND 1O1 2C4
ANGL  1C1 1O1 2C4   1O1 2C4 2C3
ANGL  1O1 2C4 2C5   1O1 2C4 2H4
DIHE  1o5 1C1 1O1 2C4       1C1 1O1 2C4 2c5
DIHE  1c2 1C1 1O1 2C4       1C1 1O1 2C4 2c3
DIHE  1h1 1C1 1O1 2C4       1C1 1O1 2C4 2h4
DIHE  1O1 2C4 2c5 2O5       1O1 2C4 2c5 2C6
```

University of Cape Town

```
DIHE  1O1 2C4 2c5 2H5       1O1 2C4 2c3 2C2
DIHE  1O1 2C4 2c3 2O3       1O1 2C4 2c3 2H3
IC    1H1  1C1  1O1  2C4    1.0882  110.64    4.85  117.87   1.4195
IC    1C1  1O1  2C4  2H4    1.4138  117.87   13.26  110.59   1.0977
IC    2C3  2C5 *2C4  1O1    1.5265  109.15 -118.85  110.91   1.4195
IC    2C5  1O1 *2C4  2H4    1.5340  110.91  121.00  110.59   1.0977
IC    1O1  2C4  2C5  2O5    1.4195  110.91 -178.47  109.01   1.4229
IC    1O1  2C4  2C3  2C2    1.4195  107.98  174.59  110.07   1.5280

!XXXXXXXXXXXXXXXXXXXXXX XYL TO XYL  JOINT XXXXXXXXXXXXXXXXXXXXXXXXX
PRES XY14        0.0     !  1-4 linkage for dixylose
GROUP                    !
ATOM 1C5  CTS    0.0     !
ATOM 1H51 HAS    0.1     !
ATOM 1H52 HAS    0.1     !
ATOM 1O5  OES   -0.4     !
ATOM 1H1  HAS    0.1     !
ATOM 1C1  CTS    0.30    !
ATOM 1O1  OES   -0.4     !
ATOM 2C4  CTS    0.10    !
ATOM 2H4  HAS    0.10    !
DELETE ATOM 1HO1
DELETE ATOM 2O4
DELETE ATOM 2HO4
BOND 1O1 2C4
ANGL  1C1 1O1 2C4   1O1 2C4 2C3
ANGL  1O1 2C4 2C5   1O1 2C4 2H4
DIHE  1O5 1C1 1O1 2C4       1C1 1O1 2C4 2c5
DIHE  1C2 1C1 1O1 2C4       1C1 1O1 2C4 2c3
DIHE  1H1 1C1 1O1 2C4       1C1 1O1 2C4 2h4
DIHE  1O1 2C4 2C5 2H51      1O1 2C4 2C5 2H52
DIHE  1O1 2C4 2C5 2O5       1O1 2C4 2C3 2C2
DIHE  1O1 2C4 2C3 2H3       1O1 2C4 2C3 2O3
IC    1H1  1C1  1O1  2C4    1.0882  110.64    4.85  117.87   1.4195
IC    1C1  1O1  2C4  2H4    1.4138  117.87   13.26  110.59   1.0977
IC    2C3  2C5 *2C4  1O1    1.5265  109.15 -118.85  110.91   1.4195
IC    2C5  1O1 *2C4  2H4    1.5340  110.91  121.00  110.59   1.0977
IC    1O1  2C4  2C5  2O5    1.4195  110.91 -178.47  109.01   1.4229
```

University of Cape Town

```
IC    1O1  2C4  2C3  2C2    1.4195  107.98  174.59  110.07   1.5280

PRES 1C4        0.0   ! patch to make the ring a 1c4 conformation
IC   O5   C1   C2   C3     1.4254  110.06  -57.64  111.08   1.5253
IC   C1   C2   C3   C4     1.5340  111.08   56.15  109.86   1.5177
IC   C2   C3   C4   C5     1.5253  109.86  -52.90  111.17   1.5288
IC   C3   C4   C5   O5     1.5177  111.17   49.15  108.69   1.4274
IC   C4   C5   O5   C1     1.5288  108.69  -52.02  113.77   1.4254
IC   C5   O5   C1   C2     1.4274  113.77   56.77  110.06   1.5340

PRES GB16     0.0       !Beta 1-6 linkage
GROUP
ATOM 1H1  HAS  0.1
ATOM 1C1  CTS  0.25
ATOM 1O1  OES -0.4
ATOM 1C5  CTS  0.1
ATOM 1H5  HAS  0.1
ATOM 1O5  OES -0.4
ATOM 2C6  CPS  0.05
ATOM 2H61 HAS  0.1
ATOM 2H62 HAS  0.1
BOND 1O1 2C6
DELETE ATOM 1HO1
DELETE ATOM 2O6
DELETE ATOM 2HO6
ANGL 1C1  1O1  2C6    1O1  2C6  2C5
ANGL 1O1  2C6  2H61   1O1  2C6  2H62
DIHE  1O5 1C1 1O1 2C6       1C1 1O1 2C6 2C5
DIHE  1C2 1C1 1O1 2C6       1C1 1O1 2C6 2H62
DIHE  1H1 1C1 1O1 2C6       1C1 1O1 2C6 2H61
DIHE  1O1 2C6 2C5 2O5       1O1 2C6 2C5 2H5
DIHE  1O1 2C6 2C5 2C4
IC    1H1  1C1  1O1  2C6    1.0950  109.00  180.00  113.00   1.4000
IC    1C1  1O1  2C6  2C5    1.3900  113.00  180.00  110.40   1.5099
IC    1C1  1C2  1C3  1C4    1.5340  111.08  -51.23  109.86   1.5177
IC    1O5  1C1  1O1  2C6    1.4254  111.55  180.00  113.00   1.4000
IC    1O1  2C6  2C5  2C4    1.4000  110.40  180.00  111.57   1.5287
IC    1O1  2H62 *2C6 2H61   1.4132  110.47 -120.32  107.85   1.0945
```

University of Cape Town

```
IC    1O1  2C5  *2C6 2H62   1.4132  110.45 -121.53  108.99   1.0959

PRES GL11        0.0     ! 1-1 linkage when the reducing ring
GROUP                    ! is an alpha-sugar
ATOM 1C5 CTS    0.10     !
ATOM 1H5 HAS    0.10     !
ATOM 1O5 OES   -0.40     !
ATOM 1H1 HAS    0.10     !
ATOM 1C1 CTS    0.30     !
ATOM 1O1 OES   -0.40     !
ATOM 2C1 CTS    0.30     !
ATOM 2H1 HAS    0.10     !
ATOM 2C5 CTS    0.10     !
ATOM 2H5 HAS    0.10     !
ATOM 2O5 OES   -0.40     !
DELETE ATOM 1HO1
DELETE ATOM 2O1
DELETE ATOM 2HO1
BOND 1O1 2C1
ANGL  1C1 1O1 2C1   1O1 2C1 2O5
ANGL  1O1 2C1 2C2   1O1 2C1 2H1
DIHE  1O5 1C1 1O1 2C1       1C1 1O1 2C1 2O5
DIHE  1C2 1C1 1O1 2C1       1C1 1O1 2C1 2C2
DIHE  1H1 1C1 1O1 2C1       1C1 1O1 2C1 2H1
DIHE  1O1 2C1 2C2 2O2       1O1 2C1 2C2 2H2
DIHE  1O1 2C1 2C2 2C3       1O1 2C1 2O5 2C5
IC    1H1  1C1  1O1  2C1    1.0882  110.64    4.85  117.87   1.4195
IC    1C1  1O1  2C1  2H1    1.4138  117.87   13.26  110.59   1.0977
IC    2O5  2C2 *2C1  1O1    1.5265  109.15 -118.85  110.91   1.4195
IC    2C2  1O1 *2C1  2H1    1.5340  110.91  121.00  110.59   1.0977
IC    1O1  2C1  2C2  2C3    1.4195  110.91 -178.47  109.01   1.4229
IC    1O1  2C1  2O5  2C5    1.4195  107.98  174.59  110.07   1.5280

END
```

University of Cape Town

Appendix C

CSFF CHARMM Parameter File

```
************************************************************************
* Carbohydrate parameter file CSFF_parm.inp
*     for pyranose simulations
* Michelle Kuttel, J. W. Brady and Kevin J. Naidoo
* Aug/Sep 2001
* Modified from PHLB (Palma-Himmel-Liang-Brady) force field parameters:
*     added CPS atom type and altered primary alcohol
*     dihedral force constants from CTS values
*     to lower energy the barriers to primary alcohol rotation
* Comments to Kevin J. Naidoo email:knaidoo,science.uct.ac.za
************************************************************************
*
!
!references
!
!This Force Field (CSFF)
!
!M.Kuttel, J. W. Brady, K. J. Naidoo
!
!PHLB force field precursor
!
!R. Palma and M. E. Himmel and G. Liang and J. W. Brady. "Molecular
!Mechanics Studies of Cellulases" in "Glycosyl Hydrolases in Biomass
!Conversion: ACS Symposium Series", published by ACS, 2001, editor
!M. E. Himmel
!
```

University of Cape Town

```
!HBFB force field precursor
!
!S. N. Ha and A. Giammona and M. Field and J. W. Brady, "A revised
!potential-energy surface for molecular studies of carbohydrates",
!Carbohydr. Res., 1988,180,207-221
! NOTE: messages about multiple dihedral terms on reading this file are
! normal and signify that the multiple dihedral terms have in fact
! correctly been read (see CHARMM docs - parmfile.doc)
BONDS
HOS  OHS    460.5000            0.9595
HAS  CTS    335.6034            1.1000
HAS  CPS    335.6034            1.1000
HAS  CBS    335.6034            1.1052
CTS  OHS    384.0792            1.4066
CPS  OHS    384.0792            1.4066
CBS  OHS    384.0792            1.3932
CTS  CTS    325.5297            1.5066
CTS  CPS    325.5297            1.5066
CBS  CTS    325.5297            1.5074
CTS  OES    385.3133            1.4165
CBS  OES    385.3133            1.4202
HSPC OSPC   450.0               1.0        ! SPC Geometry
HSPC HSPC     0.0               1.6329931  ! SPC Geometry (for SHAKE)
HT   OT     450.0               0.9572     ! TIP3P geometry
HT   HT       0.0               1.5139     ! TIP3P geom(for SHAKE)
THETAS
HAS  CTS  CTS          42.9062          109.7502
HAS  CBS  CTS          42.9062          109.7502
HAS  CTS  CBS          42.9062          109.7502
HAS  CTS  CPS          42.9062          109.7502
HAS  CPS  CTS          42.9062          109.7502
OHS  CTS  CTS         112.2085          107.6019
OHS  CBS  CTS         112.2085          107.6019
OHS  CTS  CBS         112.2085          107.6019
OHS  CPS  CTS         112.2085          107.6019
OHS  CTS  CPS         112.2085          107.6019
HOS  OHS  CTS          57.5478          109.1722
HOS  OHS  CBS          57.5478          109.1722
```

University of Cape Town

```
HOS  OHS  CPS          57.5478          109.1722
HAS  CTS  HAS          36.8220          106.1784
HAS  CPS  HAS          36.8220          106.1784
HAS  CTS  OHS          52.5070          109.3850
HAS  CBS  OHS          52.5070          109.3850
HAS  CPS  OHS          52.5070          109.3850
HT   OT   HT           55.0             104.52     ! FROM TIP3P
HSPC OSPC HSPC         55.0             109.47122  ! FROM SPC
HAS  CTS  OES          62.2500          106.4025
HAS  CBS  OES          62.2500          106.4025
CTS  CTS  CTS         167.3535          110.6156
CBS  CTS  CTS         167.3535          110.6156
CPS  CTS  CTS         167.3535          110.6156
CTS  CTS  OES         169.0276          108.3759
CBS  CTS  OES         169.0276          108.3759
CPS  CTS  OES         169.0276          108.3759
CTS  CBS  OES         169.0276          108.3759
CTS  OES  CTS          92.5892          111.5092
CBS  OES  CTS          92.5892          111.5092
OES  CTS  OHS          74.2586          115.7322
OES  CBS  OHS          74.2586          110.3385
OES  CTS  OES          37.4370          112.1882
OES  CBS  OES          37.4370          106.9160
DIHEDRALS
CTS  CTS  CTS  CTS     -1.0683    1     0.0
CTS  CTS  CTS  CTS     -0.5605    2     0.0
CTS  CTS  CTS  CTS      0.1955    3     0.0
CBS  CTS  CTS  CTS     -1.0683    1     0.0
CBS  CTS  CTS  CTS     -0.5605    2     0.0
CBS  CTS  CTS  CTS      0.1955    3     0.0
CPS  CTS  CTS  CTS     -1.0683    1     0.0 !MK - CTS value unchanged
CPS  CTS  CTS  CTS     -0.5605    2     0.0 !MK - CTS value unchanged
CPS  CTS  CTS  CTS      0.1955    3     0.0 !MK - CTS value unchanged
CTS  CTS  CTS  OES     -1.2007    1     0.0
CTS  CTS  CTS  OES     -0.3145    2     0.0
CTS  CTS  CTS  OES     -0.0618    3     0.0
CBS  CTS  CTS  OES     -1.2007    1     0.0
CBS  CTS  CTS  OES     -0.3145    2     0.0
```

University of Cape Town

```
CBS  CTS  CTS  OES     -0.0618    3      0.0
CPS  CTS  CTS  OES     -1.2007    1      0.0 !MK - CTS value unchanged
CPS  CTS  CTS  OES     -0.3145    2      0.0 !MK - CTS value unchanged
CPS  CTS  CTS  OES     -0.0618    3      0.0 !MK - CTS value unchanged
CTS  CTS  CBS  OES     -1.2007    1      0.0
CTS  CTS  CBS  OES     -0.3145    2      0.0
CTS  CTS  CBS  OES     -0.0618    3      0.0
CTS  CTS  CPS  OES     -1.2007    1      0.0 !MK - CTS value unchanged
CTS  CTS  CPS  OES     -0.3145    2      0.0 !MK - CTS value unchanged
CTS  CTS  CPS  OES     -0.0618    3      0.0 !MK - CTS value unchanged
CTS  CTS  CTS  OHS     -1.9139    1      0.0
CTS  CTS  CTS  OHS      0.2739    2      0.0
CTS  CTS  CTS  OHS     -0.0340    3      0.0
CTS  CTS  CBS  OHS     -1.9139    1      0.0
CTS  CTS  CBS  OHS     -0.3739    2      0.0
CTS  CTS  CBS  OHS     -0.0340    3      0.0
CTS  CTS  CPS  OHS     -1.7139    1      0.0 !MK - changed from -1.9139
CTS  CTS  CPS  OHS     -1.0239    2      0.0 !MK - changed from -0.3739
CTS  CTS  CPS  OHS     -0.0340    3      0.0 !MK - CTS value unchanged
CBS  CTS  CTS  OHS     -1.9139    1      0.0
CBS  CTS  CTS  OHS     -0.3739    2      0.0
CBS  CTS  CTS  OHS     -0.0340    3      0.0
CPS  CTS  CTS  OHS     -1.9139    1      0.0 !MK - CTS value unchanged
CPS  CTS  CTS  OHS     -0.3739    2      0.0 !MK - CTS value unchanged
CPS  CTS  CTS  OHS     -0.0340    3      0.0 !MK - CTS value unchanged
CTS  OES  CTS  CTS     -0.8477    1      0.0
CTS  OES  CTS  CTS     -0.3018    2      0.0
CTS  OES  CTS  CTS      0.3763    3      0.0
CTS  OES  CTS  CPS     -0.8477    1      0.0 !MK - CTS value unchanged
CTS  OES  CTS  CPS     -0.3018    2      0.0 !MK - CTS value unchanged
CTS  OES  CTS  CPS      0.3763    3      0.0 !MK - CTS value unchanged
CBS  OES  CTS  CTS     -0.8477    1      0.0
CBS  OES  CTS  CTS     -0.3018    2      0.0
CBS  OES  CTS  CTS      0.3763    3      0.0
CBS  OES  CTS  CPS     -0.8477    1      0.0 !MK - CTS value unchanged
CBS  OES  CTS  CPS     -0.3018    2      0.0 !MK - CTS value unchanged
CBS  OES  CTS  CPS      0.3763    3      0.0 !MK - CTS value unchanged
CBS  OES  CPS  CTS     -0.8477    1      0.0 !MK - CTS value unchanged
```

University of Cape Town

```
CBS  OES  CPS  CTS     -0.3018    2      0.0 !MK - CTS value unchanged
CBS  OES  CPS  CTS      0.3763    3      0.0 !MK - CTS value unchanged
CPS  OES  CTS  CTS     -0.8477    1      0.0 !MK - CTS value unchanged
CPS  OES  CTS  CTS     -0.3018    2      0.0 !MK - CTS value unchanged
CPS  OES  CTS  CTS      0.3763    3      0.0 !MK - CTS value unchanged
CTS  OES  CBS  CTS     -0.8477    1      0.0
CTS  OES  CBS  CTS     -0.3018    2      0.0
CTS  OES  CBS  CTS      0.3763    3      0.0
CTS  OES  CPS  CTS     -0.8477    1      0.0 !MK - CTS value unchanged
CTS  OES  CPS  CTS     -0.3018    2      0.0 !MK - CTS value unchanged
CTS  OES  CPS  CTS      0.3763    3      0.0 !MK - CTS value unchanged
CPS  OES  CBS  CTS     -0.8477    1      0.0 !MK - CTS value unchanged
CPS  OES  CBS  CTS     -0.3018    2      0.0 !MK - CTS value unchanged
CPS  OES  CBS  CTS      0.3763    3      0.0 !MK - CTS value unchanged
HAS  CTS  CTS  CTS      0.0000    1      0.0
HAS  CTS  CTS  CTS      0.0000    2      0.0
HAS  CTS  CTS  CTS      0.1441    3      0.0
HAS  CBS  CTS  CTS      0.0000    1      0.0
HAS  CBS  CTS  CTS      0.0000    2      0.0
HAS  CBS  CTS  CTS      0.1441    3      0.0
HAS  CPS  CTS  CTS      0.0000    1      0.0 !MK - CTS value unchanged
HAS  CPS  CTS  CTS      0.0000    2      0.0 !MK - CTS value unchanged
HAS  CPS  CTS  CTS      0.1441    3      0.0 !MK - CTS value unchanged
HAS  CTS  CTS  CBS      0.0000    1      0.0
HAS  CTS  CTS  CBS      0.0000    2      0.0
HAS  CTS  CTS  CBS      0.1441    3      0.0
HAS  CTS  CTS  CPS      0.0000    1      0.0 !MK - CTS value unchanged
HAS  CTS  CTS  CPS      0.0000    2      0.0 !MK - CTS value unchanged
HAS  CTS  CTS  CPS      0.1441    3      0.0 !MK - CTS value unchanged
HAS  CTS  CTS  HAS      0.0000    1      0.0
HAS  CTS  CTS  HAS      0.0000    2      0.0
HAS  CTS  CTS  HAS      0.1595    3      0.0
HAS  CBS  CTS  HAS      0.0000    1      0.0
HAS  CBS  CTS  HAS      0.0000    2      0.0
HAS  CBS  CTS  HAS      0.1595    3      0.0
HAS  CPS  CTS  HAS      0.0000    1      0.0 !MK - CTS value unchanged
HAS  CPS  CTS  HAS      0.0000    2      0.0 !MK - CTS value unchanged
HAS  CPS  CTS  HAS      0.1595    3      0.0 !MK - CTS value unchanged
```

University of Cape Town

```
HAS  CTS  OES  CTS      0.0000    1      0.0
HAS  CTS  OES  CTS      0.0000    2      0.0
HAS  CTS  OES  CTS      0.2840    3      0.0
HAS  CBS  OES  CTS      0.0000    1      0.0
HAS  CBS  OES  CTS      0.0000    2      0.0
HAS  CBS  OES  CTS      0.2840    3      0.0
HAS  CBS  OES  CPS      0.0000    1      0.0 !MK - CTS value unchanged
HAS  CBS  OES  CPS      0.0000    2      0.0 !MK - CTS value unchanged
HAS  CBS  OES  CPS      0.2840    3      0.0 !MK - CTS value unchanged
HAS  CPS  OES  CTS      0.0000    1      0.0 !MK - CTS value unchanged
HAS  CPS  OES  CTS      0.0000    2      0.0 !MK - CTS value unchanged
HAS  CPS  OES  CTS      0.2840    3      0.0 !MK - CTS value unchanged
HAS  CTS  OES  CBS      0.0000    1      0.0
HAS  CTS  OES  CBS      0.0000    2      0.0
HAS  CTS  OES  CBS      0.2840    3      0.0
HAS  CTS  OES  CPS      0.0000    1      0.0 !MK - CTS value unchanged
HAS  CTS  OES  CPS      0.0000    2      0.0 !MK - CTS value unchanged
HAS  CTS  OES  CPS      0.2840    3      0.0 !MK - CTS value unchanged
HAS  CPS  OES  CBS      0.0000    1      0.0 !MK - CTS value unchanged
HAS  CPS  OES  CBS      0.0000    2      0.0 !MK - CTS value unchanged
HAS  CPS  OES  CBS      0.2840    3      0.0 !MK - CTS value unchanged
HOS  OHS  CTS  CTS      1.0504    1      0.0
HOS  OHS  CTS  CTS      0.1336    2      0.0
HOS  OHS  CTS  CTS      0.3274    3      0.0
HOS  OHS  CBS  CTS      1.0504    1      0.0
HOS  OHS  CBS  CTS      0.1336    2      0.0
HOS  OHS  CBS  CTS      0.3274    3      0.0
HOS  OHS  CPS  CTS      1.0504    1      0.0 !MK - CTS value unchanged
HOS  OHS  CPS  CTS      0.1336    2      0.0 !MK - CTS value unchanged
HOS  OHS  CPS  CTS      0.3274    3      0.0 !MK - CTS value unchanged
HOS  OHS  CTS  CBS      1.0504    1      0.0
HOS  OHS  CTS  CBS      0.1336    2      0.0
HOS  OHS  CTS  CBS      0.3274    3      0.0
HOS  OHS  CTS  HAS      0.0000    1      0.0 !MK - CTS * 40% (hyd rot)
HOS  OHS  CTS  HAS      0.0000    2      0.0 !MK - CTS * 40% (hyd rot)
HOS  OHS  CTS  HAS      0.0677    3      0.0 !MK - CTS * 40% (hyd rot)
HOS  OHS  CBS  HAS      0.0000    1      0.0 !MK - CTS * 40% (hyd rot)
HOS  OHS  CBS  HAS      0.0000    2      0.0 !MK - CTS * 40% (hyd rot)
```

University of Cape Town

```
HOS  OHS  CBS  HAS      0.0677    3      0.0 !MK - CTS * 40% (hyd rot)
HOS  OHS  CPS  HAS      0.0000    1      0.0 !MK - CTS * 40% (hyd rot)
HOS  OHS  CPS  HAS      0.0000    2      0.0 !MK - CTS * 40% (hyd rot)
HOS  OHS  CPS  HAS      0.0677    3      0.0 !MK - CTS * 40% (hyd rot)
OES  CTS  CTS  HAS      0.0000    1      0.0
OES  CTS  CTS  HAS      0.0000    2      0.0
OES  CTS  CTS  HAS      0.1686    3      0.0
OES  CBS  CTS  HAS      0.0000    1      0.0
OES  CBS  CTS  HAS      0.0000    2      0.0
OES  CBS  CTS  HAS      0.1686    3      0.0
OES  CPS  CTS  HAS      0.0000    1      0.0 !MK - CTS value unchanged
OES  CPS  CTS  HAS      0.0000    2      0.0 !MK - CTS value unchanged
OES  CPS  CTS  HAS      0.1686    3      0.0 !MK - CTS value unchanged
OES  CTS  CBS  HAS      0.0000    1      0.0
OES  CTS  CBS  HAS      0.0000    2      0.0
OES  CTS  CBS  HAS      0.1686    3      0.0
OES  CTS  CPS  HAS      0.0000    1      0.0 !MK - CTS value unchanged
OES  CTS  CPS  HAS      0.0000    2      0.0 !MK - CTS value unchanged
OES  CTS  CPS  HAS      0.2086    3      0.0 !MK - changed from 0.1686
OES  CTS  CTS  OES     -2.6785    1      0.0
OES  CTS  CTS  OES      0.7851    2      0.0
OES  CTS  CTS  OES      0.2552    3      0.0
OES  CBS  CTS  OES     -2.6785    1      0.0
OES  CBS  CTS  OES      0.7851    2      0.0
OES  CBS  CTS  OES      0.2552    3      0.0
OES  CPS  CTS  OES     -2.6785    1      0.0 !MK - CTS value unchanged
OES  CPS  CTS  OES      0.7851    2      0.0 !MK - CTS value unchanged
OES  CPS  CTS  OES      0.2552    3      0.0 !MK - CTS value unchanged
OES  CTS  CTS  OHS     -3.7993    1      0.0
OES  CTS  CTS  OHS      0.5688    2      0.0
OES  CTS  CTS  OHS      0.4204    3      0.0
OES  CBS  CTS  OHS     -3.7993    1      0.0
OES  CBS  CTS  OHS      0.5688    2      0.0
OES  CBS  CTS  OHS      0.4204    3      0.0
OES  CTS  CBS  OHS     -3.7993    1      0.0
OES  CTS  CBS  OHS      0.5688    2      0.0
OES  CTS  CBS  OHS      0.4204    3      0.0
OES  CTS  CPS  OHS     -4.0193    1      0.0 !MK  changed from -3.7993
```

University of Cape Town

```
OES  CTS  CPS  OHS     -1.2688    2      0.0 !MK  changed from +0.5686
OES  CTS  CPS  OHS     -0.9704    3      0.0 !MK  CHANGED from +0.4202
OES  CTS  OES  CTS      0.1948    1      0.0
OES  CTS  OES  CTS      0.9778    2      0.0
OES  CTS  OES  CTS      0.9115    3      0.0
OES  CTS  OES  CPS      0.1948    1      0.0 !MK - CTS value unchanged
OES  CTS  OES  CPS      0.9778    2      0.0 !MK - CTS value unchanged
OES  CTS  OES  CPS      0.9115    3      0.0 !MK - CTS value unchanged
OES  CBS  OES  CTS      0.1948    1      0.0
OES  CBS  OES  CTS      0.9778    2      0.0
OES  CBS  OES  CTS      0.9115    3      0.0
OES  CBS  OES  CPS      0.1948    1      0.0 !MK - CTS value unchanged
OES  CBS  OES  CPS      0.9778    2      0.0 !MK - CTS value unchanged
OES  CBS  OES  CPS      0.9115    3      0.0 !MK - CTS value unchanged
OHS  CTS  OES  CTS      1.9193    1      0.0
OHS  CTS  OES  CTS      1.0102    2      0.0
OHS  CTS  OES  CTS      0.7294    3      0.0
OHS  CBS  OES  CTS      1.9193    1      0.0
OHS  CBS  OES  CTS      1.0102    2      0.0
OHS  CBS  OES  CTS      0.7294    3      0.0
OES  CTS  OHS  HOS      1.2936    1      0.0
OES  CTS  OHS  HOS      1.3295    2      0.0
OES  CTS  OHS  HOS      0.4323    3      0.0
OES  CBS  OHS  HOS      1.2936    1      0.0
OES  CBS  OHS  HOS      1.3295    2      0.0
OES  CBS  OHS  HOS      0.4323    3      0.0
OHS  CTS  CTS  HAS      0.0000    1      0.0
OHS  CTS  CTS  HAS      0.0000    2      0.0
OHS  CTS  CTS  HAS      0.1472    3      0.0
OHS  CBS  CTS  HAS      0.0000    1      0.0
OHS  CBS  CTS  HAS      0.0000    2      0.0
OHS  CBS  CTS  HAS      0.1472    3      0.0
OHS  CPS  CTS  HAS      0.0000    1      0.0 !MK - CTS value unchanged
OHS  CPS  CTS  HAS      0.0000    2      0.0 !MK - CTS value unchanged
OHS  CPS  CTS  HAS      0.1472    3      0.0 !MK - CTS value unchanged
OHS  CTS  CBS  HAS      0.0000    1      0.0
OHS  CTS  CBS  HAS      0.0000    2      0.0
OHS  CTS  CBS  HAS      0.1472    3      0.0
```

University of Cape Town

```
OHS  CTS  CTS  OHS     -4.9362    1      0.0
OHS  CTS  CTS  OHS      0.2907    2      0.0
OHS  CTS  CTS  OHS      0.4638    3      0.0
OHS  CBS  CTS  OHS     -4.9362    1      0.0
OHS  CBS  CTS  OHS      0.2907    2      0.0
OHS  CBS  CTS  OHS      0.4638    3      0.0

IMPROPER

NONBONDED  NBXMOD 5  atom CDIEL shift vatom VDISTANCE VSWITCH -
  CUTNB 13.0  CTOFNB 12.0  CTONNB 10.0  EPS 1.0  E14FAC 1.0  WMIN 1.5
!
!                 Emin       Rmin
!                 (kcal/mol) (A)
HSPC        0.0000    -0.0000    1.3750    ! SPC water hydrogen
HT          0.00      -0.046     0.2245    ! TIP3P
OSPC        0.0000    -0.1554    1.7766    ! SPC
OT          0.00      -0.1521    1.7682    ! TIP3P
HOS         0.00      -0.0460    0.2245    !
HAS         0.00      -0.0220    1.3200    !
CTS         0.00      -0.0200    2.2750    0.000   -0.01000   1.90000
CBS         0.00      -0.0200    2.2750    0.000   -0.01000   1.90000
CPS         0.00      -0.0200    2.2750    0.000   -0.01000   1.90000
OHS         0.00      -0.1521    1.7700       ! from para_na, on5
OES         0.00      -0.1521    1.7700       ! from para_na, on6

!
NBFIX
!
!       NBFIX the TIP3P water-water interactions
OT   OT      -0.152073  3.5365  ! TIPS3P VDW INTERACTION
HT   HT      -0.04598   0.4490
HT   OT      -0.08363   1.9927
!
!   the following will fix the SPC-sugar interaction terms
!
OSPC OSPC    -0.15539   3.5532  ! SPC VDW INTERACTION
HSPC HSPC    -0.00000   0.4490
```


University of Cape Town

```
HSPC OSPC     -0.00000    1.9927
OSPC OHS      -0.15888    3.4657
OSPC OES      -0.15888    3.4657
OSPC CTS      -0.13067    3.4733
OSPC CBS      -0.13067    3.4733
OSPC CPS      -0.13067    3.4733
OSPC HAS      -0.11015    2.7973
OSPC HOS      -0.00000    1.8
!
!
! this force field treats hbonds implicitly using charges; therefore
! the following section wildcards all the energies to 0. Note that you
! should also set the IHBFRQ to zero in all calcs, to avoid wasting
! time updating the hydrogen bond list!
!
HBOND AEXP 4 REXP 6 HAEX 4 AAEX 0 NOACCEPTORS HBNOEXCLUSIONS ALL-
  CUTHB 0.5 CTOFHB 5.0 CTONHB 4.0 CUTHA 5.0 CTOFHA 90.0 CTONHA  90.0
!
H*    O*       -0.00       2.0
END
```

University of Cape Town