Towards Designing *De Novo* (1→4)-Linked Polysaccharides

by

Jeff Yu-Jen Chen

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

07/Aug/2004

Abstract

Carbohydrates are invaluable in a variety of industrial applications ranging from biodegradable materials to pharmaceutical drug carriers. Small functional groups, such as hydroxyl, amine, *N*-acetylamide and methoxy are commonly employed for creating *de novo* saccharides with enhanced functionality. The physiochemical properties of carbohydrates are governed by their conformations and the polarity of their substituents strongly influences the overall flexibility of such biomolecules. In this thesis, two types of (1→4)-inter-glycosidic interactions, *viz.* hydrogen bonding and hydrophobic effects, are investigated.

A disaccharide is the smallest unit which incorporates inter-residue glycosidic linkages and it is such linkage that allows the larger saccharides to be formed. The strength of the hydrogen bond across the glycosidic linkage can be tampered by replacing the hydroxyl groups on the C2 and C3' position with a variety of different polar substituents. It was previously not possible to evaluate inter-molecular hydrogen bond strength in a quantitative manner. However, a novel method is used in this thesis where ΔE^{HB} vs $\Delta_{\rho(r)+Lap}$ correlation curves were used to estimate the intramolecular hydrogen bond strength which occurs between functional groups located across glycosidic linkages. The hydrogen donor (***D***) and acceptor (***A***) pair have been ranked from the strongest (*viz.* NHCOCH3(D)·····NH2(A)) to the weakest (viz. NH2(D)·····NH2(A)). However, subsequently a series of molecular mechanics based conformational analysis and solution dynamics of the disaccharides namely maltose, 2-amine-Glc-α-(1-4)-Glc, GlcNAc-α-(1-4)-3-amine-Glc, reveals that an increase in hydrogen bond strength across the glycosidic linkage does not necessarily lead to a decrease in the flexibility above the molecules.

Researchers have traditionally believed that network of inter-residue hydrogen bonds is the factor leading to the anomalous aqueous solubility of β-cyclodextrin (CD). In this thesis, 5ns molecular dynamics (MD) computational simulations of α-, β-, and γ-CD solutions have been performed to gain further insight into this anuomalos behaviour Moreover, the local water structuring around and conformational dynamics of the CDs have been investigated. In addition, translational diffusion measurements have been carried out for the three CDs using a PGSE NMR techniques and the corresponding diffusion coefficients obtained from the MD trajectories have found to be in good agreement with those observed experimentally. The MD studies further indicate that in the presence of water solvent, the macrocycle distortion and glucopyranose tilting are the preferred types of motion for relieving the steric strain of CD rings. The investigations clearly demonstrate that the β-CD is relatively more rigid than other two CDs thus initiated an increase local water structuring around the solute. In addition, the internal motion of the trimethylated β-CD, which does not form inter-glycosidic hydrogen bonds but possess a higher aqueous solubility than β-CD, has been examined. The molecular motion of three other β-CD derivatives, DIMBE, 2D2HB and 2D2MB, have also been studied. A key finding of this thesis is that the conformational flexibility of the modified cyclodextrins is greatly influenced by the bulkiness and rigidity of the hydrophobic substituents, whereas the strength of inter-residue hydrogen bonding does not appear to be an important factor.

Acknowledgements

I am privileged to have the opportunity of joining Associate-Professor Kevin J. Naidoo's research group. I am particularly thankful for his insightful guidance and valuable support towards this work. Furthermore, I am also grateful for the enthusiasm he has shown my future career development.

For their invaluable support and immeasurable encouragement, I thank my parents Nancy and Eddie, my brother George, and my girlfriend Evelyn as well as for her help in the completion of my thesis.

Throughout the course of my Ph.D, the following people have all contributed towards the progress of my studies. I would like to show my sincere appreciation to Dr. Michelle Kuttel, Dr. Anton Lopis, Dr. Robert Best and Mr. Stephan Simpson for their input and expertise.

I am very grateful to Jennie Jansson, Professor Göran Widmalm and Professor Arnold Maliniak for providing the NMR data and their insightful advice with the cyclodextrin section of this thesis.

I would also like to extend my gratitude to the University of Cape Town and the National Research Foundation of South Africa for providing the well-needed financial support.

Publications and Presentations

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

- "Molecular properties related to the anomalous solubility of β-cyclodextrin"

  J. Phys. Chem. B, 2004, 108(14), 4236-4238

- "Evaluating intramolecular hydrogen bond strengths in (1→4) linked disaccharides from electron density relationships"

  J. Phys. Chem. B, 2003, 107, 9558-9566

- "The role of water in the design of glycosidic linkage flexibility"

  Mol. Phys. 2003, 201, 17, 2687-2694

- "An *ab initio* study of inter-residue hydrogen bonding in α(1-4) linked polysaccharides"

  2001, Abst. of the 222nd Am. Chem. Soc. meeting, Chicago, IL, USA

and presented at the following conferences and institutions:

- "Evaluating intra-molecular hydrogen bond strengths in biopolymers from electron density relationships"

  Oral presentation, 37th National Convention of the South African Chemistry Institute, CSIR international Convention Centre, Pretoria, R. S. A. 2004

- "Why is β-cyclodextrin less soluble than α- and γ-cyclodextrin? "

  Oral presentation, 37th National Convention of the South African Chemistry Institute, CSIR international Convention Centre, Pretoria, R. S. A. 2004

- "Molecular properties related to the anomalous solubility of β-cyclodextrin"

  Oral presentation, International Symposium of Carbohydrates, University of Cape Town, Cape Town, R. S.A, 2004

- "Towards designing *de novo* (1-4) linked polysaccharides"

  Oral presentation, University of Debrecen, Debrecen, Hungary, 2003

- "An *ab initio* study of inter-residue hydrogen bonding in α(1→4) linked polysaccharides"

  Poster presentation, CPP5 2001 Summer School, UMIST, Manchester, UK, 2001

- "An *ab initio* study of inter-residue hydrogen bonding in α(1→4) linked polysaccharides"

  Oral presentation, 4th Materials Modelling Meeting, University of the North, SA, 2001

Lists of Abbreviations

2D2HB	heptakis(2-deoxy-2-hydro)-β-cyclodextrin
2D2MB	heptakis(2-deoxy-2-methyl)-β-cyclodextrin
Å	Angstroms
ΔE^{HB}	hydrogen bonding energy
Δ_{el+lap}	complex parameter
ρ	electron density
τ	glucopyranose monomer tilting angle
ω	macro-cyclic ring pseudo-dihedral angle
A	hydrogen bonding "acceptor" atom
AIM	atoms in molecules theory
B3LYP	Becke-Three Lee-Yang-Parr
BSSE	basis set superposition errors
CD	cyclodextrin
CG	conjugated gradient minimisation method
$-CH_3CONCH_3$	-*N*-acetylamide group
CHARMM	chemistry at Harvard molecular mechanics
CM	centres of mass
D	hydrogen bonding "donor" atom
DIMEB	heptakis(2,6-di-*O*-methyl)-β-cyclodextrin
DFT	density functional theory
E	energy
EFF	empirical force field
ESA	electronic structural approach
fs	fematoseconds
HB	hydrogen bond
HF	Hartee-Fock
K	Kelvin
kcal	kilocalories
MD	molecular dynamics
Me-	methyl group

MP2	Møller-Plesset perturbation theory
NAc	-*N*-acetylamide group
$-NH_2$	amine group
NPT	isothermal-isobaric ensemble
ns	nanosecond
NVE	microcanonical ensemble
$-OCH_3$	–*O*-methyl group
-OH	hydroxyl group
PBC	periodic boundary conditions
PDF	pair distribution function
PGSE NMR	pulse-field-gradient spin-echo nuclear magnetic resonance
ps	picosecond
RMS	root mean square
SCF	self-consistent field
SD	steepest descent minimisation method
SDF	spatial distribution function
$SD_{(Fluc)}$	fluctuation of standard deviation
TCF	time correlation function
TRIMEB	heptakis(2,3,6-tri-*O*-methyl)-β-cyclodextrin
VP	Voronoi Polyhedra

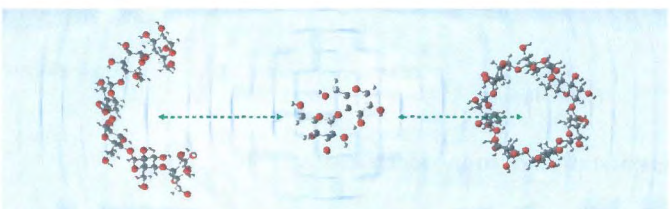
Contents

List of Figures

List of Tables

Chapter One

Simulating Saccharides in the Condensed Phase



1.1 Importance of Carbohydrates

Carbohydrates are by far the most abundant class of the biological molecules which comprise a wide diversity of biopolymers. These molecules are also known as saccharides which originate from Latin *saccharum,* meaning sugar. Normally, saccharides have the general composition of $C_xH_{2y}O_y$ and are simply constituted of carbon, hydrogen and oxygen. These molecules play a variety of significant roles in all living organisms ranging from energy storage to structural support. Saccharides are also found in the DNA which carries genetic information, while some of these biopolymers are used as pharmaceutical products.[1,2]

The classification of saccharides is based on their capabilities for hydrolysis. One can roughly divide these biomolecules into two major categories, simple and complex carbohydrates. The simple carbohydrates are monosaccharides and can not be hydrolysed into smaller units. The complex carbohydrates comprise two or more monosaccharide units which are linked together via glycosidic linkages and can be hydrolysed to yield its simple monosaccharide constituents.[1,2]

1.1.1 Simple Carbohydrates

Fructose, lactose and glucose are the three most common monosaccharides. Monosaccharides can be classified into either aldoses or ketoses and are named according to the number of the carbon atoms found within the sugar as well as the position of the hydroxyl group. Furthermore, these Simple carbohydrates can be categorised according to their stereocenters. Almost all of the natural monosaccharides have the same right turning optical stereochemical configuration at the chiral carbon farthest from the carbonyl group, these are termed D-sugars (D representing dextrorotatory). The left turning sugars, on the other hand, are known as the L-sugars (L representing levorotatory).[1,3] Figure 1.1 illustrate some of the monosaccharides and unless specified otherwise, as of this point, each small ball represents an atom constituent of the molecule and the tube joining the balls together represents the covalent bond. Colour grey, red and white represent carbon, oxygen and hydrogen respectively).

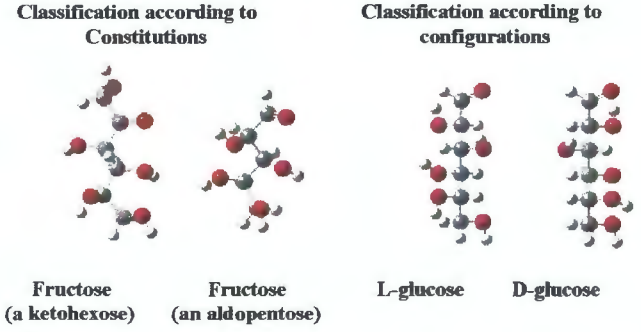
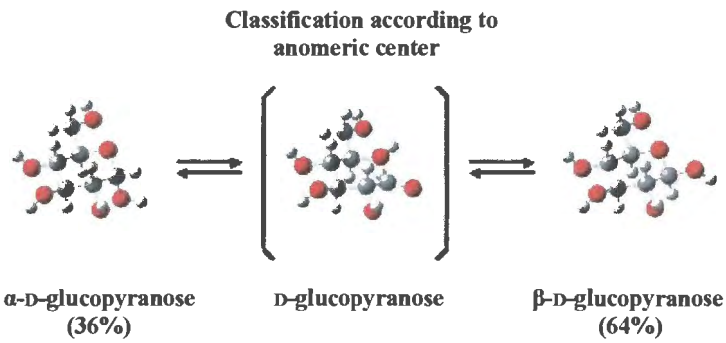


Figure 1 .1 Classification of monosaccharides

Monosaccharides can either exist as open straight chains or closed ring shaped molecules. The open chain monosaccharides forms are flexible, whereas those that occur as rings generally adopt the more stable chair conformation. In nature, only a very small fraction of the monosaccharides exist in the open-chain form which has a free aldehyde- or keto-group attached. The reason is that the carbonyl and hydroxyl groups in the open-chain sugar spontaneously undergo hemiacetal formation in aqueous condensed medium to yield a ring structure. For instance, glucose readily undertakes hemicacetal formation in to form glucopyranose in the presence of water molecules. Unless under the presence of external force compression, the pyaronse ring exist virtually only as the $^4C_1$ chair conformations in nature.[4] This hemiacetal formation in glucose can lead to two stereoisomers involving the anomeric carbon which are termed as α- or β-sugars. Even though the ring constantly undergoes mutarotation (i.e. a reversible ring opening of each anomer to the open-chain aldehyde followed by re-closure) the β-D-glucopyranose is favoured almost twice as much as than its α-anomer in aqueous solution.[2]



-Figure 1.2 Formation of two anomeric sugars (α- and β-forms).

1.1.2 Disaccharides and Complex Carbohydrates

Complex carbohydrates are difficult to study as they are large enough in size to behave in a complex manner yet they are still relatively small in size so that gross structural generalization in their conformational changes would not provide satisfactory information on its solvation nature. Hence they cannot be studied using simplified models and detailed investigations at the atomic level are required. A vast variety of complex carbohydrate can be found in nature, ranging from straight or branched long chain carbohydrates to cyclised polymers. The smallest complex carbohydrates are the disaccharides (refer to next page for more discussion). Many disaccharides are biologically important and they may join together to form more sophisticated saccharides. Larger saccharides composed of only a few monosaccharides residues are refer to as the oligosaccharides (ranging from 6 to 10) and polysaccharides.[1,2]

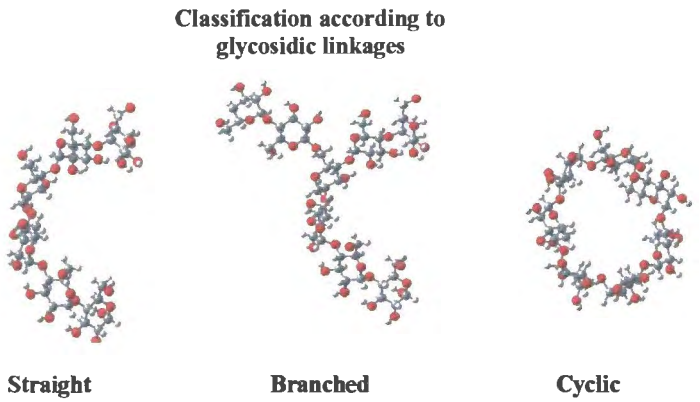


Figure 1.3 Classification of the complex carbohydrates.

Disaccharides: Most of the natural occurring disaccharides are formed by the joining of two monosaccharides via glycosidic acetal bonds between the anomeric carbon of one sugar and a hydroxyl group at any position on the other sugar. For example, a bonding can emerge from an α positioned anomeric C1 hydroxyl group of one sugar residue to a C4 hydroxyl group of the other. This bond formation is designated as α(1→4). Similarly, α(1→6) indicates a bond formation from an α positioned anomeric C1 hydroxyl group of one sugar residue to a primary hydroxyl group of the other residue. Disaccharides are classified according to the compositions of the monomer units. Disaccharides are an important class of carbohydrate for two main reasons. Firstly they have vast chemical and biological importance[1] and secondly they have a transitional position between mono- and polysaccharides since they are the smallest unit possessing of an inter-residue glycosidic linkage which are the central feature of the larger oligo- and polysaccharides.[2]

It has been shown that the glycosidic oxygen atoms play a considerable part in influencing the overall carbohydrate structure. The conformation of the saccharides in relation to the glycosidic dihedral pair (ϕ = H1-C1-O1-C4' and ψ = C1-O1-C4'-H4'; this is in accordance with those used previously by Brady *et al.*[5]) have been studied extensively by both the experimental[6-9] as well as the theoretical methods[5,10]. As a result of this, the overall topology of a large polysaccharide is influenced significantly by the nature of each the individual glycosidic dihedral angle found within the polymer. By rotating the above two glycosidic dihedrals systematically, one can generate its energy hypersurface known as an adiabatic map (refer to Chapter 5 for further details). In this case the strength of the cross-glycosidic intramolecular hydrogen bond has a significant effect on the degree of rotational flexibility of the glycosidic dihedral angles. Due to the pervasiveness of hydrogen bonding in carbohydrates, this can be critically influenced by the dihedral and the atomic negativity of the substituted functional groups methods [5,10]

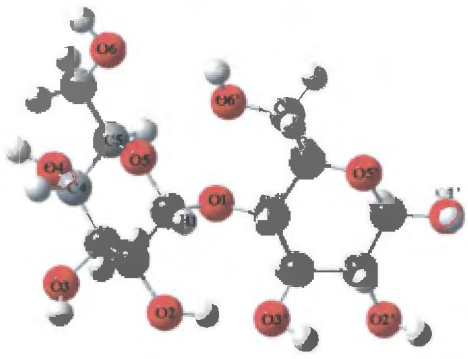


Figure 1.4 The atom naming convention of a disaccharide.

Linear and Branched Saccharides: Starch is one of the most abundant polysaccharides found in nature. It occurs naturally as discrete granules. This polymer is relatively dense, insoluble and hydrate only slightly within the aqueous medium. Being the end product of photosynthesis, starch is the major carbohydrate reserve in plant tubers and seed endosperm. Starch is composed of a mixture of two types polymers, amylose and amylopectin. Amylose is the constituent of starch in which each anhydroglucose units are linked by α-D-(1→4) glucosidic bonds to form essentially linear chains. Whereas amylopectin is a polysaccharide having a polymeric but branched structure, in addition to α-D-1,4 bonds, an α-D-(1→6) bonds occurs in every 20-30 anhydroglucose units.[1,2]

Cyclised Saccharides: Cyclodextrins (CD) are the most common cyclic oligosaccharides. The first report concerning the isolation of cyclodextrin can be dated back to 1891 when Villiers isolated a small amount of a crystalline substance from a starch cultured medium used for growing the Bacillus amylobactor. Between 1903 and 1911, Schardinger successfully characterised this crystalline substance.[11] Cyclodextrin generally contain a minimum of six α-(1→4) linked D-(+)-glycopyronose units but other analogues comprising a larger number of monomer constituents have also been made.[12] (Refer to session 1.3 and chapters 7 and 8 for more in-depth discussion)

1.1.3 Effects influencing the saccharide conformational behaviour

For comprehensive investigation of the biophysical functions of a particular molecule, it is essential to understand its structural changes and its properties in its natural environment and how the environment affects the molecule itself. This is especially important in the study of carbohydrates considering their conformations are highly sensitive to hydration. Hydrogen bonding and hydrophobic effects both play a crucial role on the structural transformation of the carbohydrate.[13]

Hydrogen bonding is an interaction between molecules in which a hydrogen atom that is bound to an electronegative atom interacts with an electron-rich atom on another molecule. Despite much research and advances in knowledge, hydrogen bonding has remained as a controversial topic over the years. Hydrogen bond has known to comprise both covalent and ionic characteristics. The true nature of the hydrogen bond has been extensively analysed and discussed in the literature.[14,15] From a modelling standpoint, the pervasiveness of the hydrogen bonding sites found in carbohydrates makes the study of carbohydrates in the condensed phase challenging. This is because of the free rotating nature of the hydroxyl groups that can simultaneously act as the hydrogen bond donor and acceptor depending on their interaction orientation. The variations of the exo-cyclic functional groups can subsequently alter the overall physiochemical properties of the carbohydrate and are therefore the central feature of the macromolecular conformation influences.[2,13] (Refer to chapter 4 for more in-depth description)

The hydrophobic effect is another factor that makes a substantial contribution to the thermodynamic stability of a saccharide. The hydrophobic effect arises when solvating substances of less polarity in an polar aqueous medium. This causes the water molecules neighbouring the solute to arrange themselves in a more orderly structure and thus produce ice-like orientations. This orederly arrangement of solvent molecules corresponds to a total loss of configurational variation in the system resulting in a decrease of entropy. It is widely believed that hydrophobic interactions provide the driving force for various important physicochemical processes such as protein folding and biomolecular recognition processes.[16] In conjunction with the hydrophilic effects, the hydrophobic effect also play a determining role the intrinsic aqueous solubility the carbohydrate. (This part will be explored further in chapter 8)

1.2 Computational Chemistry Principles Employed in This Project

It is in general challenging to evaluate properties of the biopolymer within the condensed phase medium using the conventional experimental techniques. Certain conformational changes or reaction pathways are generally below or beyond the experimental time scale and are therefore out of the instrumental limitations. Other times, biopolymers may hold complicated structural shapes or unusual properties making it almost impossible to draw conclusions from the experimental data. Computer modelling has recently emerged as a the discipline of choice in the study of these intricate biopolymer due to its capability of providing detailed descriptions.[17] The thorough use of computational modelling techniques can promote ones understanding of the nature of biopolymers and consequently aid the design of *de novo* methods or materials of interest.

There are numerous computation chemistry techniques known to date. Currently, in this work, two different fields within computational chemistry are employed for modelling the structural properties

of saccharides: the electronic structural approach (EAS) and empirical force field (EFF) methods. [17,18,19] The fundamental concepts behind these two methods are discussed in chapter 2 and 3.

At this point, it should be emphasized that all of these simulation techniques are not a replacement for conventional experimental chemistry. Instead, the computational work should be used to complement the experimental data. With any simulation, the computation techniques must be validated by comparing the results generated with those of conventional experiments.

1.3 Project Motivations

As was previously mentioned, the structural characteristics of the carbohydrate influence its physicochemical properties. In general, normal carbohydrates, are only able to adapt a limited number of conformational changes in the aqueous medium due to the steric hindrances imposed on the rotation of sugar units about the inter-glycosidic bonds.[13] The allowed conformations are dependent on the type of carbohydrate linkages as well the intra- and inter-molecular forces. The most stable conformations of the polysaccharide will be that which balanced both the intra- and inter-molecular forces. In this project, we aim to explore the effect of these two intra-molecule forces (hydrogen bonding and hydrophobic effect) on the carbohydrate.[2,13] Due to the following commercial reasons, we focus our research particularly on the conformational flexibility of some α(1→4)-linked carbohydrates:

(1) Starch is widely available in almost unlimited quantities from agricultural products. As was previously mentioned, this biopolymer comprises α(1→4)-linked α-D-glucopyranose polysaccharide with α(1→6)-branches. Starch can be genetically modified to improve its functionality and commercial uses.[20] Recently owing to environmental pressures, starch based polymeric materials have gained of considerable interest as the preferred replacement to the traditional non-biodegradable plastic products. These starch-based polymeric materials may range from packing materials to eating utensils. Studies have shown that extrusion of starch is made possible when crystallinity is removed under the condition of a relatively high temperature and low water content. By cross-linking or co-melting this substance with other polymers, for instance polyethylene and poly(ethylene-co-acrylic acid), one can produce materials suitable for applications such as packing materials or injection molded biodegradable plastics.[21,22] However, in the presence of water, these materials gain undesirable plasticity and this limits their practicality. Various important studies have shown that this as a result of the bridging water phenomenon which increase the total flexibility of the carbohydrates at the molecular level.[23,24] A possible method of reducing the flexibility is by altering the hydroxyl functional groups on saccharides to substituents that capable of initiating stronger inter-glycosidic interactions. By doing so, this may perhaps hinder or prevent the formation of the intermolecular hydrogen bonding between the saccharides and water. In order to make an intelligent guess as about which

modifications might prove fruitful in the *de novo* design of biodegradable materials, it is essential to understand the dynamic properties of starch in under a water environment. After examining the complex carbohydrates and their corresponding disaccharides, it is clear that the constituents sugar residues and glycosidic linkages are essentially comparable. Therefore one can expected the disaccharides to exhibit some similarities to the larger complex saccharides in their basic physical-chemical properties. We commence our studies by examining the conformational behaviour of some chemically modified α(1→4)-linked disaccharides (chapters 4, 5 and 6) in the hope that the results found may then be extrapolate to the larger systems.

(2) Cyclodextrins, as previously mentioned, are the cyclic α(1→4)-linked α-D-glucopyranose oligomers. Due to it structure, these cone-shape molecules have a hydrophilic exterior and hydrophobic cavity (Figure 1.5). It is this unique structure and properties of cyclodextrins which resulted in their extensive applications in the fields ranging from pharmaceutical industries to food and cosmetic research[25-27]. If a less hydrophilic molecule (commonly drugs with low molecular weight) can fit entirely or partially into the cyclodextrin cavity, an inclusion complex of the guest-host complexation may occur owing to the participation of various intermolecular interactions. These interactions include hydrophobic interactions, van der Waals forces, hydrogen bonding, release of high energy water molecule in complex formation and relief of stain energy in the macromolecular cyclodextrin ring. The complexation phenomenon generally increased the stability as well as the aqueous solubility of the drugs. Even though three major parent cyclodextrins, designated α, β, and γ, comprising from six, seven and eight glycopyranose units respectively, have been regularly studied, it is their chemically modified derivatives that have also been more frequently utilized in the pharmaceutical field. These modified cyclodextrins may show less toxicity and often possess a higher degree of intrinsic solubility and are hence able to complex with larger amount of guest molecules.[25] Often, these cyclodextrins are derivitised with hydrophobic functional groups and still show an improvement of the aqueous solubilities when comparing to its precursor. To our knowledge, all of these cyclodextrins are important to the pharmaceutical industry, research concerning these preposterous differences in the trend of their solubilities has somehow been neglected.[28,29,30] Theoretical as well as experimental studies exclusively targeting this puzzle are somewhat scarce and comprehensive theoretical explanations on these problems have not been fully resolved. An understanding of this paradox could lead toward the design of pristine chemically modified cyclodextrins which would be useful for a wider range of pharmaceutical application.

This research targets the effect of hydrophilic and hydrophobic substitutents on the physicochemical properties of above mentioned carbohydrates and subsequently enabling researcher to invent materials with improved properties. However, this is certainly not the ultimate objective of this research. In truth, these molecules are themselves of scientific significance; results obtaining from this

work are useful in its own right. The more sophisticated problems such as cell-mediated glycoproteins or carbohydrate assembly[1] can only be answered when the fundamental properties of carbohydrate solvation behaviour are fully understood. Therefore the thorough investigation concerning the intricate nature of the hydrogen bonding and hydrophobic effect must be carried out. The aim of this project is to contribute a fraction of the knowledge towards the foundation of solving such problems.

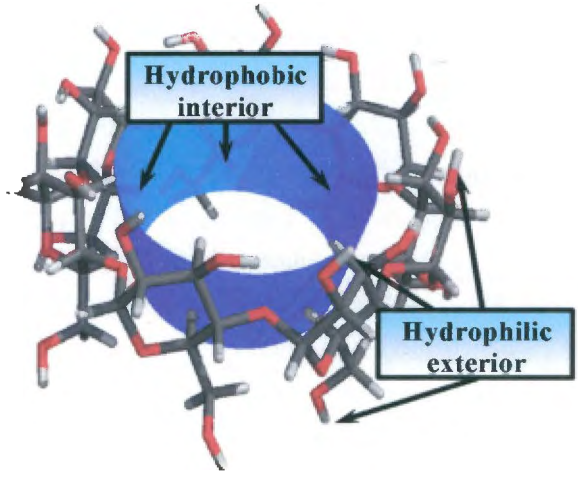


Figure 1.5 The unique property of a cyclodextrin.

1.4 Project Protocol

The pervasiveness of the hydrogen bonding is common to almost every type of saccharides. While evaluating inter-molecular hydrogen bonding strength can be carried out in a fairly straightforward procedure, estimating intra-molecular hydrogen bonding strength appropriately has been a challenging task in the past. Relying on the fundamental ESA principles (refer to chapter 2 for a brief discussion relating to some of the fundamental concepts), a novel method of estimating intra-molecular hydrogen bonding strengths was consequently developed by us for the purpose of rapid charactering any intra-molecular hydrogen bond strength initiated by any combination of substituents. In chapter 4, we demonstrate the usefulness of this method for predicting the intra-molecular hydrogen bonding strength of various chemically modified disaccharides mimics. Furthermore, we were able to predict the conformational determining role of inter-glycosidic hydrogen bonds within the aqueous medium based on the results obtained from this method.

Bioploymers are vibrant in nature when immersed within the aqueous condensed phase and the EFF dynamics simulations are well-suited to investigate their relativistic behaviour (refer to chapter 3 for further details). Developing a set of adequate force field parameters is a prerequisite prior to any EFF

dynamic simulations, a set of parameters must be developed. The outline of the procedures of developing some force field parameter is documented in chapter 5. Furthermore, chapter 5 also includes some fundamental conformational studies in vacuum of certain disaccharides which were specifically chosen based on the findings from chapter 4. These studies lead to the need of comparing the condensed phase molecular dynamics simulations of two α(1→4) linked disaccharides which is documented in chapter 6. By comparing the results obtained from chapter 4, 5 and 6, we were able to confidently deduce the conformational determining role of inter-glycosidic hydrogen bonds within the aqueous medium.

The native and the chemically modified cyclodextrins are commercially important since they are used for a wide range of industrial applications. Their anomalous aqueous solubilities have been a topic of controversy over the years. In chapters 7 and 8, the conformational behaviour of each individual cyclic-oligomers was simulated within the large solvent box. The dynamics properties of each cyclic-oligomer in solution is examined using various simple yet effective approaches which share light on the issue of problems relating to their solubilities. Particularly in chapter 8, we rely on these approaches to stuy the hydrophobic effect on the conformational flexibility of certain cyclodextrins that are derivertised with various hydrophobic functional groups. The results obtained from these two chapters enable us to rationalise the conformational determining role of hydrophobic groups.

1.5 Conclusion

Carbohydrates are conformationally complex molecules and this complexity is responsible for their interesting physicochemical properties. It is unusual features of carbohydrates which allow them to play crucial roles in a variety of fields. Knowledge of carbohydrate molecular structure, particularly their structures in solution, is the key of understanding their macroscopic properties. Detailed studies of carbohydrate solution conformational changes are proven to be difficult to investigate experimentally. Computational chemistry is a fast growing field of modern chemistry with simulation methodologies and with the computer efficiency being constantly improved, this field of discipline becomes the invaluable tool for assisting such studies.

REFERENCES

(1) Mattew, C. K.; van Holde, K. E. *Biochemistry*, The Benjamin & Cummings Pub. Co.: Redwood City, 1991.

(2) French, A. D. *Carbohydrate*, In *Encyclopedia of computational chemistry*, 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, 1998; pp 209.

(3) McMurry, J. *Fundamentals of organic chemistry*, 4th ed.; Brooks & Cole Pub. Co.: Oxford, 1998.

(4) O'Donoghue, P.; Luthey-Schulten, Z. A. *J. Phys. Chem. B* **2000**, *104*, 10398.

(5) Ha, S. N.; Madsen, L. J.; Brady, J. W. *Biopolymer* **1988**, *27*, 1927.

(6) Quigley, G. J.; Sarko, A.; Marchessault, R. H. *J. Am. Chem. Soc* **1970**, *92*, 5834.

(7) Shashkov, A. S.; Lipkind, G. M.; Kochetkov, N. K. *Carbohydr. Res.* **1986**, *147*, 175.

(8) Gress, M. E.; A., J. G. *Acta Cryst.* **1977**, *B34*, 213.

(9) Stevens, E. S.; Sathyanarayana, B. K. *J. Am. Chem. Soc* **1989**, *111*, 4149.

(10) Momany, F. A.; Willett, J. L. *J. Comp. Chem* **2000**, *21*, 1204.

(11) Schardinger, F.; Unders, Z. *Nahr. Genussm* **1903**, *6*, 109.

(12) Saenger, W.; Jacob, J.; Gessler, K.; Steiner, T.; Hoffmann, D.; Sanbe, H.; Koizumi, K.; Smith, S. M.; Takaha, T. *Chem. Rev.* **1998**, *98*, 1787.

(13) Brady, J. W. Solvation: Carbohydrates. In *Encyclopedia of computational chemistry*, 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, 1998; pp 2609.

(14) Del Bene, J. E. Hydrogen bonding: 1. In *Encyclopedia of computational chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, 1998; pp 1263.

(15) Lii, H. J. Hydrogen bonding:2. In *Encyclopedia of computational chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, 1998; pp 1271.

(16) Pratt, L. R. Hydrophobic effect. In *Encyclopedia of computational chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, 1998; pp 1286.

(17) Grant, G. H.; Graham Richards, W. *Computational Chemistry*, Oxford University Press: Oxford, 1995; Vol. 29.

(18) Leach, A. R. *Molecular modelling: Principles and applications*, 1st ed.; Addison Wesley Longman Limited: Singapore, 1996.

(19) Foresman, J. B.; Frisch, Æ. *Exploring Chemistry with electronic strucure methods*, 2nd ed.; Gaussian, Inc.: Pittsburgh, 1996.

(20) Jobling, S. *Curr. Opinion Plant Biol.* **2004**, 210.

(21) Donane, W. M. *Starch/Stäerke* **1992**, *44*, 293.

(22) Shogren, R. L.; Fanta, G.; Donane, W. M. *Starch/Stäerke* **1993**, *45*.

(23) Naidoo, K. J.; Kuttel, M. *J. Comp. Chem.* **2001**, *22*, 445.

(24) Hulleman, S. H. D.; Janssen, F. H. P.; Feil, H. *Polymer* **1998**, *39*, 2043.

(25) Szejtli, J. *Chem. Rev.* **1998**, *98*, 1743.

(26) Breslow, R.; Doug, S. D. *Chem. Rev.* **1998**, *98*, 1825

(27) Schurig, V.; Nowotny, H. *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 939.

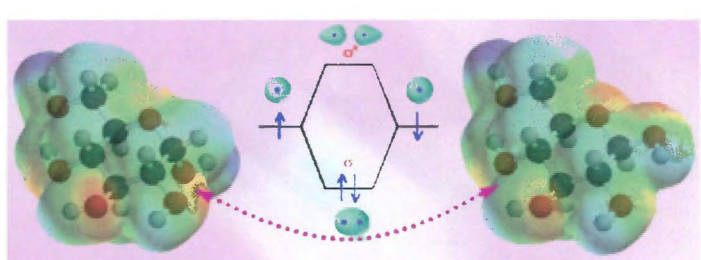
(28) Linert, W.; Margl, P.; Renz, F. *J. Chem. Phys.* **1992**, *161*, 327.

(29) Kozár, T.; Venanzi, C. A. *J. Molec. Struct. (Theochem)* **1997**, *395-396*, 451.

(30) Lipkowitz, K. B. *Chem. Rev.* **1998**, *98*, 1829.

Chapter Two

An Overview of the Electronic Structure Approach: Theories and Applications



2.1 Theoretical Background

By the beginning of the twentieth century, most scientists were content with the fundamental theories. Apart from a few minor difficulties, which were referred to in a lecture given by Lord Kelvin on April 1900, most researchers were satisfied that all problems could be accounted for by methods derived from the principles originally published by Sir Issac Newton in his famous book titled *Philosophiae Naturalis Principia Mathmatica* in 1687. Thereafter, Newtonian mechanics was soundly established as the fundamental discipline that was believed to be capable of describing everything in the world of physics and physical chemistry. It was not until the mid-1920's, with the discovery of modern quantum theory pioneered by Plank, Einstein, Bohr, Schrödinger, Heisenberg and Dirac, that a reconstruction in our views towards the nature of the world came about. During development period of quantum theory, a variety of other scientists have also contributed significantly (refer to Figure 2.1 for example).[1] This chapter provides an overview of the fundamental concepts and applications of quantum chemistry.

Quantum chemistry methods are also referred to as the electronic structural approaches (ESAs). The basic idea these methods involves determining the electron distribution associated with a fixed set of nuclear positions. ESAs allow one to calculate the energy and related properties of that particular system by solving the Schrödinger wave equation along with auxiliary quantum mechanical laws. Convergent *ab initio* methods and density functional theory are the two electronic structure approaches. These two methods have been extensively used in this project. Simulations employing these two approaches require only the aid of a limited number of physical constants; *viz*., the speed of light (c), Plank's constant (ħ), and basic information concerning the atoms involved (i.e. masses, charges electronic configuration and the relative positions).[2,3] In the remainder of this chapter, only those quantum mechanical procedures that are relevant to this project will be discussed.

2.1.1 Schrödinger's Wave Equation

The Schrödinger wave equation is generally considered to be the origin of modern science. This equation yields the energy as a function of the electronic properties. In Schrödinger's wave equation, instead of trying to find solutions which pin-point the exact positions and velocities of the particles of a system in a given state of motion, this equation approximate the solution by subjecting the system into the appropriate boundary condition.[2,3] The full, time-dependent form of Schrödinger equation is:

$$\left\{ \frac{-\hbar^2}{8\pi^2 m} \nabla^2 + V \right\} \psi(\vec{r}, t) = \frac{i\hbar}{2\pi} \frac{\partial \psi(\vec{r}, t)}{\partial t} \qquad \text{eq 2.1}$$

The denotation ψ, $\vec{r}$, m, $\hbar$, V and t represent for wavefunction, position vector, mass of the particle, Plank's constant, potential field in which the particle is moving and time respectively. In addition to this the gradient ∇ is given by:

$$\nabla = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k} \qquad \text{eq 2.2}$$

When V is independent of time, one can use the separation variables technique and rewrite the wavefunction as the product of a spatial and a time components:

$$\psi(\vec{r},t) = \psi(\vec{r})\,\tau(t) \qquad \text{eq 2.3}$$

Normally the left-hand side of the Schrödinger equation can be abbreviated to $H\psi$ where H is the Hamiltonian operator. A time-independent form of Schrödinger equation can be written as following:

$$\hat{H}\psi(\vec{r}) = E\psi(\vec{r}) \qquad \text{eq 2.4}$$

where $\vec{r}$ is the component vector describing the position of each electron. The energy (E) of the particle can be determined by solving the Hamiltonian ($\hat{H}$) which is made up of kinetic (T) and potential (V) energy terms.[2-4]

$$H = T + V \qquad \text{eq 2.5}$$

$$T = \frac{-h^2}{8\pi^2}\sum_k \frac{1}{m_k}\left(\frac{\partial^2}{\partial x_k^2} + \frac{\partial^2}{\partial y_k^2} + \frac{\partial^2}{\partial z_k^2}\right) \qquad \text{eq 2.6}$$

$$V = \frac{1}{4\pi\varepsilon_o}\left[-\sum_i^{\text{electrons}}\sum_I^{\text{nuclei}}\left[\frac{Z_j e^2}{\Delta r_{iI}}\right] + \sum_j^{\text{electrons}}\sum_{j<i}\left[\frac{e^2}{\Delta r_{ij}}\right] + \sum_I^{\text{nuclei}}\sum_{J<I}\left[\frac{Z_I Z_J e^2}{\Delta R_{IJ}}\right]\right] \qquad \text{eq 2.7}$$

It contains three energy contributing functions, Coulomb, exchange and core terms, where *e*, *Ze* and Z denotes the electronic charge, nuclear charge and atomic number.[2,4]

The product of the ψ and the conjugated ψ' is interpreted as the probability density distribution of the particle which is also referred to as the complex conjugate $|\psi|^2$. The integration is performed and normalised over entire space. The probability of integration thereof must be equivalent to the number of the total particles within the system. Subject to the appropriate boundary condition, the properties of the system can be derived by solving for the ψ using the equation 2.8. Any physical observable for a particular state of a system may be determined using this scheme once the wavefunction is known.[2-4]

$$\int_{-\infty}^{+\infty} |c\psi|^2 \, dv = n_{particles}$$ eq 2.8

The Schrödinger wave equation provides an accurate description of the dual nature of electron particles. However, the biggest drawback is that, whilst the Schrödinger equation is powerful enough to describe almost all chemistry, the exact electronic wavefunction of a system can not be easily resolved by solving of this equation analytically without employing other auxiliary approximations. These approximations can be categorised into two classes; the first class involves with the choice of basis functions to represent the one-electron functions, also known as the molecular orbitals, and second entails with the choice of electron functions to represent the electronic wavefunction. It is with the aid of these approximations (refer to session 2.1.2 to 2.1.11), that the Schrödinger equation can be transformed into an algebraic equation which permits the utilisation of the numerical methods.[2-4]

2.1.2 Born-Oppenheimer Approximation

Prior to solving the wave equation, the Born-Oppenheimer approximation, sometimes known as the Born-Oppenheimer separation, has to be applied. This approximation involves the two parts of the wave equation to be resolved independent of each other. This means that the nuclear and electronic motions can be discretely treated.

The charge of electrons and the nuclei are of the same order of magnitude, therefore the forces exerted on both electrons and nuclei must also be the same. As result, one can assume the changes of the momenta of both electrons and nuclei are also of equal magnitude. However, since the masses of the nuclei are so much larger than the masses of the electrons and the nuclear motion is slow with respect to electron's motion, nuclei must therefore have much less velocities. It is commonly accepted that electrons adjust their positions almost instantaneously to any change of the position of the nuclei. The Born-Oppenheimer approximation assumes that the electronic wavefunction depends solely on the position of the nuclei and independent to their momenta. It allows an electronic Hamiltonian to be constructed while neglecting the kinetic energy terms of the nuclei.[2-4]

$$H = T^{elec}(\vec{r}) + T^{nucl}(\vec{R}) + V^{nucl\text{-}elec}(\vec{R}, \vec{r}) + V^{elec}(\vec{r}) + V^{nucl}(\vec{R})$$ eq 2.9

Where T, V, $\vec{R}$ and $\vec{r}$ are the terms describing the kinetic energy, potential energy and vector position of the electron and the electrons. The electronic structure calculation is performed under the fixed nuclear configuration condition. Hence the positions of all the atoms must be specified clearly before one can compute the electronic energy by solving the electronic Schrödinger equation.[2-4]

2.1.3 Evolution and Classification of Different Methods Relying on Electronic Structural Approach

The total electronic energy of the chemical system can be conceptually partitioned into the following four terms listed below.

$$E = E_T + E_V + E_J + E_{XC} \qquad \text{eq 2.10}$$

E_T represents the kinetic energy of the electrons. E_V represents the Coulombic energy arises from the electron-nuclear attractions. E_J denotes the Coulombic energy of the electrons moving independently with respect to one electron within their own electric field. However, since the movement of each electron does interfere with path of other electrons according to the Pauli Exclusion Principle and each electron does not repel itself, a correction term is therefore required. E_{XC} is the term used to correct these false assumptions generated by (E_T, Ev and E_J). In fact, E_{XC} consists of two terms: E_C (the exchange energy for correcting self-repulsion) and E_C (the correlation energy for correcting motions of electrons with different spins).[5]

The self-consistent field (SCF) concept is the cornerstone of the ESAs. It comprises both the *ab initio* methods and density functional theories in which the E_V and the E_J terms are evaluated using the same expression. However, these methods formulate the E_T and the E_{XC} terms very differently.[5] Figure 2.1 illustrates some of the mainstream quantum chemistry developer. It also shows the extension of the Hartree idea by different developers lead to the range of current quantum expressions.

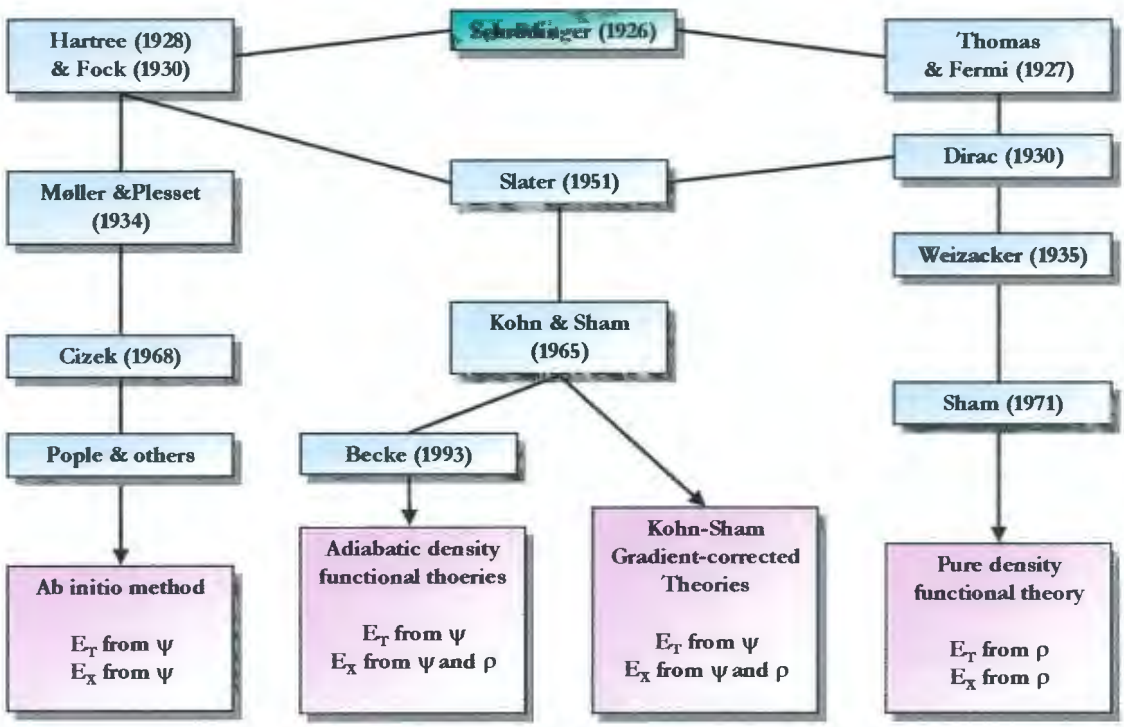


Figure 2.1 Evolution of the electronic structural theories.[5]

2.1.4 Hartree-Fock Self Consistent Field (HF-SCF) Approximation

The HF method is usually considered to be the first step of most *ab initio* chemistry calculations. Hartree introduced the self-consistent field principle in 1928 by proposing a model in which the electron belonging to an atom moves completely independently of the other electrons within the molecule. According to this theory, the total kinetic energy is simply the sum of the kinetic energies of each and every individual electron.[5]

$$E_T = -\frac{1}{2}\sum_i \int \psi_i(\vec{r}_1)\nabla^2\psi_i(\vec{r}_2)\,\delta r \qquad \text{eq 2.11}$$

In 1930, Fock pointed out that Hartree wavefunction did not compliant with Pauli Exclusion Principle due to the expression did not counter for antisymmetry. He resolved the problem by incorporating an exchange functional (E_x, refer to Equation 2.12) into Hartree's expression:

$$E_X = -\frac{1}{2}\sum_i^n \frac{\psi_i^2(\vec{r}_1)\psi_i^2(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|}\delta r_1\,\delta r_2 \qquad \text{eq 2.12}$$

In 1930, Lennard-Jones, Mulliken and Hund demonstrated that the self-consistent field can be extended molecules comprising several atoms. This was the beginning of molecular orbital theory.[5]

2.1.5 Molecular Orbitals

A molecular orbital is the term used for wavefunction that satisfies the Schrödinger wave equation. One conceptualised the $|\psi|^2$ as a probability density of the electron in the system. Molecular orbital theory assumes that ψ can be decomposed into a combination of molecular orbitals. In quantum mechanics, an orbital is written as a three-dimensional mathematical representation which expresses the behaviour of a single electron. A linear combination of a pre-defined set of one-electron functions is known as basis function or basis sets. When solving any polyelectronic system, one can implement the orbital approximation which involves treating each electron independently. Each of the electrons will be subsequently described with its own electronic wavefunction.[2-4] In general, the Hartree-Fock product for an atom, ψ, is defined as the product of all the single functions electron associated with the atomic orbitals, ϕ, for example:

$$\psi = \phi_1\phi_2\phi_3 \ldots\ldots \phi_n \qquad \text{eq 2.13}$$

2.1.6 Electron Spin and Anti-symmetric Wavefunction

The Pauli's exclusion principle states that "no two electrons can have the same set of four quantum numbers, n, l, m_l and m_s". Even though electrons are indistinguishable, electrons can either have spin

As was illustrated in equation 2.32, the implementation this scheme involves the evaluation of, firstly, the total energy of the complex and the energy of each molecule found within the complex. Secondly, molecules within the complex and each possible combination of subsystem were set as the "ghost system" signifies a molecule or complex which only has orbitals but no electrons. The total energy of the complex is then subtracted with the sum of the energies of all the individual molecules as well as the subsystem under the influence of the other ghost system(s). It is important to keep all the energetic terms computed with the exact same basis set or else the scheme is invalid.[13]

2.1.13 Geometry Optimisation

The manner in which the energy of a molecular system depends on its structure or orientation may be represented in terms of a potential energy surface plot. Each point on the surface represents a particular molecular configuration. Geometry optimisation is usually employed to find minima along the potential energy surface, both locally or globally. Hence this technique predicts equilibrium structures of the molecular systems. For ESAs, the geometry optimisation employed numerical minimisation algorithms in conjunction with the quantum mechanical theorem to solve the stability of the wavefunction. Unless a stable wavefunction is acquired, the properties of that particular molecular system cannot be correctly studie.

Once the stationary structure is obtained after the optimisation, the system should be characterised using a frequency calculation. The information yielded from the frequency calculation enables one to one to evaluate the competency of the optimisation by comparing with the actual experimental data. Furthermore, frequency calculation also tests the presence of the imaginary frequency. In the case of locating one or more imaginary frequency numbers, the system computed does not represent a minimum on the potential energy surface. Unless one is searching for a transitional state structure, would yield one imaginary frequency, the presence of any negative frequency node indicates one should perform a re-optimisation with different convergence criteria.[2]

2.1.14 Molecular Partitioning

Quantum mechanics can be used to derive a wide variety of structural related properties. Regardless which quantum mechanical methods employed, the calculation of these properties are basically dependent on a fundamental factor, the electronic distribution. Various applications on analysing the distribution of the electron have been proposed over the years. Since there is no actual quantum mechanical operator for partitioning the molecule, these methods are purely arbitrary. In this section, a few of the more popular electronic partitioning methods are listed. It is important to emphasise that these methods relied on the result obtained from the above mentioned quantum mechanical method to justify the validity of the calculations.

2.1.11.4 Diffuse basis functions

In some cases the above mentioned basis functions can not adequately describe the subtle electronic distortions that occurred within certain systems. These systems include the excited states, molecules with lone pairs (where electrons are relatively far away from the nucleus), anions or compounds with significant negative charges (where the electronic density is more spread out over the molecule) and hydrogen bonding complexes (where slight interaction of the electron cloud can induce vast amount of energetic stabilisation). In order to describe these systems involving the electronic interaction between the groups contribute to a considerable change of overall outcome, one would be compelled to employ more "spread out" basis functions for which are able to account for a larger region of space. These additional basis functions are called diffuse functions and they are essentially large-size version of s- and p-shell functions. Such basis sets are normally termed as 6-31+G(d) or 6-31++G(d) which simply means adding diffuse functions to heavier atoms and to both heavier atoms and hydrogen respectively. It has been suggested that including diffuse function on hydrogen atoms seldom make a significant difference in accuracy in most of these cases. However studies have shown that using the 6-311++(d,p) does lead to a significant improvement on the accuracy for investigating of hydrogen bonding phenomena.[2-4]

2.1.12 Basis Set Superposition Error (BSSE)

Basis set superposition error is an artificial mathematical effect resulting in an overestimation of the interaction energy. This overestimation of energy has two causes:

(1) Accurate evaluation of the interaction energy of a complex system require the use of sufficiently large basis set to correctly describe the multipole moments and the polarizability of the systems. However, when the basis set employed in the calculation is contracted, the margin of error increases.[13]

(2) Within a complex, the wavefunctions of one molecule experiences the presence of the basis functions on the other entity and this non-physical phenomena leads to a false increase in the stabilisation of the wavefunction of the first molecule.[13]

The magnitude of BSSE within a complex is equivalent to the strength of the binding energy.[13,15] There are two approaches to correct the BSSE and subsequently to estimate the binding energy: the chemical Hamiltonian approach (LDH) and counterpoise correction (CP). Salvador and co-workers have shown both methods essentially yield equivalent results.[16,17] Owing to its simplicity, this project employes the Boy-Bernardi's counterpoise correction scheme to estimate the non-bonded interaction.[18]

$$\Delta E_{BSSE} = E(X;Y;Z) - [E(X) + E(Y) + E(Z) + E(X;Y) + E(X;Z) + E(Y;Z)] \qquad \text{eq 2.32}$$

2.1.11.2 Split-valence basis sets

The basis set may be improved by replacing each primitive minimal basis set orbital with two orbitals of different sizes, one large exponent and one small exponent. This is known as the split-valence basis set for GTOs (for example, 3-21G and the 6-21G). Almost equivalently the double zeta basis of the Dunning-Hunzinaga functionals (D95) also induce the same effect. When using this kind of basis set, both orbitals of the set appear to mix in the ratio that gives the lowest energy. Due to the fact that by combining a large orbital and a small orbital, one essentially yields an equivalent of orbitals with intermediate size. Such type of basis set can provide more acceptable fits to the molecular condition.[14]

Similarly the triple-split valence basis sets utilize three different size contracted functions as the replacement of a single STO. Common examples are the 6-311G or other triple zeta basis sets referred to as TZ, TZP and TZ2P.[2-4]

2.1.11.3 Polarisation Basis Functions

Although the split valance basis sets allows the size of orbitals to alter but they do not modify the in shape of the orbital. In certain study cases, the orbital needs to be able to change their shape in order to accommodate molecules containing heavier atoms or more polarisable atoms. It has been argued that the orbital of one atom within a molecule can become smaller when they are attracted by another atom with high electronegativity. This can significantly influence the atoms to distort or polarise the electron density. The problem can be eradicated by adding polarisation basis functions, the basis functions of higher angular momentum quantum numbers beyond what is required for the ground state description of atoms. Polarisation functions allow the distortion of the spherical orbital by mixing orbitals with higher valence shell symmetry. The orbital lobes within this type of basis set can take shapes of orbitals with higher sub-shell number.[2-4]

One normally incorporates these polarisation functions by adding them to a single GTOs and not by contracting them together. Results obtained by adding such polarisation functions to the split valence basis sets can significantly improve the outcome of the calculation, particularly for any system with higher than ground state electronic configuration.

Addition of the d-shell polarisation functions for heavy atoms to the double zeta basis sets is expressed using nomenclature such as 6-31G(d) or 6-31*. The 6-31G(d,p) or 6-31** are used to designate for basis sets which incorporate extra p-shell functions to the hydrogen atoms while adding the d-shell functions onto the heavy atoms.[2-4]

atomic orbitals within a molecule are now more correctly referred to as the "basis functions. The basis set effectively defines the restriction on each electron to a particular region of space around the molecule. Larger basis sets impose less restrictions.[2,3,14]

$$\phi_i^c = \sum_{i\mu}^{N} (S^{-1/2})_{kl} \phi_\mu \qquad \text{eq 2.30}$$

The $c_{\mu i}$ are known as the molecular orbital expansion coefficients.

2.1.11.1 Gaussian-type Minimal Functions

The Schrödinger equation for the hydrogen atom and other hydrogen-like one-electron systems can be solved directly by using atomic orbitals. These atomic orbitals are the product of several radial functions which depend on the distance of the electron from the nucleus as well as a spherical harmonic. However, it is too complicated to use these functions when trying to compute systems with more than two particles. An alternative is to use Slater-type orbitals (STOs) which better represent the real situation of electron density in the valence region and beyond. The STOs only require functions that consist of spherical harmonics and an exponential term, but may show poor accuracy when describing electrons that occur nearer to the nucleus.

Even though STOs do speed up the task of evaluating orbital integrals that are centred on three or four different atoms, it remains a very time-consuming. Pople suggested an alternative approach to evaluate the orbital integrals which employs Gaussian-type orbital (GTO) functions. These GTOs are fitted to the STO by using three linear combination of Gaussian function. As a result of this, the functionals become less complex while still being applicable to since almost all the elements in the periodic table. This type of function is known as the minimal basis set as it contains the fewest function required to describe every atomic orbital within the atom. Such minimal basis set is simple and is fairly computationally inexpensive. However, it fails to yield accurate results when comes to deal with if more sophisticated problems are dealt with. More flexible functions are necessary to account for the different molecular environments. Therefore, it has been recommended that the GTOs should be constricted into separate functions. In this approach, each basis function would then be comprised of several linear combination of functions with fixed coefficients. It is therefore advisable to expand the total number of the GTOs within a basis set, but usually kept to less than six.[2-4] The additional auxiliary functions to the STO-nG basis employed for improving the accuracy of the solutions is commonly as expressed as following:

$$g(\alpha,r) = cx^n y^m z^l e^{-\alpha r^2} \qquad \text{eq 2.31}$$

r is composed of x, y and z coordinates α is a constant determining the radial extent of the function.

2.1.11.2 Split-valence basis sets

The basis set may be improved by replacing each primitive minimal basis set orbital with two orbitals of different sizes, one large exponent and one small exponent. This is known as the split-valence basis set for GTOs (for example, 3-21G and the 6-21G). Almost equivalently the double zeta basis of the Dunning-Hunzinaga functionals (D95) also induce the same effect. When using this kind of basis set, both orbitals of the set appear to mix in the ratio that gives the lowest energy. Due to the fact that by combining a large orbital and a small orbital, one essentially yields an equivalent of orbitals with intermediate size. Such type of basis set can provide more acceptable fits to the molecular condition.[14]

Similarly the triple-split valence basis sets utilize three different size contracted functions as the replacement of a single STO. Common examples are the 6-311G or other triple zeta basis sets referred to as TZ, TZP and TZ2P.[2-4]

2.1.11.3 Polarisation Basis Functions

Although the split valance basis sets allows the size of orbitals to alter but they do not modify the in shape of the orbital. In certain study cases, the orbital needs to be able to change their shape in order to accommodate molecules containing heavier atoms or more polarisable atoms. It has been argued that the orbital of one atom within a molecule can become smaller when they are attracted by another atom with high electronegativity. This can significantly influence the atoms to distort or polarise the electron density. The problem can be eradicated by adding polarisation basis functions, the basis functions of higher angular momentum quantum numbers beyond what is required for the ground state description of atoms. Polarisation functions allow the distortion of the spherical orbital by mixing orbitals with higher valence shell symmetry. The orbital lobes within this type of basis set can take shapes of orbitals with higher sub-shell number.[2-4]

One normally incorporates these polarisation functions by adding them to a single GTOs and not by contracting them together. Results obtained by adding such polarisation functions to the split valence basis sets can significantly improve the outcome of the calculation, particularly for any system with higher than ground state electronic configuration.

Addition of the d-shell polarisation functions for heavy atoms to the double zeta basis sets is expressed using nomenclature such as 6-31G(d) or 6-31*. The 6-31G(d,p) or 6-31** are used to designate for basis sets which incorporate extra p-shell functions to the hydrogen atoms while adding the d-shell functions onto the heavy atoms.[2-4]

atomic orbitals within a molecule are now more correctly referred to as the "basis functions. The basis set effectively defines the restriction on each electron to a particular region of space around the molecule. Larger basis sets impose less restrictions.[2,3,14]

$$\phi_i^c = \sum_{i\mu}^{N} (S^{-1/2})_{kl} \phi_\mu \qquad \text{eq 2.30}$$

The $c_{\mu i}$ are known as the molecular orbital expansion coefficients.

2.1.11.1 Gaussian-type Minimal Functions

The Schrödinger equation for the hydrogen atom and other hydrogen-like one-electron systems can be solved directly by using atomic orbitals. These atomic orbitals are the product of several radial functions which depend on the distance of the electron from the nucleus as well as a spherical harmonic. However, it is too complicated to use these functions when trying to compute systems with more than two particles. An alternative is to use Slater-type orbitals (STOs) which better represent the real situation of electron density in the valence region and beyond. The STOs only require functions that consist of spherical harmonics and an exponential term, but may show poor accuracy when describing electrons that occur nearer to the nucleus.

Even though STOs do speed up the task of evaluating orbital integrals that are centred on three or four different atoms, it remains a very time-consuming. Pople suggested an alternative approach to evaluate the orbital integrals which employs Gaussian-type orbital (GTO) functions. These GTOs are fitted to the STO by using three linear combination of Gaussian function. As a result of this, the functionals become less complex while still being applicable to since almost all the elements in the periodic table. This type of function is known as the minimal basis set as it contains the fewest function required to describe every atomic orbital within the atom. Such minimal basis set is simple and is fairly computationally inexpensive. However, it fails to yield accurate results when comes to deal with if more sophisticated problems are dealt with. More flexible functions are necessary to account for the different molecular environments. Therefore, it has been recommended that the GTOs should be constricted into separate functions. In this approach, each basis function would then be comprised of several linear combination of functions with fixed coefficients. It is therefore advisable to expand the total number of the GTOs within a basis set, but usually kept to less than six.[2-4] The additional auxiliary functions to the STO-nG basis employed for improving the accuracy of the solutions is commonly as expressed as following:

$$g(\alpha,r) = cx^n y^m z^l e^{-\alpha r^2} \qquad \text{eq 2.31}$$

r is composed of x, y and z coordinates α is a constant determining the radial extent of the function.

$$\rho(r) = 2\sum_{i}^{\infty} |\psi_i(r)|^2$$ eq 2.28

LDA works well in solid-state physics but is ineffective to compute most chemical applications that do not satisfy the restriction of slowly varying electron density. Moreover, the exchange correlation energy functional term of the LDA sometimes tends to over-estimate the total energy. This leads to the necessity for the development of other improved functionals. The generalized gradient approximations (GGAs) demonstrated a much better accuracy in predicting the chemical properties. This method computes the density fluctuations via the gradient of the density.[5]

A more popular class of GGA is the hybrid functionals proposed by Becke. This class of functionals combine the exact Hartree-Fock exchange with conventional correlation functionals. These hybrid functionals can provide significant improvement on the estimation of various molecular properties over other GGAs. The most frequently used hybrid density functional theorem is conceivably the Becke-three-Lee-Yang-Parr (B3LYP) functionals[11] where there exchange functionals are combined with the Hartree-Fock exchange (E_x^{HF}) and the Slater local-density exchange ($E_x^{Slater}[\rho] \approx \alpha\rho^{4/3}$) together with the addition of the correlation functional that included both uniform (local) and gradient (non-local) terms.[2-4] Moreover, three scaling parameters are customized into the algorithm as is shown in the following equation:

$$E_{xc}(B3LYP) = A \times E_x^{Slater} + (1\text{-}A) \times E_x^{HF} + B \times DE_x^{Becke} + E_c^{VWN} + C \times DE_c^{Lee,Yang,Parr}$$ eq 2.29

where the ΔE_x^{Becke} is the exchange "correction" term, E_c^{VWN} represents the correlation functional based on a uniform electron gas[12] and $\Delta E_c^{Lee,Yang,Parr}$ denotes the non-local corrections to the correlation functional taken from the LYP expression. The "best" scaling parameters that found to fit a number of calculations were declared to be 0.80, 0.72 and 0.81 for A, B and C respectively.[5,12]

Density functional theory is generally most suitable for ground state theory. It performs poorly on excited state modelling. Apart from the highest occupied state, the eigenvalues obtained from DFT do not always necessarily coincide with the physical eigenvalues of the system. Furthermore, it has long been argued that this level of theory is unsuitable for calculations involving systems with substantial degree of hydrogen bonding.[13] Further discussion regarding the comparisons between the MP2 and B3LYP calculations will be dealt with in chapter 4.

2.1.11 Basis Sets

The use of basis sets to represent the solution of the electronic Schrödinger equation can be dated back to 1927 when Heitler and London made use of a basis set of hydrogen orbitals to describe the electronic structure of some simple molecules. Molecular orbitals can be expanded as a linear combination of atomic orbitals. The mathematical representations of such linear combination of

2.1.10.2 Density Functional Theory

An alternative electronic correlation method is density functional theory (DFT). Pure density functional theory was originally developed by Hohenberg and Kohn and later modified by Kohn and Sham. The pure density functional theory was subsequently further modified by Becke and co-workers. They created a new category of density functional theory in which the functionals are tailored by the addition of hybrid theory and few parameters.[8] Recently the application of DFT in the computational chemistry has gained overwhelming popularity as these methods offer a significant amount of electron correlation with only a minimal fraction increase on the computational cost in comparison with Hartree-Fock.[9,10] Despite its simple origins, DFT is not a post-Hartree-Fock method sincere here the electronic charge density is the fundamental variable rather than the wavefunction. The electron spin and spatial parts of the equation are different in DFT than to those implemented in Hartree-Fock theory.[5]

For a given wavefunction, ψ_n, the Hamiltonian, H_n, can be represented by the following equation:

$$H_n = T + U + V_n \qquad \text{eq 2.25}$$

T and U are the common kinetic and interaction energy operators, respectively. V_n is the ground state electron density corresponding with the external potential. Density functional theory is based on the conception that within a many-electron system there is a direct connection between the external potential and the electron density:[2-4]

$$V(r) \leftrightarrow \rho(r) \qquad \text{eq 2.26}$$

According to this theory, all properties are a derivation of the density. The electron density yields the potential, which then determines the Hamiltonian and the wavefunction $\psi[\rho]$. All physical properties can then be determined. Kohn and Sham showed that the exact ground-state electronic energy for an N-electron system could be written as:

$$\langle\psi[\rho]|H|\psi[\rho]\rangle = \int dr\, v(r)\rho(r) + T[\rho] + V_{ee}[\rho] = E_v[\rho] \geq E_0(\psi) \qquad \text{eq 2.27}$$

with $T[\rho]$ is the electronic kinetic energy functional and $V_{ee}[\rho]$ is the electron-electron interaction energy functional.[2-4]

A variety of density functional theorems have been established. Kohn-Sham's local density approximation, LDA, is noteworthy. This method stated that in a one-electron formalism, there exists a number (n) of non-interacting electrons in the system. Each electron in the system will experience only a local potential $V_L(r)$ that yields exactly the same density as the system of interacting electrons with potential $V(r)$. The electron density is represented by the sum of the densities of the $n/2$ doubly occupied single-particle spatial orbitals. Since the exact expression for the exchange correlation energy functional which compensates for the electron-electron interaction is not yet been proposed, therefore this forced one to use an approximate functional.[2-4]

larger one-electron basis sets than Hartree-Fock method to yield converged results. Therefore CI and CC calculations generally required extensive amount of computational cycles.[2-4]

In 1933, Møller and Plesset established a particularly successful method which was able to resolve the correlation problem. This method is also based on the variational principle to account for the correlation energy. When examining a system where static correlation effects are negligible, the remaining correlation can be dealt with as a small perturbation of the original Hartree-Fock method. Calculation of the energy using the perturbation theory is based on the following principles. In the basic perturbation method, one decomposes the Hamiltonian matrix in the following way:

$$H = H_0 + \lambda V \qquad \text{eq 2.21}$$

Perturbation theory works on the idea that if one has possible solutions to the answer to the problem, one can work out an answer which is closely to the related problem. Thus, one would like to make the simplest possible choice for H_0. λV is the perturbation applied to H_0. Given a choice of, the CI secular equation becomes:

$$(H_0 - E)C = -VC \qquad \text{eq 2.22}$$

If the outcome generated from the modified Hamiltonian operator is not satisfactory, one can continue solving the difference between the Hamiltonian operators as a small perturbation to the first solution by literately treating the above equation according to the following expression:

$$C^n = (E - H_0)^{-1}\, VC^{(n-1)} \qquad \text{eq 2.23}$$

A higher approximations to the desired C vector can be generated as its eigenvalue E can be corrected at each iteration with the using the equation stated below:

$$(E - E_0)(C^0)^T C = -(C_0)^T\, VC \qquad \text{eq 2.24}$$

Even though Møller-Plesset perturbation theory may sometimes overcorrect Hartree-Fock results, this form of electron correlation is still relatively economical in terms of the computational processing time when compared to the CC or CI methods. The second order Møller-Plesset perturbation theory (MP2) does provide a reasonable proportion of the correlation energy. Although a higher order Møller-Plesset perturbation methods, such as the third and fourth order Møller-Plesset perturbation methods (MP3 and MP4) may yield better results. However, they are much more computationally demanding than MP2 but only able to yield an improvement of results about a few tenths of a kcal/mol^{-1}. Thus, for this the reason, MP2 remains as the most widely used method for computing correlated structures and binding energies of hydrogen bonded complexes.[2,4]

requires the user to predefine the electron spin of the state. This causes problems for some study cases where one needs to investigate what that spin state is and how it relates to structure and reactivity.[7]

The unrestricted Hartree-Fock (UHF) method is the popular approach in simulating open-shell systems. This method adopts what has been called the "Different Orbitals for Different Spins" approach. There are two sets of spatial orbitals used within a computation: those with α- and β- spin electrons. This separation of orbitals has been found to reproduce attributes of the open-shell system most closely.[7]

2.1.10 Electron Correlation Methods

In general, there are four sources that can lead to an incorrect energy estimate when using a Hartree-Fock level calculations:

(1) relativistic effects

(2) deviations from the Born-Oppenheimer approximation

(3) the neglect of the complete treatment of the electron correlation

(4) the incompleteness of the basis set assumption

The first type of error is only important for molecules containing heavy atoms. The second is negligible for ground-state systems. Hence, these types of errors do not affect the study of carbohydrates. However, the third and the forth problem may significantly interfere with the accuracy of the computation. The third source of errors is discussed in this section and the fourth source of error will be explored further in the discussion of basis sets section.[2]

In order to compensate for the lack of electron correlation of Hartree-Fock theory, many electron correlation methods employ different schemes to approximate and optimise the wavefunction. Møller-Plesset (MP) perturbation theory, configuration interaction (CI), or coupled cluster (CC) methods are the three post-Hartree-Fock electron correlation methods. Density functional theory (DFT) also takes electron correlation into account, although strictly speaking, it is not a true post-Hartree-Fock electron correlation method.[2-4]

2.1.10.1 Møller-Plesset (MP) Perturbation Theory

The MP perturbation theory is the more popular electron correlation method due to the fact that the CI and CC methods are both based on the principle of involving large matrices in which a wavefunction built with linear combination of Slater determinants. It is crucial to bare in mind that in order to yield acceptable results: most of the explicit electron correlation treatments require much

$$E(\Xi) \geq E(\psi) \quad ; \Xi \neq \psi \qquad \text{eq 2.17}$$

Where Ξ is the symbol chosen for antisymmetric normalised function of electron coordinate.[2-4]

2.1.8 Roothaan-Hall Approximations

The Hartree-Fock equations are a set of integro-differential equations using basis vectors. The Hartree-Fock equation can be transformed into a set of algebraic equations but gives a generalized eigenvalue which is a problem to solve. Roothan and Hall derived the following coefficient expansion equations using the variational principle,

$$FC = SCe \qquad \text{eq 2.18}$$

where C is the molecular orbital expansion coefficient matrix, e is the diagonal matrix of orbital energies and F is called *Fock matrix* and is defined as follows:

$$F_{uv}=H_{uv}+ P_{\lambda\sigma} [(\mu v|\lambda\sigma)-1/2(\mu\lambda|v\sigma)] \qquad \text{eq 2.19}$$

where H_{uv} represents the energy of an electron in the field of the nuclei. $P_{\lambda\sigma}$ is called the density matrix element that can be defined by the following expression:

$$P_{\lambda\sigma} = 2\Sigma C_{\lambda i} C_{\sigma i} \qquad \text{eq 2.20}$$

It has been found that the Fock matrix and the molecular orbital theory depend on their expansion coefficients and these equations can only be solved in the iterative manner. The $(\mu\lambda|v\sigma)$ term represents the two-electron-repulsion integrals and under the Hartree-Fock treatment each electron experiences the average field of all the other electrons without there being any instantaneous electron-electron interaction.[2-4]

2.1.9 Multiplicity and Open Shell Methods

The multiplicity of a molecule is determined by the number of unpaired electrons the molecule contains. In this project, only the normal ground-state equilibrium structures consisting of only paired electrons were studied. These systems are termed "singlets" or also commonly known as "closed-shell molecules". However other methods have been developed to handle the "open-shell system" which contains either free radicals with one unpaired electron or biradicals with two unpaired electrons respectively. The assumptions made in the restricted Hartree-Fock method are inappropriate for to these open-shell molecules. There is more than one way of resolving this type of problem.[7]

The Restricted Open-Shell Hartree Fock method (ROHF) is one method that extends the HF principle to treat open-shell systems. It uses the open-shell wavefunction that is well approximated with a single, spin-adapted linear combination of Slater determinants. However, this method requires

up (alpha) or spin down (beta) state. Since electrons are fermions and the orbitals which describe the system must satisfy the fermion wavefunctions characteristics. Electrons must obey Fermi Dirac statistics and a sign change is necessary if two electron are interchanged .[2-4] In 1930, Fock pointed out that this invalidity of Hartree wavefunction since it did not satisfy the Pauli Exclusion Principle. The wavefunction must be antisymmetric with respect to electron interchange. Fock demonstrated that a Hartree product could be made antisymmetric by adding and subtracting the permutations of the Hartree product.[6]

$$f(i,j) = -f(i,j) \qquad \text{eq 2.14}$$

$$E_X = -\frac{1}{2}\sum_{i}^{n}\sum_{j}^{n}\iint \frac{\psi_i(r_1)\psi_j(r_1)\psi_i(r_2)\psi_j(r_2)}{|\, r_1 - r_2 \,|}\,\delta r_1\,\delta r_2 \qquad \text{eq 2.15}$$

Later, Slater illustrated that the resulting wavefunction is simply the determinant of a particular determinant matrix. This matrix is now known as the Slater matrix determinant. For example, in a closed shell system containing n electrons, the electrons are assigned two at a time to a set of molecular orbitals.[5] In the Hartree-Fock self consistent field (HF-SCF) method, the wavefunction will increase with a subset of all possible of Slater determinants.

$$\psi = \frac{1}{\sqrt{n!}}\begin{vmatrix} \phi_{n/2}(1)\alpha(1)\phi_{n/2}(1)\beta(1)\ldots\ldots\ldots\ldots & \phi_{n/2}(1)\alpha(1)\phi_{n/2}(1)\beta(1) \\ \phi_{n/2}(2)\alpha(2)\phi_{n/2}(2)\beta(2)\ldots\ldots\ldots\ldots & \phi_{n/2}(2)\alpha(2)\phi_{n/2}(2)\beta(2) \\ \phi_{n/2}(3)\alpha(3)\phi_{n/2}(3)\beta(3)\ldots\ldots\ldots\ldots & \phi_{n/2}(3)\alpha(3)\phi_{n/2}(3)\beta(3) \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ \phi_{n/2}(n)\alpha(n)\phi_{n/2}(n)\beta(n)\ldots\ldots\ldots\ldots & \phi_{n/2}(n)\alpha(n)\phi_{n/2}(n)\beta(n) \end{vmatrix} \qquad \text{eq 2.16}$$

where α and β denote the two electrons of different spin within an orbital. Swapping two electrons simply corresponds to interchanging two rows of the determinant which subsequently lead to swapping the signs. The possible permutations are represented by n! term. The wavefunction can be thus solved using a single Slater determinant is used when the orbitals is variationally optimised.[2-4]

2.1.7 Variational Principle

The variational principle states that the expectation energy value obtained from using an approximate wavefunction is always equal to or higher than the exact solution of the Schrödinger equation for a particular Hamiltonian operator. An improvement of the wavefunction description results on lesser margin of error between the calculated value and actual value. This means that if one has a wavefunction that contains flexible parameters, one can then adjust the parameters to minimise error in the expectation value towards the exact solution.

2.1.14.1 Mulliken Population Analyses and Merz-Singh-Kollman Scheme

A convenient way of partitioning the molecular electron distribution and gaining an idea of its charge distribution has its origin from the Mulliken population analysis.[19] However, even though Mulliken population analysis is quick to perform, it does suffer from a few drawbacks. This method rigidly assigns an atomic charge to the atom around which the orbital is centred. It apportions electrons between the pair of atoms equally which is somehow problematic for electrons sharing between p, d or f shell orbitals and s shell orbital. Furthermore, this method is highly basis set dependent and the calculation schemes can sometimes allocated a negative number of electrons or more than two electrons within an orbital. This clearly violates the Puali's exclusion principle.[4,20]

The Merz-Singh-Kollman scheme uses the principle proposed by Cox and Williams of fitting the electrostatic potential to points selected on a set of concentric van der Waals envelope around each atom.[21] This method is more acceptable than the Mulliken-Löwdin population analysis. We employed this method to roughly estimate the atomic charges assignment scheme in force field parameterisation (refer to chapter 5). However, the Merz-Singh-Kollman analysis has a disadvantage of being conformationally dependent. In other words, it tends to be defective for studying atoms which are buried within a molecule.[4,21] Therefore one can not depend on this method to solve problems which requires accurate analysis. Bader's Atoms in Molecule theory (AIM) overcomes these drawbacks by offering a more theoretically sounded approach of partitioning the electron population.

2.1.14.2 Bader's Atoms in Molecule Theory (AIM)

From a chemistry level, atoms are the fundamental constituents of all matter and within each atom, the positively charged nuclei are immersed in a cloud of negatively charged electrons. The electron clouds which may be represented as electron densities, $\rho(r)$, are distributed throughout space but mostly are confined within the region of space bounded by the attractive field exerted by the nuclei. Once the optimised geometry and the wavefunctions of a system are known, AIM can transform the distribution of electron density of the atoms found at any particular part of the molecule into a 2-dimensional gradient vector map. This vector map is also known as a relief map (see Figure 2.2) and it allows one to transforms the electron density into a measurable property by examining the topology of the map.[22]

One can slice a part of molecule (shown in Figure 2.2(i) highlighted with blue lines) and transform it into a relief map (Figure 2.2(ii)). In Figure 2.2(ii), the $\rho(r)$ is iteratively traced out based on the gradient vectors trajectory as well as the direction of the density trajectory at each point of the map iteratively. Then $\rho(r)$ gradient of the neighbouring point is calculated and a new direction is obtained. By repeating of this process, one can trace out the total trajectory of $\nabla\rho(\mathbf{r})$ of the system and a relief map can be generated. The positively charged nuclei, also known as the attractors, exhibit a strong attracting force on the negatively charged gradient vector field of the electron density resulting the

maximum $\rho(r)$ to be found near the position of each nucleus. All trajectories terminate at each nucleus. At the same time, one can also derive the minima and saddle points along the gradient vector trajectory. By joining all the saddle points on the map, one can successfully partition the "atoms" found on that particular molecular slice according to the topology of the $\rho(r)$.[23]

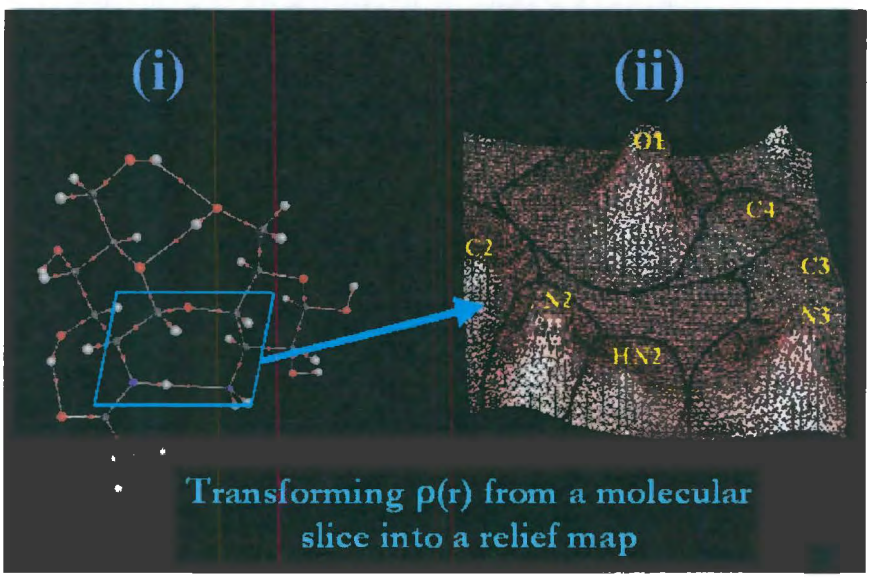


Figure 2.2 **Graphical representation of transforming a slice of the GlcNAc-α-(1→4)-3-amine-Glc into a 2-dimensional relief map and partitioning the atoms base on the $\rho(r)$ topology.**

Since AIM enables the partitioning of the "atoms" found on a particular molecular slice according to above mentioned procedure, thus similarly by stacking all the slices that made up of the whole molecule, one accurately determine the amount of $\rho(r)$ found within each 3-dimensional atom in the 3-dimensional molecule. The 3-dementional characterisation of $\rho(r)$ defines the atomic boundary which is often referred as an atomic basin. Each basin is the region of space cross through by the trajectories but terminating at a given attractor. Since a single attractor is associated with each basin, the merger of an attractor and its basin can be regarded as an atom. Within this boundary, the atomic volume can be calculated.

Alternatively, an atom can be defined as a region of space bounded by surfaces through which there is a zero flux in the gradient vector field of the electron density (refer to equation 2.38). An inter-atomic surface is defined by the set of trajectories that terminate at a point where the "zero-flux" boundary condition is satisfied for every point r_s on this surface $S(r_s)$.

$$\nabla\rho(r)n(r_s) = 0 \qquad \text{eq 2.33}$$

The $n(r_s)$ is the unit vector normal to the surface at r_s. The thick black lines found in Figure 2.2(ii) indicate the region that satisfy this "zero-flux" gradient condition and the atoms are partitioned accordingly.

Each topology of $\rho(r)$ is associated with a bond critical point (BCP). A BCP denotes a point where the $\nabla\rho(r) = 0$. There are four types of possible points associated with a three-dimensional system and they are defined as following:[23]

(3,-3) All curvature of $\rho(r)$ at the critical point is negative and $\rho(r)$ is a local maximum at r_c. These type of points represents the location of the nucleus.

(3,-1) Two curvature of $\rho(r)$ is negative with local maximum $\rho(r)$s at the r_c in the plane of the two axes. The third curvature is positive and $\rho(r)$ is a local minimum at the r_c along the axis perpendicular to the plane. This type of points represents the location of the bond-critical point.

(3,+1) Two curvature of $\rho(r)$ is negative with local minimum $\rho(r)$s at r_c in the plane of the two axes. The third curvatures are positive and $\rho(r)$ is a local maximum at r_c along the axis perpendicular to the plane. This type of points represents the location of the ring-critical point.

(3,+3) All curvature of $\rho(r)$ at the critical point is positive and $\rho(r)$ is a local minimum at r_c. This type of points represents the location of the cage-critical point.

2.1.14.3 Investigating Hydrogen Bonding using AIM

AIM is particularly useful when considering the controversial hydrogen bonding phenomenon and this technique has been extensively applied in characterising hydrogen bonding.[24-27]. For the purpose of our study, we direct out attention to the fundamental meaning of (3,-1) critical point. Associated with each bond-critical point, there is a set of trajectories that start at infinity (i.e. the nucleus) and terminate at the bond-critical point. Therefore, between two nuclei, an atomic interaction line with maximum $\rho(r)$ can be traced along the gradient vectors trajectory of the electron density. This atomic interaction line is known as the bond-path. Hence, the existence of a (3,-1) critical point implies that $\rho(r)$ occurs between two nuclei and linked by the associate bond path. In other words, the existence of a (3,-1) critical point indicates the formation of a bond. One can study the relationship and the relative characteristics of the atomic interaction of any two atoms in the molecule based on the bond-critical point. The covalent characteristic of the hydrogen bonding has clearly been illustrated over the past using this theory[24-27] (refer to chapter 4 for more the discussion of the physical characteristic of hydrogen bonding).

Apart from the presence of (3,-1) BCP, it was found that a true hydrogen bond must comply with these two sets of properties listed below:[28,29]

(1) The following topological changes must be found within the (3,-1) BCP: (a) the $\rho(r)$ locates within the range of 0.002 to 0.04 a.u, (b) The $\nabla\rho(r)$ value fall between the range of 0.015 to 0.15 a.u. and (c) mutual penetration of hydrogen and acceptor atoms.[28,29]

(2) The following integrated properties must be found between the hydrogen atom and the acceptor: (a) an increase of the positive atomic charge of hydrogen atom along with a decrease of the atomic charge of the acceptor atom, (b) the occurrence of a energy stabilisation, (c) a total decrease of dipolar polarisation and (d) decrease in the atomic volume of hydrogen atom along with an increase of the atomic volume of the acceptor.[28,29]

2.2 Applications

Electronic structural approaches resolve the molecular wavefunctions in terms of basis function coefficients. Along with a number of approximations, these applications involve calculating the energy of a given arrangement of molecular coordinates generally obtained from geometrical optimisation and they are based on the fundamental physical principles. Therefore ESAs do not depend on the nature of the system under investigation. Hence, making correct use of the electronic structural techniques for predictive calculations leads to a much reliable investigation. The ESAs are non-biased and the results obtained are generally highly comparable to the experimental findings. Chapters 4 and 6 rely profoundly on the electronic structural methods to analyse intra- and inter-molecular hydrogen-bonding within carbohydrates. The ESA methods were also to chapter 5 for assisting the force field parameterisation process.

2.3 Conclusion

ESAs are most useful for simulating the detailed electronic structure of a system or predicting the mechanism of a reaction. Traditionally, these methods are limited to small molecules owing to the high computational cost of the calculations. As the availability of computing power increases, the range of problems addressed by electronic structural methods have widened. However, the computational cost for study of a large chemical system on the atomic level using electronic structural approach still remains exceedingly high. Therefore, there is still a need for empirical force field methods in order to model such system. The principles and the techniques relating the empirical force fields (EFF) are covered in the next chapter.

REFERENCES

(1) Laidler, K. J. *The World of Physical Chemistry*, Oxford University Press Inc.: New York, **1995**.

(2) Foresman, J. B.; Frisch, Æ. *Exploring Chemistry with Electronic Structure Methods*, Gaussian, Inc.: Pittsburgh, **1996**.

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

(4) Leach, A. R. *Molecular Modelling: Principles and Applications*, Addison Wesley Longman Ltd.: Singapore, **1996**.

(5) Gill, P. M. W. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p678.

(6) Hohenberg, P.; Khon, W. *Phys Rev A* **1964**, *136*, 864.

(7) Jensen, F. *Introduction to Computational Chemistry*, John Wiley & Sons Ltd.: Chichester, **1999**.

(8) Becke, A. D. *J. Chem. Phys.* **1996**, *98*, 11623

(9) Hu, C., H.; Chong, D. P. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p664.

(10) Chen, J. Y. J.; Naidoo, K. J. *J. Phys. Chem. B* **2003**, *107*, 9558.

(11) Lee, C.; W., Y.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785.

(12) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200.

(13) Del Bene, J. E. In*Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p1263.

(14) Dunning, J. T. H.; Peterson, K. A.; Woon, D. E. *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p88.

(15) Lii, H. J. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p1271.

(16) Salvador, P.; Paizs, B.; Duran, M. *J. Comput. Chem.* **2001**, 22(7), 765.

(17) Salvador, P.; Duran, M.; Fradera. *J. Chem. Phys.* **2002**, 116(15), 6443.

(18) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, 19, 553.

(19) Mulliken, R. S. *J. Chem. Phys.* **1955**, 23, 1833.

(20) Cornell, W. D.; Chipot, C. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p 258.

(21) Singh, U. C.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, 5, 129.

(22) Bader, R. F. W. *Atoms in Molecules- A Quantum Theory*, Oxford University Press: Oxford, **1990**.

(23) Bader, R. F. W. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p64.

(24) Klein, R. A. *J. Comput. Chem.* **2003**, *24*, 1120.

(25) Grabowski, S. J. *J. Phys. Chem. A* **2001**, *105*, 10739.

(26) Kosov, D. S.; Popelier, P. L. A. *J. Phys. Chem. A* **2000**, *104*, 7339.

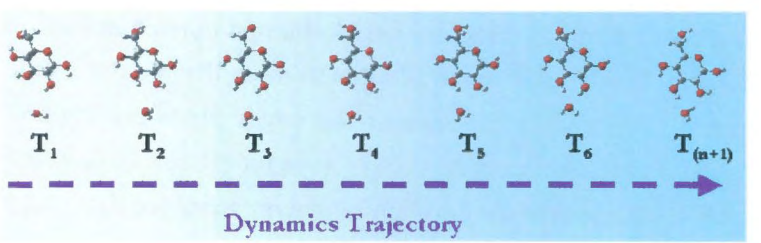
(27) Vila, A.; Carbello, E.; Mosquera, R. A. *Can. J. Chem.* **2000**, *78*, 1535.

(28) Pacios, L. F.; Gómez, P. C. *J. Comput. Chem.* **2001**, *22(7)*, 702.

(29) Popelier, P. L. A. *J. Phys. Chem. A* **1999**, *103*, 2883.

Chapter Three

Empirical Force Field Methods: Principles and Techniques



3.1 Theoretical Background

Empirical force field (EFF) methods are the fast-growing discipline of computational chemistry. The conformational behaviour and energetic properties of molecules at atomic level can be evaluated using EFF methods provided that the composition of the system and a set of spatial coordinates for all the atoms within the system are both defined properly. The reason EFF works at all is due to the validity of several assumptions. The Born-Oppenheimer approximation is perhaps the most important assumption amongst all. Without it, one would not be able to describe the system as the function of the nuclear coordinates. In the EFF methods, molecules are treated at the atomic level and the electrons are not explicitly treated. The original concepts of these methods aimed to mimic the vibrational spectroscopy. The atoms and bonds are assumed to behave like balls and spring. Commonly, the EFF methods employ several algorithms derived from Newtonian principles to account for the dynamic nature of the system. Using EFF methods in combination with the principle of statistical mechanics results in the conversion of microscopic properties of the vibrant system into macroscopic deterministic observables.[1,2]

A variety of simple potential energy functions were implemented in the EFF calculations resulting in an empirical model. CHARMM (Chemistry at HARvard Molecular Mechanics) is such an EFF program. Modelling of isolated macromolecules, condensed phase systems, crystalline and amorphous solids to are all possible using CHARMM. It features a wide range of functions and analytical tools that enables researches to study the relativistic molecular properties. Even though the program has been extensively tested and modified since the original version of CHARMM published in 1983, the fundamental structure of the program is still unchanged.[3,4] In this chapter, the important aspects of the CHARMM program are overviewed along with selected principles and theories that are central in EFF simulation methods.[3,4]

3.1.1 The Potential Energy Functions

The potential energy functions are the mathematical expressions used to describe the dependence of the energy of a system on the coordinate of its atomic composition. These functions along with their parameters are known as the force field. The energy, E, is a function of the atomic positions, R, of all the atoms in the system which is usually expressed in terms of Cartesian coordinates. There are two broad classes of energy functions, *viz* 'bonded' and 'non-bonded' interactions. For the purpose of this work, bonded interactions are taken between atoms that are connected covalently by no more than three bonds (i.e. not more than four atoms) whereas the non-bonded interactions are taken as the atoms are more than two bonds apart and in addition to those that are not chemically bounded. The value of the energy is calculated as the sum of the covalently bonded terms (i.e. E_{bonded}, which describes the all the bonded interactions such as stretching, angles bending and dihedral angle torsions in a molecule), and a sum of the non-bonded energy terms (i.e. $E_{non\text{-}bonded}$, which takes van der

Waals, electrostatic and other higher terms interactions into consideration).[5] The pictorial representation of the empirical force field terms are shown in Figure 3.1.

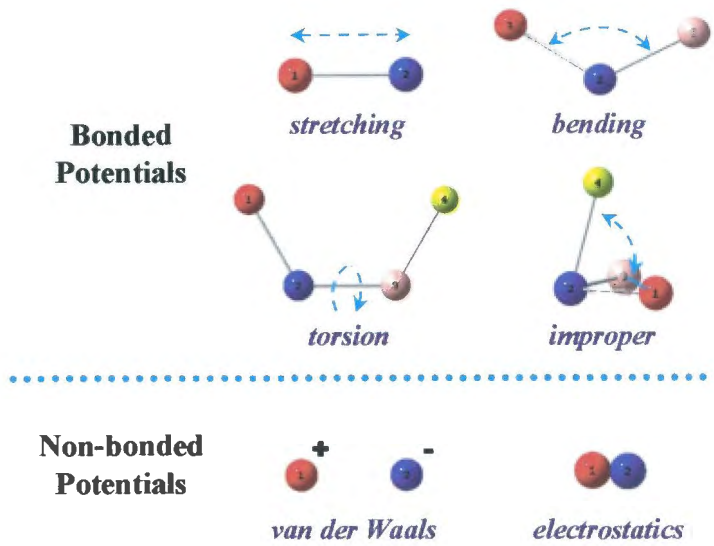


Figure 3.1 Pictorial representation of the terms included in a force field.[1]

The values of the force constants and the corresponding parameters of the various potentials are often estimated from experimental data such as infrared-Raman stretching frequencies, microwave spectroscopy data, NMR plots and high resolution crystal structures. However, sometimes these experimental data are unavailable and deriving force field parameters can be achieved through comparing the EFF calculated properties with the results generated from high level quantum mechanical calculations (refer to chapter 5 for more examples).[1,2,5]

3.1.1.1 Covalently-Bonded Potential Functions

Depending on the empirical force field methods used, the implementation of covalently-bonded potential functions may differ slightly. Typically, the principle of deriving the total bonded energy generally comprises the terms below:

$$E_{bonded} = E_{stretching} + E_{bending} + E_{torsion} + E_{improper} + E_{supplementary\ terms} \qquad \text{eq 3.1}$$

These potential functions behaves like penalty functions, whereby an energy penalty is imposed once the specific term deviates from its ideal geometry.[1,2]

Bond Stretching: In many force fields, harmonic bond stretching functions are employed. The Morse potential function is known to be the most accurate model for bond stretching but has the

disadvantages of increasing computational time. CHARMM describes the bond stretching using the simple harmonic functions.

$$E_{stretch} = \sum_{bond} K_b (b - b_0)^2 \qquad \text{eq 3.2}$$

In this approximation, the energy of a bond is a quadratic function of the deviation of its length, b, from its ideal bond length, b_0. The force constant, K_b, governs the strength of the bond. The values of both the ideal bond lengths and the force constants are different depending on the chemical constituents of the covalently boned atom pairs. The harmonic function suffers from the limitation as it only roughly describes the actual behaviour of the covalent bond. This function tends to deviate too steeply when the bond length is over stretched from its ideal value. It provides a fairly inaccurate representation of large deformation dissociation, but overall, it produces satisfactory results.[1,2]

Angle Bending: In CHARMM, the angle bending motions, $E_{bending}$, involving three bonded atoms modelled using the simple harmonic function.

$$E_{bending} = \sum_{angle} K_a (\theta - \theta_0)^2 \qquad \text{eq 3.3}$$

$E_{bending}$ thus depends on the amount that the actual bond angle, θ, deviates from its equilibrium value θ_0. Once again the values of θ_0 and K_a are specific to the chemical types of the atoms that makes the angle. The only drawback of using an harmonic potential function to describe the E_{angle} is the lack of accuracy when modelling the strained ring systems. Nevertheless, the overall outcome of employing this function is satisfactory.[1,2]

Dihedral Rotation: The torsional interactions are represented by the bond dihedral potential functional. This term accounts for the behaviour of steric barriers occurred between atoms separated by three covalent bonds. A Fourier series is generally used and several functions are summed together in order to reproduce the appropriate torsional rotation motion. CHARMM's torsional angle terms are defined to be periodic and is often expressed using cosine functions as follows:

$$E_{torsional} = \sum_{dihedral} K_\omega (1 - \cos (n\omega)) \qquad \text{eq 3.4}$$

K_ω is the rotational energy barrier height and n is the periodicity of the torsion angle ω. It should be emphasized that unless the non-bonded interactions (van der Waals interaction and the atomic charge assignments), described in section 3.1.1.2, are specified correctly, the K_ω parameter does not fully correspond to the actual complete profile of the molecule.[1,2] Hence in order to model the

torsional interactions correctly, both the dihedral and van der Waals parameters need to be appropriated chosen. Careful selection process of these values must be carried out and this will be demonstrated in chapter 5.

Improper Torsion Angles: Many chemical groups involve arrangements of 4 or more atoms in a plane. The improper dihedral term, also known as the out-of-plan bonding terms, is used to uphold the planarity and the chirality of the molecule. CHARMM employs a harmonic energy term to represent the improper dihedral potentials. The topology of the bonding atoms is different from the scheme of the 'proper' dihedrals.[1,2]

$$E_{improper} = \sum_{improper} K_i (\varphi - \varphi_0)^2 \qquad \text{eq 3.5}$$

Urey-Bradley Potential Terms: One additional potential function is also implemented in CHARMM namely the Urey-Bradley potential function. This term is an interaction based on the distance between atoms separated by two bonds. It takes the 1-3 interaction into account rather than an explicit potential. The Urey-Bradley term is used to overcome the difficulty of mimicking the force constant matrix within a molecular vibrational spectrum.

$$E_{Urey\text{-}Bradley} = \sum_{UB} K_{UB} (S - S_0)^2 \qquad \text{eq 3.6}$$

Unless this term is included, the evaluation of the molecular vibrational energies would not yield reasonable agreement with experimental data.[3,4]

3.1.1.2 Non-bonding Interactions

Commonly in EFF, any non-bonded interaction involving more than two atoms is neglected. The non-bonded interactions potential function mainly consists of two components; the van der Waals and the electrostatic interaction energy. Sometimes, other terms such as hydrogen bonding, stretch-bend, bend-bend, torsion-bend and/or polarization functions can be specifically introduced in CHARMM. But it is generally sufficient to employ only the van der Waals and the electrostatic interaction for the condensed phase modelling.[3,4]

$$E_{non\text{-}bonded} = E_{van\ der\ Waals} + E_{electrostatic} + E_{additional\ terms} \qquad \text{eq 3.7}$$

Van der Waals Interactions: This term models the atomic size parameter while trying to mimic the energy induced from the electron correlation effect between the atoms. The van der Waals interaction between any two atoms is known to be attractive at long range between all atoms. This is

due to the London dispersion forces initiated from fluctuations in the charge distribution in the electron cloud on one atom which further induced a direct dipole in a second atom or a molecule. In the short interatomic range where overlapping of the electron clouds occurs as the distance between the atoms is less than their atomic radii, the interaction is strongly repulsive. The Buckingham potentials and Lennard-Jones potentials, also known as 6-12 potential, are the two expressions that have been widely accepted to model van der Waal interaction. The Buckingham potential is particularly useful for investigating short contacts. This is due to the fact that the Lennard-Jones term can sometimes over-estimate the steepness of the repulsion between two atoms that are positioned less than the optimal inter-atomic distance. However, a drawback of using Buckingham potential is that at the very short inter-atomic distance, the function inverts and yields a negative infinity value. It is this reason CHARMM utilizes the Lennard-Jones potential to model the van der Waals interaction. CHARMM describes the interaction energy using the atom-type dependent constants, r_i and r_j in which r^{-12} term represents the short range repulsion and the London dispersion forces are accounted for by the r^{-6} term.

$$E_{van\ der\ waals} = \sum_{\substack{Non\text{-}bonded \\ pairs}} 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \qquad \text{eq 3.8}$$

The overall shape of the curve below is influenced by the collision diameter, σ (the distance at which the energy required is zero) and the well-depth, ϵ. The separation of atoms r* is roughly equal to the sum of the van der Waals radii of that atoms. The depth of the curve determines the interaction energy, E*. Both r* and E* are atom type dependent.[1,2]

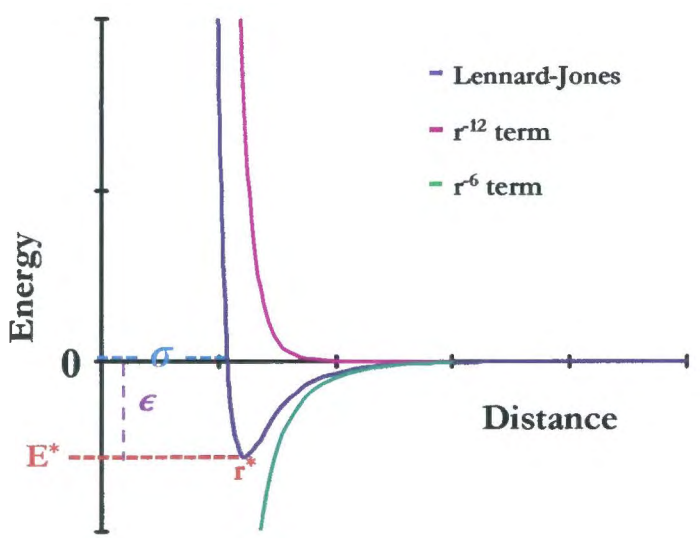


Figure 3.2 A typical van der Waals interaction energy plot.

Coulombic Interactions: CHARMM employs the Coulomb potential to model the electrostatic interaction between charged atom pairs.

$$E_{Coulombic} = \sum_{\substack{Non\text{-}bonded \\ pairs}} \frac{1}{4\pi\varepsilon} \frac{q_i q_j}{r_{ij}^2} \qquad \text{eq 3.9}$$

where D is defined as the effective dielectric function for the specific medium and r is the interatomic distance between the atom pair having charges q_i and q_k. The influence of the atomic charge on the overall potential energy will be discussed further in Chapter 5.

3.1.2 Energy Minimisation

Energy minimisation is the technique used to search for the conformation of the molecule with the lowest potential energy. Such technique iteratively refines the geometry of a molecule based on the relative energy of the geometrical conformers. It is advisable to perform an energy minimisation before starting a molecular dynamics simulation as such these technique are useful in remove any undesirable steric clashes existing between atoms or molecular groups which can otherwise cause the distortion of the structure and result in an unstable simulation.

Energy of the system varies according to atomic positions. Thus conceptually the geometrical change of a system can be visualised as moving along on a multi-dimensional "surface" called the potential energy surface. All energy minimisation techniques are dedicated on searching along a downhill path on the potential energy surface. Hence these techniques tend to find the local minimum on the energy surface that is nearest to its starting point. These approaches may fail to reach the global minimum due to the high conformational energy barrier. A thorough conformational search should therefore be carried out using simulated annealing procedures (described in chapter 5).

The "simplex" and "sequential univariate" methods are the two common non-derivative minimisation approaches. However, this type of approaches is somehow not as effective and reliable as the first and second order derivative minimisation methods. Many modern minimisation approaches are now based upon the principle of higher order algorithms which require only a fractional increase of computationally cycles but yielding more dependable results. CHARMM offers five different energy minimisation algorithms and each has its own strengths and weaknesses.[6,7] Since only two minimisation methods are employed in this project, the discussion will thus be limited to these two methods only.

Steepest Descent: The method of steepest descent (SD) employs the first derivatives of the potential energy of the molecule with respect to the molecular atomic coordinates. A new atomic configurations is generated from the current configuration using the following equation:

$$X_{(new)} = X_{(old)} + \lambda \times s \qquad \text{eq 3.10}$$

where s is the unit vector that points the search direction along the force ($F = -dE/dx$), λ is the step-size that governs the amount of change allowed per step along the search path while X is denoted for the molecular atomic coordinates.

SD is a fast and practical of method when dealing with the initial stages of an energy minimization where the forces are strong. However, it suffers from poor convergence. This algorithm sometimes yields structures that are trapped in a local energy minimum of relatively high energy since the algorithm may fail to cross the potential energy barriers.[6,7]

Conjugated Gradient: The conjugated gradient (CG) method employs first-derivatives. It produces a set of directions prevents the minimisation oscillates within the narrow valley of the global minima. In conjugated gradient method, the gradient vector at each point is orthogonal. Furthermore, the search direction is a combination of that given by the force and the previous search direction:

$$s = g + \gamma \times s_{(old)} \qquad \text{eq 3.11}$$

where γ is given by

$$\gamma = (g_{(new)} \times g_{(new)}) / (g_{(old)} \times g_{(old)}) \qquad \text{eq 3.12}$$

The CG method is slightly slower than SD, but the method can produce larger coordinate alterations. This can often lead to atomic configurations with lower energies.

Other high-order minimisation methods are also available in CHARMM. These are based on the principle of the Newton-Raphson method. Multidimensional Newton-Raphson methods require both first and second derivatives of the energy about the energy surface. However, other second-derivatives minimisation methods are usually computationally costly in terms of both the memory and speed. Thus for large molecular systems, they are less frequently employed than the first-order minimisation methods.[6,7] In this project, the minimisation scheme generally employs a combination of SD and CG.

3.1.3 Statistical Mechanics

The goal of molecular dynamics (MD) is to gain insight and to predict the macroscopic properties of a system based on its microscopic behaviour. In order to achieve the goal, one needs to convert the microscopic information obtained from simulations to macroscopic observables using the rigorous theoretical formulation of statistical mechanics.

Thermodynamics, structural and transport properties can be calculated from microscopic observables combined with the statistical mechanics methods. Within the microscopic state, the system can be completely described by the phase space parameters: the three atomic positions, q_x, q_y and q_z, and the three momenta description, p_x, p_Y and p_z. For a system with N particles, the system consists of a $6N$ dimensional phase space. Each point of the phase space represents a state of the system. A collection of all possible points in the phase space which have different microscopic states but have an identical macroscopic state is known as an ensemble.[2,8]

3.1.4 Calculating Averages from a Molecular Dynamics (MD) Simulation

An ensemble average (< A >) is the averages value of the quantity A, determined for a large number of replicas of the system. MD simulates an ensemble as a function of time thus a time average property ($<A>_{time}$) to be computed as a function of the microscopic states of the system in the available phase space. The basic idea is that if one allows the system to evolve in time indefinitely, the system will eventually sample through all possible states. A molecular dynamics simulation must be sufficiently long in order for all representative configurations to be sampled.

$$<A>_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t))\mathrm{d}t \approx \sum_{t=1}^{M} A(p^N, r^N) \qquad \text{eq 3.13}$$

In equation 3.13, the <A> term is expressed as a function of the momenta, p, and the positions, r, of the system. The integration is over all the possible values of r and p. M is the number of time steps in the simulation and $A(p^N,r^N)$ is the instantaneous value of A.

Based on statistical mechanics, the Ergodic Hypothesis is a method for computing molecular properties of the system from MD simulations. It states that statistical averages obtained from ensembles are equivalent to the "time-averaged" properties since a computer simulation cannot follow a path in phase space over an infinite amount of time. The simulation time is limited to a finite period and the phase space may not thus be completely sampled. Hence one must be satisfied with the following statement:

$$\overline{A} \approx <A> \qquad \text{eq 3.14}$$

A number of different ensembles exists, each possessing its own unique characteristics. In this work, only two different statistical namely the ensembles, microcanonical and the isothermal-isobaric ensemble are employed.

In the basic MD setting the macroscopic observables N (number of atoms), V (volume) and E (energy) are held constant. This is known as the "microcanonical ensemble" (NVE). The NVE ensemble can be shown to be thermodynamically equivalent to the other ensembles. The NVE

ensemble can be shown to be thermodynamically equivalent to the other ensembles. The NVE ensemble is particularly useful for simulating *in vacuo* dynamics owing to the fact that this ensemble does not need to assign pressure the individual atoms. Since *in vacuo* dynamics generally deals with system consist of a small quantities of atoms, assigning pressure over each atoms would normally lead to an exceeding of the limits of pressure fluctuation tolerance.

Previously, the stochastic procedures were used to regularly reset the particle velocities to random values picked from the appropriate Maxwell-Boltzmann distribution. More recently, deterministic and reversible equations of motion such as Gaussian and Nosé-Hoover dynamics have been developed. Instead of completely constraining the kinetic energy, Nosé has shown that under quite general conditions, a canonical distribution can be approached by introducing some kind of thermostat for maintaining the energy near a desired value. In 1980 Andersen introduced an additional energy term that could be coupled with the Nosé-Hoover algorithms and enable the system to change its volume.[9] On the same year, Parrinello and Rahman extended this method to allow for non-isotropic stretching and shrinking of box sides. Morris and Evans further extended these ideas and made it possible to retain both the pressure and the kinetic energy near a desired value. This brought about the isothermal-isobaric ensemble (NPT).[8]

Recently, NPT ensembles have been shown to successfully mimic the experimental conditions by fixing the number of atoms, N, pressure, P, and temperature, T of the system. The NPT simulations employs a thermostat and phase transitions are known to best studied using this ensemble.[2,8]

3.1.5 Molecular Dynamics (MD)

In the late 1950's, Alder and Wainwright presented one of the very first molecular dynamics simulations involving the interactions of hard spheres. Their studies drew attention to many important aspects concerning the behaviours of simple liquids. Starting with a suitable initial conformation, molecular dynamics makes use of Newton's equations of motion to determine how the system will transform over a period of time. A long dynamics simulation also provides insight into its various thermodynamic properties. MD can inform us of the various conformational dependent behaviours of the system, such as what conformational states are the system accessible to, and how the conformational fluctuations varies with the timescale.

A variety of MD algorithms have been developed extensively in the past few years. The default simulations using under conditions the NVE ensemble. However, it is preferable to perform simulations under the more realistic NPT ensemble. CHARMM is capable of utilising a wide range of dynamics algorithms and in addition to the original Verlet algorithm, it has now implemented with the Velocity Verlet and the Leapfrog Verlet algorithms. Moreover, apart from the original approach based on the work perform by Berendsen *et al.* for controlling temperature and pressure, the Nosé-

3.1.5.1 Equations of Molecular Dynamics

During a molecular dynamics simulation, the positions, velocities and accelerations of the particles at each time increment can be obtained by integrating Newton's equations of motion. Using one of the numerical integrators described in the next section and the potential energy functions described below, one can generate a configurational time series is known as a trajectory.

Newton's equation of motion is given by:

$$F_i = m_i a_i \qquad \text{eq 3.15}$$

where F_i, m_i and a_i represent the force exerted on particle i, its mass and its acceleration respectively. Since the force is also given by the gradient of the potential energy, these two equations can be combained:

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \qquad \text{eq 3.16}$$

where V is the potential energy of the system. This permits one to obtain the derivative of the potential energy to the changes in position as a function of time.

Newton's Second Law of motion states:

$$F_i = m_i a_i = m_i \frac{dv}{dt} = m_i \frac{d^2 x}{dt^2} \qquad \text{eq 3.17}$$

The velocity can be derived from integration as follows:

$$v = a_i t + v_o \qquad \text{eq 3.18}$$

Further integrates yields:

$$r = v_i t + r_o \qquad \text{eq 3.19}$$

Combining the equations 3.18 and 3.19 would yield the value of position (r) at time t as a function of the acceleration, a_i, the initial position, r_o, and the initial velocity, v_o:

$$r = a_i t^2 + v_o t + r_o \qquad \text{eq 3.20}$$

The acceleration is the derivative of the potential energy with respect to the atomic position, *r*:

$$a_i = -\frac{1}{m_i}\frac{dV}{dr_i} \qquad \text{eq 3.21}$$

Hence if the initial positions of the atoms, the distribution of velocities and the acceleration are known, the deterministic trajectory can easily be simulated using with various the numerical integration algorithms (listed in the section below).[2,10]

3.1.5.2 Dynamic Integration Algorithms

The equations of motion must be solved numerically, owing to the complicated relationship between the potential energy (V) and the phase space (r_i) of the system. Numerous numerical algorithms have been developed for integrating the equations of motion. These integration algorithms are derived from the Taylor series expansions where r is the position, v, the velocity, a, is the acceleration and the constants are denoted by A, B, C, D and E so forth:[10]

$$r(t + \delta t) = r(t) + v(t)\delta t + 1/2A(t)\delta t^2 + \ldots\ldots \quad \text{eq 3.23}$$

$$v(t + \delta t) = v(t) + B(t)\delta t + 1/2C(t)\delta t^2 + \ldots\ldots \quad \text{eq 3.24}$$

$$a(t + \delta t) = D(t) + E(t)\delta t + \ldots\ldots \quad \text{eq 3.25}$$

Basic Verlet: The Verlet method employs a second-order differentials equation and formed by combining the Taylor series expansions of $t+\delta t$ and $t-\delta t$, in which the velocities are generated by combining the two expansions about the position at time, t, such as is shown below:

$$r(t + \delta t) = r(t) + v(t)\delta t + 1/2a(t)\delta t^2 \quad \text{eq 3.26}$$

$$r(t - \delta t) = r(t) - v(t)\delta t + 1/2a(t)\delta t^2 \quad \text{eq 3.27}$$

this produces a new expression:

$$r(t + \delta t) = 2r(t) + r(t + \delta t) + \delta t^2\, a(t) + \ldots\ldots \quad \text{eq 3.28}$$

The new positions at time $t+\delta t$ are calculated by gathering the positions and accelerations at time t along with the positions from time $t-\delta t$. The Verlet algorithm is known to be straightforward and it requires only a modest computational storage capacity since this algorithm integrates the positions directly, thus avoiding having to explicitly integrating velocities. However, the Verlet algorithm does suffer from a shortcoming of only yielding result with moderate precision and it may also overestimates the kinetic energy.[1,10]

Leap-Frog Verlet: When using the Verlet algorithm, numbers of the order of δt^2 are added to numbers of the order δt^0. This may introduce numerical error. The leapfrog Verlet method was developed to overcome such problem. This algorithm determines the velecities on odd steps, and positions on even steps:

$$r(t + \delta t) = r(t) + \delta t\, v(t + 1/2\delta t) \quad \text{eq 3.29}$$

$$v(t + 1/2\delta t) = v(t - 1/2\delta t)r(t) + \delta t\, a(t) \quad \text{eq 3.30}$$

First a new mid-step velocity is generated from the velocity and then used to calculate the new position. The new velocity is then calculated from:

$$v(t) = (1/2)v(t + 1/2\delta t) + (1/2)v(t - 1/2\delta t) \qquad \text{eq 3.31}$$

The temperature scaling is much more practical when using the leap-frog algorithm since the temperature control is determined by the velocity scaling.[1,10]

Velocity Verlet: The handling of the velocity terms is not as close to ideal in both of the above mentioned algorithms. The velocity Verlet algorithm is known to produce a more accurate velocity description. The velocity Verlet approach is equivalent to basic Verlet algorithm when one eliminates the velocities terms. This method stores positions, velocities, and accelerations using the equations below that are implemented in two stages. First, the new positions at time $t + \delta t$ are calculated

$$r(t + \delta t) = r(t) + \delta t\, v(t) + 1/2\delta t^2 a(t) \qquad \text{eq 3.32}$$

$$v(t + \delta t) = v(t) + 1/2\delta t[a(t) + a(t + \delta t)] \qquad \text{eq 3.33}$$

Then the velocities at mid-step are calculated using:

$$v(t + 1/2\delta t) = v(t) + 1/2\delta t a(t) \qquad \text{eq 3.34}$$

The forces and acceleration at time $t + \delta t$ are then calculated and consequently a new velocity is calculated:

$$v(t + \delta t) = v(t + 1/2\delta t) + 1/2\delta t a(t+\delta t) \qquad \text{eq 3.35}$$

The velocity Verlet algorithms may be more precise, but for molecular dynamics, calculating the forces becomes time-consuming. On the other hand, the basic Verlet's can yield various predictions with a low precision.[1,10] Hence, the Leap-frog Verlet algorithm is the preferred choice for our dynamics simulation.

3.1.6 Deriving Ensemble Variables

The initial velocities assignments in NVE ensemble are usually performed with the following equation to confirm the required temperature. The velocity v_x in the x direction of atom i is randomly assigned by using a Maxwell-Boltzmann or Gaussian distribution at a given temperature T:

$$P(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{K_B T}\right] \qquad \text{eq 3.36}$$

Using the conventional NVE ensemble can lead to various problems when one attempts to simulate the more complex systems. For example, the total density of a system consisting of a liquid/liquid interface is difficult to determine. In some other cases, even when the macroscopic properties of the system are properly determined, the appropriate volume of the microscopic simulation cell is unable to be resolved easily. In recent years, the NPT ensemble dynamics has produced simulation results close to the experimental findings. This ensemble permits the total volume of the box to fluctuate with the particle coordinates being adjusted instantaneously to the fluctuation of the box size.[8]

All solution simulations in this project were performed using a NPT ensemble where the pressure and temperature were maintained constant using the Langevin Piston method. In this scheme the pressure is controlled using an equation of the following type:

$$\frac{d^2V}{dt^2} = \frac{1}{W}[P(t) - P_{ext}] - \gamma\frac{dV}{dt} + R(t) \qquad \text{eq 3.37}$$

The volume, V, is treated as a dynamic variable, while W is the "mass" of the piston with a collision frequency of γ and applied with a random force $R(t)$ that is determined from a Gaussian distribution having a zero mean and variance.

For a system containing N number of atoms, the temperature (T) then can be obtained from the velocities using the following expression:

$$T = \frac{1}{(3n)}\sum_{i=1}^{N}\frac{|p_i|}{2m_i} \qquad \text{eq 3.38}$$

Moreover, the average potential energy can be acquired using the equation stated below:

$$E_{potential} = \langle E_{potential} \rangle = \frac{1}{M}\sum_{i=1}(E_{potential})_i \qquad \text{eq 3.39}$$

where M is the number of configurations in the molecular dynamics trajectory and $E_{potential}$ is the potential energy of each related conformation and the average kinetic energy ($E_{kinetic}$) is obtained from following equation:

$$E_{Kinetic} = \langle E_{Kinetic} \rangle = \frac{1}{M}\sum_{j=1}^{M}\left[\sum_{i=1}^{N}\frac{m_i}{2}v_i \times v_j\right]_j \qquad \text{eq 3.40}$$

in which N is the number of atoms, and m_i and v_i represent the mass and the velocity of particle i.[8]

3.1.7 Treatment of Non-bonded Energy Terms

Since there are N×(N-1)/2 non-bonded interactions to calculate in an N atom system, evaluating the non-bonded interactions is the most time consuming part during the dynamic simulation. To reduce computational cost, the non-bonded interactions that involve atom pairs separated by more than a given distance are usually neglected. This distance range is known as the cutoff distance. Employing such a non-bonded cut-off does not affect the van der Waals interactions accuracy, since that when the distance between the two atoms increases, the interaction rapidly decay to minute magnitude. However, the charge-charge interactions are of much longer range due to the 1/r dependence on the interaction energy. For the charged atoms, the use of a cut-off is more challenging. Mathematical tricks such as the Ewald Particle Mesh Summations method has been used to approximately adjust for long-ranged electrostatic interactions.[2,8]

In an ideal situation, the non-bonded energy between every atom pair should be calculated. However, the total number of pairwise interactions increases by the square of the number of atoms. To increase the efficiency of the simulation, the interactions between two atoms separated by a distance greater than the cutoff distance is neglected. Several different approaches have been developed for terminating the van der Waals interaction above the cutoff distance and these methods are being described and illustrated as follows:

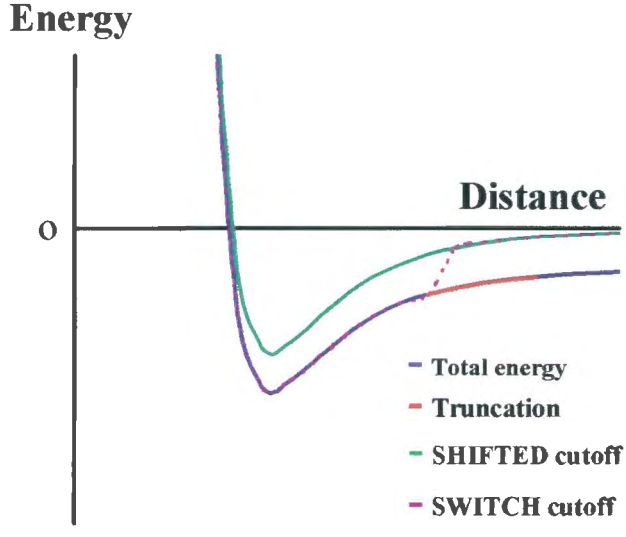


Figure 3.3 Several cutoff methods principles used in molecular dynamics.

Truncation: interactions are completely ignored once the inter-atomic distances are greater than the cutoff range (represented by the red mesh lines in the above figure). This method is not often used as it can lead to large fluctuations in energy.[2]

The SHIFT cutoff method: this method modifies the entire potential energy surface so that the interaction potential is set to zero at the cutoff distance. However, the equilibrium interatomic distances will be slightly decreased when applying this method.[2]

$$S_{shift}(r_{ij}) = \left(1 - \frac{2r_{ij}^2}{r_{cut}^2} + \frac{r_{ij}^4}{r_{cut}^4}\right) \quad r_{ij} < r_{cut}$$
$$= \quad 0 \quad r_{ij} > r_{cut} \qquad \text{eq 3.41}$$

The SWITCH cutoff method: this method reduces the interaction potential over a range of distances. In other words, two cutoff values are specified. When the inter-atomic distance is less than the first cutoff, the potential energy is calculated as normal. Between the first and last cutoff, the energy is gradually switched to zero. In principle, it must be ensured that the cutoff distance should be sufficiently large when using the SWITCH function. A weakness of using this method is that sometimes the equilibrium structure can be perturbed when the forces in the switching region are strong.[2,11]

$$S_{switch}(r_{ij}) = \quad 1 \quad r_{ij} \le r_{on}$$
$$= \left(\frac{(r_{off}^2 - r_{ij})^2 \ (r_{off}^2 - 2r_{ij}^2 + 3r_{on}^2}{(r_{off}^2 - r_{on})^3}\right) \quad r_{on} < r_{ij} \le r_{off}$$
$$= \quad 0 \quad r_{ij} > r_{off} \qquad \text{eq 3.42}$$

3.1.8 Harmonic Constraints of Hydrogen Atoms

Hydrogen atoms have substantially lower molecular weight than other atoms and hence their motion tends to be faster. If one allows hydrogen atoms to vibrate during a simulation, one needs to set the dynamics time step to be a fraction of a femtosecond in order for the algorithms to be soundly integrated. This solution is somehow time consuming and in order to resolve this problem all bond distances that involve hydrogen are fixed with harmonic constraints. SHAKE[12] is one such method that is commonly employed. This algorithm constrains the coordinates with a set of harmonic constraints. The constraining tolerance of the SHAKE algorithm constraints are normally set to be lower than other molecular fluctuation arises from other sources. However, on the same time, the motion of the molecule studied should not be affect by SHAKE.

3.1.9 Solvent Models in a Molecular Dynamics Simulation

Water is essential to all biochemical processes and thus it remains as the most studied condensed phase system. However, it is also one of the most difficult solvents to study since it is an associative liquid wherein hydrogen bonding gives rise to the water's unique characteristics. The water models in CHARMM generally take the form of effective pair potentials. Numerous models have been developed for water such as ST2, TIPS (transferable intermolecular potential functions) and SPC (simple point charge).[13,14] An improved version of SPC was used in some of our simulations, the SPC/E model. The SPC/E model has been known to reproduce the important structural features determined experimentally. It has been shown that the difference between the TIP3P and SPC/E water models in carbohydrate simulation is inconsequential.[13,14] Therefore both these water model were employed in our simulation.

MD simulation may be performed using either implicit or explicit solvent models. The Self-Consistent Reaction Field methods is one of the approaches which enables dynamics simulations to perform within the implicitly solvent medium. There are several available models and they are all based on the principle of treating solvent models as a reaction field with a continuum of uniform dielectric constant. When the solute is modelled within the "field" of the solvent, such methods require less computational cycles. However, frequently it is necessary to include the solvent model explicitly. This is especially important in the case of studying carbohydrate in an aqueous medium. The exocyclic functional groups of sugars establish hydrogen bonds readily with the adjacent water molecules in specific orientations which affect the water properties significantly. This solvent structuring is generally responsible for much of the behaviour of the carbohydrate solvation. The three-dimensional analyses of such anisotropy of the solvent ordering has proven to be a useful tool in predicting the solvation nature of the carbohydrate. It is worth noting that when solvent molecules are strongly bound to the carbohydrate solute, significant changes in the overall carbohydrate topology result owing to the bulkness of the solvent molecules. The bridging water effect can only be studied by MD simulations under explicit solvent conditions.[15]

In principle, simulations performed using fully explicit solvents to account for the physicochemical solvation environments should yield more realistic behaviour of the solute than simulations which employ implicit solvating environments. If water molecules are explicitly included in a simulation a solvent boundary conditions must be imposed. Periodic boundary conditions (PBC) ensure that the water molecules do not diffuse away from the solute and that no a void would be created around the solute during the simulation. The PBC also provide a means for calculating the macroscopic properties of a system using a relatively small number of solvent molecules from MD simulations.[2,15]

3.1.10 Periodic Boundary Conditions (PBC)

Condensed phase simulations using PBC allow the modelled particles to experience force similar to those in a bulk solution. In this approach, system under this condition is treated as the central simulation box and the box is replicated in all three dimensions (i.e a cubic box has a total of 26 neighbouring boxes.) Molecules situated near at the edge of the real box can experience the interaction with the molecules on the neighbouring box (or known as the image cell) as well as the nearby molecules. Whenever a certain amount of molecule leaves the central box during the MD simulation, the same amount of molecules will enter the central box from the image cell. Hence the molecules in the real system will always be constant and will never experience the edge effect. Figure 3.4 illustrates for the case of in two dimension PBC.

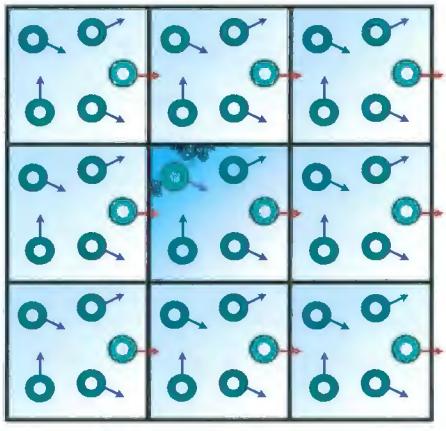


Figure 3.4 **Periodic boundary conditions for a two-dimensional system. The molecule leaving the central box is replaced an image of the same molecule entering from the from the neighbouring image cell.**

There are two important considerations when setting up the PBC for MD simulations. First, the box length should always be more than twice that of the cutoff distance. This ensures that each particular molecule in the real box does not interact with its own image in the surrounding boxes. Second, the magnitude of the cutoff distances should be considered. If the cutoff distances is less than 10Å, to ensure a reasonable degree of convergence of the electrostatic terms. However, increasing the magnitude of the cutoff distance increases the computation costs as has been previously discussed.[2,8]

3.2 Analytical Methods

A molecular dynamics simulation typically produces large quantities of data. The vectors describing the phase space, force and a variety of instantaneous values of all calculated properties, is often saved

at regular intervals to permit thorough analyses. Structural parameters of interest and time dependent properties can be computed using the analytical tools such as those described in following subsections. Some of these tools serve as means of interpreting the complicated information obtained from the molecular dynamics configuration into comprehensible graphical visualisation format.

3.2.1 Pucker Analysis

The pyranose ring is able to adopt the same conformational changes as cyclohexane, namely chair, boat envelop and twist ring conformations. These conformations can be conveniently analysed by the generalized puckering parameters developed by Cremer and Pople.[16] Experimental findings have shown that the glucopyranose rings are virtually rigid in solution and exist only in the $^1C_4$ chair conformation. Puckering analysis has been used one to verify the conformational changes of the sugar ring during the simulations[17]. We make use of this procedure throughout all the dynamics runs involving glucopyranose to ensure the fluctuation of the cyclic ring dihedrals remains within the acceptable range.

3.2.2 Pair Distribution Function (PDF)

Pair distribution function (PDF), sometimes known as the radial distribution function, is a simple analytical method used to describe the structure of a system. Particularly in the case of investigating liquids, PDF make uses the spatial atomic coordinates from each MD trajectory frame and a probability density function to represent a probability distribution in terms of integrals. The PDF g(r), indicates the probability of finding a pair of atoms of specified types at a distance r from one another. To calculate such a function from a trajectory, atoms or molecules specified are sorted into histograms of distance bins. The number of atoms or molecules in each bin, $N(r)$, is then averaged over the entire simulation taking the bulk water number density ρ in to account. This is mostly defined using equation listed below

$$g(r) = \frac{1}{4\pi\rho r^2}\frac{dN(r)}{dr_i} \qquad \text{eq 3.43}$$

A probability density function can be seen as a "smoothed out" (continuous) version of a histogram. The PDF can be determined experimentally by evaluating the diffraction of X-rays, neutrons or electron diffraction. One can validate a computed trajectory by comparing the PDF obtained from simulation with the corresponding data obtained experimentally.[2]

PDF provide insight into the structuring of liquids. However, since they are radially averaged, they are unsuitable to for investigating anisotropic liquid structuring using such method. Other techniques must be applied in order to study systems with non-spherical symmetry.[15]

3.2.3 Spatial Distribution Function (SDF)

The spatial distribution function (SDF) is also not computed as a continuous function, but is obtained by dividing the three dimensional space into small but discrete blocks. Each solvent molecule (or atom specified) can be assigned and "binned" into the corresponding block. After the process has been performed for the entire simulation, the probability of the SDF can be approximated. The equations 4.44 and 4.45 are generally employed for SDF analyses and a graphical representation this process is illustrated in Figure 3.5:

$$g(r) = Z(\alpha/\pi)^{3/2} \times e^{-\alpha r^2} \qquad \text{eq 3.44}$$

only the oxygen water atoms were employed for the calculation. Gaussian distribution function centred on the electrons for each oxygen atom was used to approximate the distribution of electron of the oxygen. Notation Z is the total number of electrons for the atom and α is calculated constant so that the function drops to 10% of its maximum value when found within the atom's van der Waals radius. Ideally, the maximum density binning grid size was found to be approximately 0.5A wide. The densities in each box are summed and then analogously normalised using the expression listed below:

$$Dens_{norm}(i, j, k) = Dens(i, j, k) \times [(n_{bin(x)} \times n_{bin(y)} \times n_{bin(z)}) / (n_{atom} \times n_{electrons} \times n_{frame})] \qquad \text{eq 3.45}$$

Notation $n_{bin(x)}$, $n_{bin(y)}$, and $n_{bin(z)}$ are the number of grid allocate in the x, y and z axis of the boundary box respectively where n_{atom}, $n_{electrons}$ and n_{frame} define the number of atoms whose density is to be calculated, the number of electrons for that particular atom and the number of frames selected for calculations respectively.

Mapping the distribution of the solvent molecules allows one to attentively examine the characteristics of solvent structuring around a particular solute or solvent molecule. By predefining a molecular spherical polar coordinate frame with respect to a reference molecule, the probability density of solvent around that reference molecule during the simulation can be determined. It is important to perform the following procedures prior to the analyses.[15,18,19]

(1) In order to account of changes in conformation as well as the diffusion and rotation of the reference molecule, each atom need to be treated separately. The conformations of the molecule analysed are superimposed to match a reference molecule in by means of translation and/or rotation.

(2) It is important to employ a sufficiently long MD trajectory in order to ensure adequate dynamics sampling.

(3) It is important to select an equal number of MD frames for SDF analyses. The density mappings are sensitive to the amount of frames used in the calculation. Using unequal amounts of frames allows more motion and the possibility of more density blurring of probability density one solute than the others.

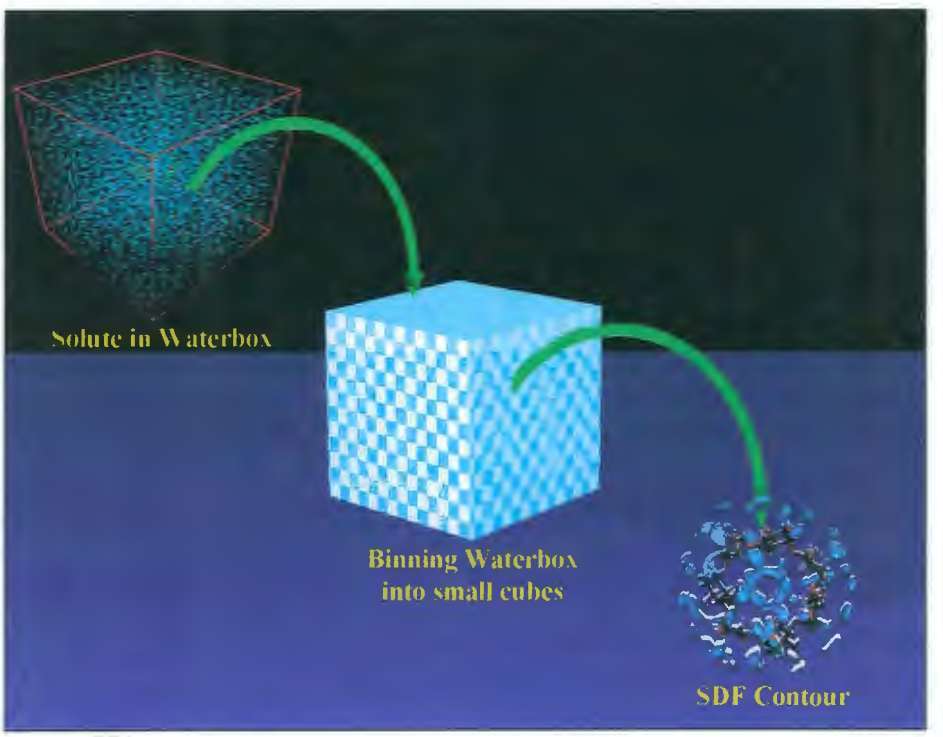


Figure 3.5 Graphical representation of solvent density mapping.

3.2.4 Voronoi Polyhedra (VP) Analysis

This analysis partitions the three-dimensional point space into convex Voronoi polyhedra (VP). This partitioning analysis provides the number of faces and vertices for each point such that all points of this region are closer to an particular atom than to any other atom of the system. Within a group of four atoms, their Voronoi polyhedra meet at one vertex describing the coordination of the nearest atomic environment. Water structure can be characterized using this statistical analysis of irregular polyhedra to obtain a specific tessellation in three-dimensional point space. An ice-like structure has 16 faces while for normal water, the number of faces ranges between 9 and 24.

Voronoi polyhedra analysis is purely a geometric approach to the problem of describing the microstructural disordering of the system. It is particularly useful in the study of liquids. The average shape of these polyhedrons indicates an important indication of the level of disordering within the system. The nature of the first hydration water shell can be evaluated for further determining if their collective structure is different from that observed for the bulk water. An accurate measure of local

water structure can be performed by calculating an asphericity parameter (η) for each of the water VPs. The asphericity parameter is defined in following equation:

$$\eta = \frac{A^3}{36\pi V^2} \qquad \text{eq 3.46}$$

where A is the total area of all the faces and V is the volume of the VP. The more spherical a VP the closer η is to 1. Thus a perfect sphere takes on a value of 1 by definition while a cube would have η=1.91. In the case of ice I_h η=2.25 for the VP of each water molecule illustrating the anisotropy of the distribution of solvent molecules in a tetrahedrally coordinated arrangement. On the other hand for liquid water at room temperature there is a Gaussian distribution of η for the solvent molecules' VPs that lies within the range 1.3 to 2.0.[20]

3.2.5 Correlation Functions

A correlation function enables one to understand the degree of correlation between the same observable of two quantities x and y. Using this concept, the time correlation functions indicate the correlation between the x and y at different times during the simulation, in other words how quickly x and y becomes uncorrelated within a dynamic trajectory. The correlation of two quantities x and y is often measured by the <x.y> for vector quantities. If the properties are uncorrelated, the correlation functions will have means of zero whereas if the properties are correlated, the function will be either positive or negative.

$$C_\phi(t) = \langle \Delta\phi(t)\Delta\phi(0) \rangle \qquad \text{eq 3.47}$$

One tend to normalize the functions by dividing the product of the mean values by x and y.[8]

3.2.5.1 Autocorrelation Function

The intramolecular fluctuations such as the rotation about the glycosidic bonds are of general interest in carbohydrate science. Experimentally, one can use NMR to determine the Lamour frequency of this phenomenon using T1 internuclear relaxation measurements. Computationally, one can employ the root mean square (RMS) fit autocorrelation function as defined in equation 4.49. The rate of fluctuation decay indicates the "uncorrelated" rate of the internal degrees of a molecule. The function RMS(t_a,t_b) is the measurement of the best RMS fit for superimposing the different conformers at time t_a and t_b.[8]

$$C_\phi(t) = \langle \mathrm{RMS}(0, t) \rangle = \langle \left(\frac{\sum_{i=1}^{N_{atoms}} d^2}{N_{atoms}} \right)^{-2} (0,t) \rangle \qquad \text{eq 3.48}$$

4.1.1 Geometrical Criteria of Hydrogen Bonding

Hydrogen bonding has long been regarded as an important type of interaction in the simulating biopolymers. When a hydrogen atom is covalently bonded to an atom (the "donor" atom, (***D***)) with high electronegativity, the high electron withdrawing tendency of that particular atom causes the hydrogen atom to be partially deshielded. Consequently this hydrogen atom is then able to interact with a region of high electron density from another electron-rich atom (the "acceptor" atom (***A***)). However, such interaction is only possible when under certain geometrical criteria. Hydrogen bonding is abundant in nature and their importance can not be overlooked specially in the studies of carbohydrates (see chapter one). From a historical perspective, In 1920, Latimer and Rodebusch published one of the very first proper documented reference which draws attention specifically to this ambiguous phenomena.[16] However, the concept of hydrogen bonding has been noticed long before and can be traced back to 1902 when Werner disclosed the formation of ammonia hydroxide from water and ammonia.[17]

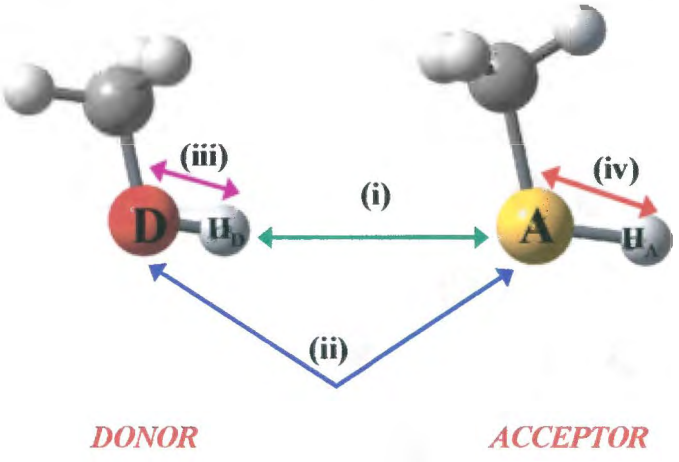
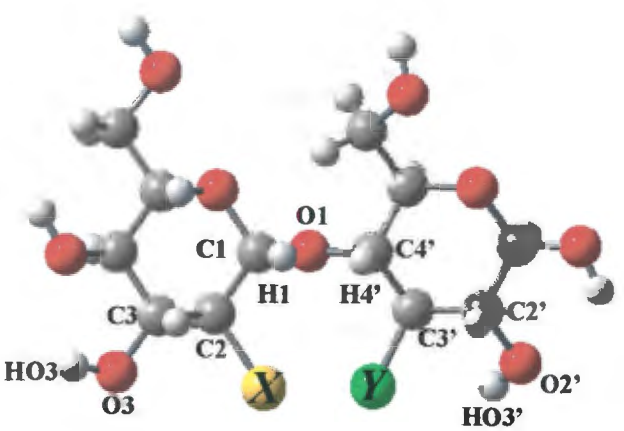


Figure 4.2 Hydrogen bonding interaction induced several geometrical alterations within a dimer complex.

Hydrogen bonding can be detected experimentally using NMR, IR and UV spectrospcopies and the observable outcomes have been summarised very adequately by Lii.[18] Geometrically, hydrogen bonding determination can be achieved by examining the following changes (illustrated in Figure 4.2) which occurs when two molecules interact via a hydrogen bond.[18]. The changes includes (i) Distance between the donor hydrogen ($\mathbf{H}_D$) and acceptor atoms (***A***) are brought together closer than the sum of their van der Waals nuclei radii. (ii) The degree of linearity between the ***D***, $\mathbf{H}_D$ and ***A*** increases, (iii) the bond length between ***D*** and $\mathbf{H}_D$ increases and (iv) so does the bond joining the acceptor hydrogen ($\mathbf{H}_A$) to the acceptor.

however, is due to a fortuitous cancellation of errors stemming from limitations in the basis set and the wavefunction models. Moreover a more serious criticism of the HF method is its neglect of electron correlation, which results in an over-estimation in bond lengths.[13] In order to accurately investigate the intramolecular hydrogen bond, perturbation methods and Density Functional Theory (DFT) therefore must be employed as they both offer electron correlation. MP2 and Becke-Three Lee-Yang-Parr (B3LYP)[14] have been shown to provide reliable trends of binding energies even though the small energy differences often observed in relative binding energies can be overshadowed by the DFT quadrature errors.[13] This weakness of DFT can be overcome by increasing the basis set size as has been recommended by Gálvez *et al.* that the B3LYP/6-311++G(d,p) is the preferred method for calculating hydrogen bond strengths.[15]

Our main objective here is to rank the hydrogen bond strength of various combinations of the following functional groups: OH, NH_2, and $NHCOCH_3$ as they are found in 1-4 linked saccharides and illustrated in Figure 4.1. We are able to achieve this objective and arrive at a ranking of intramolecular hydrogen bonds by quantifying the relationship between hydrogen bond strength and electron density within the hydrogen bonds occurring between pairs of functional groups.



$\Phi = H1 - C1 - O1 - C4'$

$\Psi = C1 - O1 - C4' - H4'$

-X	-Y
-OH	-OH
$-NH_2$	-OH
$-NH_2$	$-NH_2$
$-NHCOCH_3$	-OH
$-NHCOCH_3$	$-NH_2$
$-NHCOCH_3$	$-NHCOCH_3$
$-OCH_3$	$-OCH_3$
$-OCH_3$	-OH
$-CH_3$	$-CH_3$

Figure 4.1 **Atomic, dihedral angle and functional group labels for the disaccharides and their derivatives only the β(1→4) however the same labels applies to the α(1→4) analogues.**

4.1 Introduction

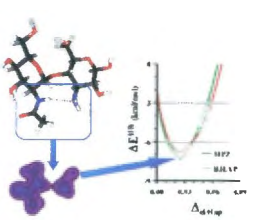
Hydrogen bonding is the cornerstone of macromolecular secondary and tertiary structure. There are two classes of hydrogen bonds that determine the conformation and intermolecular interactions of biomolecules in solution. These are the internal or *intramolecular* hydrogen bonds that often compete with the external solute to solvent *intermolecular* hydrogen bonds in a biomolecular folding environment. In the case of globular proteins 90% of the hydrogen bonds are due to the interactions between the main chain N—H and main chain C=O groups forming secondary structure (α-helices, β-sheets etc).[1] The majority of intramolecular hydrogen bonds in polysaccharides take place between small functional groups (e.g., hydroxyls, amines and acetylamides) branching off pyranose or furanose rings that are in close proximity to each other. More important are the hydrogen bonds that form across glycosidic linkages joining adjacent sugar rings along a saccharide chain. These account for the macromolecular structure of polysaccharides[2] and the structural specificity, of oligosaccharides found in glycoproteins, required for molecular recognition in cell-cell interactions.[3]

There have been some experimental[4] and *ab initio*[5-8] investigations into the nature of the hydrogen bond between functional groups of biological relevance. At the same time molecular mechanics studies often invoke hydrogen bond explanations to rationalize conformational and binding observations.[9] The overall topology of a large polysaccharide is influenced significantly by the performance the individual glycosidic dihedral angles in the polymer. Here the strength of the cross-glycosidic intramolecular hydrogen bond has a significant effect on the degree of rotational flexibility of the glycosidic dihedral angles. However within the condensed medium, this intramolecular hydrogen bond is in competition with water molecules that form hydrogen bonded bridges between residues across the linkages.[10] The introduction of a bridging intermolecular bonded water molecule that is in regular exchange with the intramolecular cross glycosidic linkage hydrogen bonds will significantly affect the conformation and properties of an oligosaccharide. One possible means of controlling the strength of a polysaccharide material would be to prevent the occurrence of such phenomenon through increasing the strength of the intramolecular hydrogen bonds across the glycosidic linkage so as to exclude the intermolecular hydrogen bonds formed with the bridging water.[11]

In general it is not possible to accurately estimate the strength of an intramolecular hydrogen bond from ab initio calculations. As a result there has to date not been a systematic attempt to rank the hydrogen bond strength across the glycosidic linkage apart from our preliminary studies.[12] In the meantime there have been much theoretical investigations into intermolecular hydrogen bonding in small organic molecules. Most studies, on medium sized molecules of the type we are interested in, employed the Hartee-Fock (HF) level of theory. Hydrogen bond strengths between pairs of functional groups such as hydroxyls and amines are determined from intermolecular binding energies. It is possible to approach the HF limit of the binding energy using a sufficiently large basis set and in some cases these energies are in agreement with experimental binding energies. This

Chapter Four

Evaluating Intramolecular Hydrogen Bond Strengths in (1→4) linked Disaccharides from Electron Density Relationships



"The theory of atoms in molecules enables one to take advantage of the single most important observation of chemistry, that of a functional group with a characteristic set of properties."

R. F. W. Bader

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

(4) MacKerell, J. A. D.; Brooks, B.; Brooks, I. C. L.; Nilsson, L.; Roux, B. *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p271.

(5) Maple, R. J. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p1015.

(6) Schlegel, H. B. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p1136.

(7) Schlick, T. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p1142.

(8) Allen, M. P.; Tildesley, D. J. *Computation Simulation in Liquids*; Clarendon Press Ltd.: Oxford, **1989**.

(9) Andersen, H. C. *J. Chem. Phys.* **1980**, *72(4)*, 2384.

(10) Thompson, D. L. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p3056.

(11) Tasaki, Y.; McDonald, S.; Brady, J. W. *J. Comput. Chem.* **1993**, *14*, 278.

(12) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.

(13) Soper, A. K. *Chem. Phys.* **1986**, *107*, 61.

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

(15) Brady, J. W. In *Encyclopedia of Computational Chemistry*, van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, **1998**, p2609.

(16) Cremer, D.; Pople, J. A. *J. Am. Chem Soc.* **1975**, *97*, 1354.

(17) Brady, J. W.; Schmidt, R. K. *J. Phys. Chem.* **1993**, *97*, 958.

(18) Kuttel, M., *MSc Thesis*, Chemistry Department; University of Cape Town, Cape Town, **1999**.

(19) Best, R., *MSc Thesis*, Chemistry Department; University of Cape Town, Cape Town, **2000**.

(20) Ruocco, G.; Sampoli, M.; Vallauri, R. *J. Chem. Phys.* **1992**, *96*, 3857.

3.2.5.2 Diffusion Coefficient

In condensed phase, the molecules within the system diffuse randomly and a modified time correlation can be used to calculate the diffusion properties. The self-diffusion coefficient can be obtained by the following equation:

$$D_s = \int_0^{+\infty} \langle \vec{v}(t) \cdot \vec{v}(0) \rangle \, dt \qquad \text{eq 3.49}$$

The vector notation v is the velocity of the particles. With a sufficiently long trajectory, one can approximate the displacement of the particle (r) using the Einstein relation:

$$D_t = \lim_{t \to \infty} \frac{\left\langle |r(0) - r(t)|^2 \right\rangle}{6t} \qquad \text{eq 3.50}$$

where the r(t) denotes the displacement of the particle.[8]

3.3 Applications and Conclusion

For the last four decades, empirical force field methods have proven to be exceedingly useful for the modelling and simulation of various phenomena. The advantages of these methods are clearly demonstrated in chapter 5, 6, 7, and 8 where a variety of complicated problems can be resolved using the EFF techniques of molecular mechanics and molecular dynamics.

CHARMM is a comprehensive platform which uses the EFF methods as its basis of calculation. A wide range of algorithms and analytical tools are available in CHARMM which makes investigations of lager variety of chemical systems possible. Numerous extensions and upgrades of CHARMM ensure the applicability and capability of this program. The simulation of biological molecules can be performed efficiently with CHARMM. However, unless a set of accurate force fields is acquired, reproducing the proper conformational and dynamisc behaviours of the condensed phase system would not be possible. The validity of the force fields parameters will be discussed and examined in chapter five.

REFERENCES

(1) Grant, G. H.; Graham Richards, W. *Computational Chemistry*, Oxford University Press: Oxford, **1995**.

(2) Leach, A. R. *Molecular Modelling: Principles and Applications*, Addison Wesley Longman Ltd: Singapore, **1996**.

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

4.1.2 Mathematical Representation of Hydrogen Bonding

Refer to Chapter 2, mathematically, the common expression for estimating hydrogen bond strengths between a hydrogen donor molecule and the molecule accepting the hydrogen is to divide the energy of the molecular pair into a contribution from the two molecules plus the overall lower interaction energy such that

$$\Delta E^{HB} = E(\boldsymbol{D}\text{-}\mathbf{H}_{\boldsymbol{D}}\cdots\boldsymbol{A}) - [E(\boldsymbol{D}\text{-}\mathbf{H}_{\boldsymbol{D}}\cdots\boldsymbol{A}) + E(\boldsymbol{A})] \qquad \text{eq 4.1}$$

Decomposing the interaction energy (ΔE^{HB}) into electrostatic and quantum mechanical components is somewhat artificial in terms of molecular orbital theory; however, this approach has been used to facilitate a chemical understanding of the make up of hydrogen bonds. This decomposition of ΔE^{HB} can be investigated using the framework of perturbation theory.[19] In perturbation theory the contribution to the self-consistent field (SCF) interaction energy (ΔE_{SCF}^{Def}) is associated with the deformation of the zeroth order perturbation wavefunctions of the hydrogen donor (Ψ_D^0) and the hydrogen acceptor (Ψ_A^0) molecules which results from the interaction between the electrons and the nuclei on the two different molecules. The interaction operator ($\hat{V}$) acts on the relaxed, fully symmetrized SCF wavefunction of the hydrogen-bonded complex ($\Psi_{D,A}$) giving the deformation energy

$$\Delta E_{SCF}^{Def} = \left\langle \Psi_{D,A} \middle| \hat{V} \middle| \Psi_{D,A} \right\rangle - (ES + EX) \qquad \text{eq 4.2}$$

which is equivalent to ΔE^{HB}. The second term in equation 4.2 is the combined electrostatic energy (E_S) and the quantum mechanical exchange energy (E_X) arising from the antisymmetry operator ($\hat{A}$) acting on the zeroth order perturbation wavefunctions i.e.,

$$\left\langle \hat{A}\Psi_D^0\Psi_A^0 \middle| \hat{V} \middle| \hat{A}\Psi_D^0\Psi_A^0 \right\rangle \qquad \text{eq 4.3}$$

The strength of the hydrogen bonds can vary considerably with the type of donor and acceptor atoms presented but commonly within the range between 3 to 12 kcal/mol amongst the functional groups found biologically.[6] There are two main contributions to the hydrogen bonding energy: covalent and electrostatic. The covalent attribution arises as a result of overlap of the orbital of hydrogen with the acceptor atom whereas the interaction between the electronegative acceptor atom and the electropositive hydrogen leads to the electrostatic attraction.[20] During the last few decades,

the hydrogen boding energy estimations were mostly based on the decomposition scheme developed by Morokuma[21]. This scheme sub-partitioned the total Hartree-Fock (HF) derived binding energy (ΔE_{SCF}^{Def}) into electrostatic, polarization, exchange repulsion, charge transfer and coupling components. Using procedures of this type it has become widely accepted that energetically moderate hydrogen bonds involved in bio molecular systems (e.g., proteins, polysaccharides, nucleic acids etc.) are principally electrostatic in nature. However, this procedure is flawed since differentiating the orbital space between the above terms is largely arbitrary[19,22] and so it is not possible to distinguish between the charge transfer and polarization terms.[23] Furthermore the lack of electron correlation at the Hartree-Fock level of theory makes the response of electrons in one part of the hydrogen-bonded system to the positions of electrons in another part less accurate.

A more direct approach to assessing the quantum mechanical nature of the hydrogen bond is to calculate directly the relation between electron density and hydrogen bond strength. The problem is to define the electron density in the volume between the ***D*** and ***A*** participants in the hydrogen bond. The overall electron density can be calculated from the wavefunction of the hydrogen-bonded complex over all N atomic coordinates.

$$\rho(\vec{r}_{D-H---A}) = \int \left|\Psi_{D,A}(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N)\right|^2 d\vec{r}_1, d\vec{r}_2, \ldots, d\vec{r}_N \qquad \text{eq 4.4}$$

An absolute electron density (ρ(r)) associated with the hydrogen bond, that is comparable for all combinations of functional groups capable of hydrogen bonding, is best calculated using Bader's Atoms in Molecules (AIM) theory).[24] The electron density may be analyzed in terms of its topology (maxima, minima and saddle points) by partitioning the molecules into atomic basins. A gradient vector map ($\nabla\rho(\hat{r})$) is generated and an interatomic surface (a surface that separates the atomic basins of neighbouring atoms) can then be analysed. The bond path is then the line through space between ***D***-H_D and ***A*** along which the electron density is a maximum while the point of lowest $\rho(r)$ on this line is the bond critical point (BCP). There are two sets of trajectories associated with the BCP, (i) the set that terminates at the critical point and defines the interatomic surface and (ii) the pair that originates at the critical point and defines the line of maximum density.[25] An accurate estimate of the strength a hydrogen bond can be determined when the electron density distribution satisfies eight topological criteria and the hydrogen atom complies with four properties. These have been derived in detail by Koch and Popelier[26] and conveniently summarized in a paper by Pacios and Gómez.[8] The major advantages of the AIM method is its ability to produce system independent robust results and that the method is not highly basis set dependent.[24] Refer to chapter 2 for more in depth details.

While intermolecular hydrogen bond energies between independent (i.e., not bonded) functional groups can be estimated by calculating the binding energy between two molecules, an accurate

measure of intramolecular hydrogen bond strengths for multifunctional molecules is not available. Estimates of the energy of intramolecular hydrogen bonds have been achieved by taking the difference in energy of the maximally oriented hydrogen bonded configuration i.e., E(***D***-$\mathbf{H}_D$···***A***) and the energy of an anti-hydrogen bonded configuration that has the ***D***-$\mathbf{H}_D$ proton-donating bond rotated away from the acceptor atom by 180° i.e., E($\mathbf{H}_D$-***D***···***A***).[27,28] The difference in energy between these two configurations do not correctly reveal the energy of the hydrogen bond (ΔE^{HB})[28] since strain effects and electronic steric repulsions are embedded in these energy differences. Moreover, procedures of this kind are not suitable for molecules with multiple hydrogen bonds found in close proximity to each other, as is the case in disaccharides. Rotating a single hydrogen bond, in the way described above, in the presence of a set of adjacent hydrogen bonds will result in an energy difference between the conformations that is artificially higher than the energy resulting from only the electrostatic and electronic contributions to hydrogen bonding. Grabowski proposed that the hydrogen bond strength may be adequately described based on either the geometrical data, or the topological parameters or the Laplacian values of the electron densities exhibited by the hydrogen bonded complex.[29] By combining these three factors a complex parameter ($\Delta_{geo+el+lap}$) was developed and shown to have a linear relationship with the hydrogen bond energy.[29] We find that it is sufficient to include only the electron density in the hydrogen bond and the Laplacian of the electron density to achieve a strong linear relationship with hydrogen bond strength. In this modified Grabowski complex parameter shown in equation 4.5 the ρ_{BCP} and $\nabla^2\rho_{BCP}$ correspond to the electron density and Laplacian of the electron density of a proton donor bond (***D***-**H**) involved in hydrogen bonding. While the ρ^0_{BCP} and $\nabla^2\rho^0_{BCP}$ correspond to the electron density and Laplacian of the electron density of bond (***X***-**H**) not involved in hydrogen bonding.

$$\Delta_{el+lap} = \frac{\left(\rho^0_{BCP} - \rho_{BCP}\right)}{\rho^0_{BCP}} + \frac{\left(\nabla^2\rho^0_{BCP} - \nabla^2\rho_{BCP}\right)}{\nabla^2\rho^0_{BCP}} \qquad \text{eq 4.5}$$

Here we develop correlation curves between hydrogen bond strengths (ΔE^{HB}) and a modified Grabowski's complex parameter (Δ_{el+lap}) from all hydrogen donor and acceptor combinations of the following functional groups: OH, NH_2, and $NHCOCH_3$) by varying distances between ***D*** and ***A***. These ΔE^{HB} vs Δ_{el+lap} correlation curves are used to estimate the ΔE^{HB} value for the cross glycosidic intramolecular hydrogen bond by calculating the Δ_{el+lap} parameter for the minima conformation of each of the 1-4 linked disaccharide derivatives (as shown in Figure 4.1).

4.2 Computational Details

Full geometry optimisation was performed under the restricted closed shell ground electronic state for all the pairs of functional groups using Gaussian 98.[30] All the geometry optimisation of the pairs of functional groups were performed using the second order Møller-Plesset perturbation theory (MP2)[31] and Becke-Three Lee-Yang-Parr (B3LYP)[14] density functional methods. Three different levels of basis sets, 6-31G(d), 6-31+G(d,p) and 6-311++G(d,p), were used in the optimisation of the functional group pairs. Due to the shallow potential curves of these systems, the keyword OPT=Tight was included to increase the convergence criteria. This ensures a better reliability of the final geometries. For each optimisation, vibrational frequency calculations were carried out to confirm that the global minimum state was reached for each pair of molecules. It was necessary to calculate all the donor-acceptor orientations to allow us to estimate the magnitude of the hydrogen bonds as observed in the disaccharides. As a result we optimised Me-NH_2(***D***)---Me-OH(***A***) and Me-NAc(***D*****(H)**)---Me-NAc(***A*****(N)**) (amide hydrogen of the first NAc group as hydrogen bond donor and amide nitrogen of the second NAc group as the hydrogen acceptor) for which a single imaginary frequency was observed. This is explicable since it was previously shown that oxygen is a superior hydrogen bond donor compared with nitrogen and so locating the global minimum state when nitrogen is the hydrogen bond donor is a near impossiblity.[32] All other functional group pairs reached the global minima on optimisation.

Binding energy evaluations were performed with a Boys-Benardi functional counterpoise scheme.[28] The standardized ΔE^{HB} vs Δ_{el+dsp} correlation curves were constructed by using the restricted closed shell electronic ground state method with 6-311++G(d,p) basis set for all functional group pairs including both Me-NH_2(***D***)---Me-OH(***A***) and Me-NAc(***D*****(H)**)---Me-NAc(***A*****(N)**). Both MP2 and B3LYP electron correlation theorems were employed and the results were directly compared. It was found that the rankings of hydrogen bonding strength were comparable when using both of these electron correlation theorems. Thus the disaccharides were only optimised with B3LYP/6-31G(d) initially due to the fact that MP2 is highly computational costly. After the probing the results of all possible arrangements of functional groups orientations, only selective disaccharides with the lowest energy conformation were optimised with the B3LYP/6-311++G(d,p) basis set to allow better accuracy. The topological properties of the electronic density were characterised using the AIM methodology[24] for all the calculations as implemented in Gaussian 98[30] and AIM2000.[34]

4.3 Results and Discussion

We have calculated the binding energies of all combinations of the following functional groups: Me-CH_3, Me-OCH_3, Me-NH_2, Me-OH and *trans* Me-N(H)-C(=O)-CH_3 (Me-N-acetylamide, Me-NAc). Alkylamides show planar symmetry as the electrons are delocalised along the amide bond therefore exhibiting partial double bond character. Two distinct species of geometric isomers can be found for any of the alkylamides. The trans *N*-alkyl isomer is highly preferred over the *cis* isomer regardless of the solvation medium.[35] Consequently only hydrogen bonding involving the *trans* Me-NAc group have been investigated here.

There have been many studies on the various hydrogen bonded complexes involving the functional groups, Me-OH, Me-NH_2, and Me-NAc.[6,7,36-39] using electronic structural approach. Early ab initio work for the methanol dimer was pioneered by Del Bene[36] followed by numerous HF calculations differing only in the level of basis set used. Dixon *et. al.* performed a HF/6-31G(d) calculation on a methanol dimer complex[37] and found a binding energy of –4.10 kcal/mol which is within the range of –3.2 to –4.1 kcal/mol limit measured from experiments.[40] These calculations did not include a correction for Basis Set Superposition Error (BSSE) and electron correlation therefore the comparison with experimental energies and geometries are predictably poor. When BSSE were corrected for and electron correlation included as in the Hagemeister *et. al* B3LYP/6-31+G(2d,p) Zero Point Energy (ZPE) level calculations[39] the resulting binding energies were comparable to the experimental values.

To date, one of the most comprehensive computational studies on moderate hydrogen bonds of biomolecular significance has been that of Kim and Freisner.[6] The objective of this study was to provide a resource for developers of molecular mechanics protein force fields. They optimised the dimer geometries with HF/6-31G** followed by MP2/cc-pVTZ(-f/d) energy evaluations. We optimized pairs of functional groups with basis sets (I) B3LYP/6-31G(d), (II) B3LYP/6-31+G(d,p) and (III) B3LYP/6-311++G(d,p) and compared the geometries and binding energies with published DFT and MP2 data[6,7,36-39,41] where possible. On the whole the B3LYP results are in agreement with the MP2 ones. The relative hydrogen bond strengths for the functional group pairs derived from B3LYP counter poise corrected binding energies are shown in Table 4.1. Of particular note is the result predicting the Me-OH(**D**)---Me-NH_2(**A**) hydrogen bond to be more than 3.5kcal/mol stronger than the Me-NH_2(**D**)---Me-OH(**A**). This is consistent with previous studies that show hydroxyls to be better hydrogen bond donors than amines.[6]

Table 4.1 **Hydrogen bonded functional pairs that have optimised and evaluated with three different methods (I) MP2/6-31G(d), (II) MP2/6-31+G(d,p) and (III) MP2/6-311++G(d,p) are ranked from strongest to weakest hydrogen bond.**

Rank	Donor	Acceptor	Geometry			Basis set I	Basis set II	Basis set III
			A ··· D Dist. (Å)	H ··· D Dist. (Å)	A-H ··· D Angle (°)	ΔE^{HB} kcal/mol	ΔE^{HB} kcal/mol	ΔE^{HB} kcal/mol
1	Me-NAc	Me-NH2	3.0584	2.0410	172.73	-7.2003	-7.0882	-6.9151
2	Me-OH	Me-NH2	2.8712	2.1784	170.02	-7.1748	-6.8332	-6.7354
3	Me-OH	Me-NAc	2.8396	1.8960	160.63	-7.0122	-7.0512	-6.6988
4	Me-NAc	Me-NAc	2.9802	1.9792	168.82	-6.7920	-6.9676	-6.4982
5	Me-NAc	Me-OH	2.9611	1.9801	161.70	-5.2892	-5.4297	-4.9932
6	Me-OH	Me-OH	2.8691	1.9033	169.41	-5.2510	-5.1481	-4.7865
7	Me-OH	Me-O CH3	2.9062	1.9343	173.37	-4.5181	-4.5018	-4.3955
8	Me-NH2	Me-NAc	3.1121	2.1416	168.09	-3.8537	-4.0158	-3.6110
9	Me-NH2	Me-NH2	3.1505	2.1945	155.06	-3.3004	-3.1837	-3.2064
10†	Me-NH2	Me-OH	3.0334	1.9240	150.39	-3.0926	-3.1039	-2.6374
11*	Me-OCH3	Me-OCH3	3.581	2.699	138.58	-0.687	-0.919	-1.169
12*	Me-CH3	Me-CH3	--	--	--	-0.010	-0.006	-0.005

* improper hydrogen bond

† conformation with 1 imaginary frequency (transitional state).

Table 4.2 **Hydrogen bonded functional pairs that have optimised and evaluated with three different methods (I) B3LYP/6-31G(d), (II) B3LYP/6-31+G(d,p) and (III) B3LYP/6-311++G(d,p) are ranked from strongest to weakest hydrogen bond.**

Rank	Donor	Acceptor	Geometry			Basis set I	Basis set II	Basis set III
			A ···· D Dist. (Å)	H ···· D Dist. (Å)	A-H ···· D Angle (°)	ΔE^{HB} kcal/mol	ΔE^{HB} kcal/mol	ΔE^{HB} kcal/mol
1	Me-OH	Me-NH2	2.8849	2.1895	171.13	-7.5737	-7.3499	-7.0751
2	Me-NAc	Me-NH2	3.0604	2.0562	175.41	-7.0447	-6.9111	-6.7348
3	Me-OH	Me-NAc	3.0195	1.8603	174.96	-6.7352	-6.7334	-6.5814
4	Me-NAc	Me-NAc	2.9905	1.9801	172.34	-5.7075	-5.6277	-5.3372
5	Me-NAc	Me-OH	2.8179	2.0012	175.96	-5.3567	-5.3350	-4.8528
6	Me-OH	Me-OH	2.8280	1.8975	168.18	-5.0323	-4.7325	-4.6709
7	Me-OH	Me-OCH3	1.887	2.858	178.58	-4.865	-5.333	-5.273
8	Me-NH2	Me-NAc	3.1051	2.0957	160.12	-3.5514	-3.7443	-3.2098
9†	Me-NH2	Me-OH	3.1058	1.8883	148.38	-3.4188	-3.2015	-2.8874
10	Me-NH2	Me-NH2	3.2256	2.2106	170.60	-3.2552	-3.1231	-2.7626
11*	Me-OCH3	Me-OCH3	3.601	2.769	132.58	-0.777	-1.119	-1.569
12*	Me-CH3	Me-CH3	--	--	--	-0.008	-0.005	-0.002

* improper hydrogen bond

† conformation with 1 imaginary frequency (transitional state).

4.3.1 Comparison Between MP2 and B3LYP

We used both MP2/6-31G(d) and B3LYP/6-31G(d) methods to optimise the dimer structures with the aim of measuring the reliability of the computationally faster, DFT method in comparison with the accepted reliability of the MP2 method for hydrogen bond analysis. The geometry and the binding energies of the optimised monomer and dimer complexes calculated here strongly correlate to the data reported by various researchers.[6,7,36-40,42] The data for the MP2 calculations is listed in Table 4.1 while the B3LYP data is shown in Table 4.2. In general the B3LYP/6-31G(d) derived energies and geometries are similar to those produce from the MP2 method with a few notable exceptions. We find that the B3LYP method consistently produces binding energies that are on average 0.3 kcal/mol lower in energy than the MP2 derived energies. The results for the Me-NAc dimer leads to the greatest disagreement between the two methods with the B3LYP calculations predicting a binding energy of more than 1.2 kcal/mol lower than that found with the MP2 calculations. None-the-less the trends in binding energies predicted by both methods are the same except for two cases. The MP2 calculations predicts Me-OH(***D***)----Me-NH_2(***A***) to be marginally stronger than Me- NAc (***D***)----Me-NH_2(***A***) with an overall relative ranking of 2 and 3 respectively while the B3LYP calculation showed a reverse ranking for these hydrogen bond pairs. Further down the ranking the MP2 procedure predicts the Me-NH_2 dimer to be approximately 0.6kcal/mol stronger than the Me-NH_2(***D***)----Me-OH(***A***) complex while the B3LYP method shows the latter hydrogen bond pair to be stronger by 0.12 kcal/mol. These differences between the two methods are small and the results disagree by less than 1 kcal/mol. On the whole the B3LYP results are in agreement with the MP2 ones. Of particular note is the common result predicting the Me-OH(***D***)---Me-NH_2(***A***) hydrogen bond to be more than 3.5kcal/mol stronger than the Me-NH_2(***D***)----Me-OH(***A***). This is consistent with previous studies that show hydroxyls to be better hydrogen bond donors than amines.[6]

4.3.2 Electron Density Hydrogen Bond Correlation

We utilised the AIM procedure to calculate the electron density of the hydrogen bonds of the pairs of functional groups listed in Tables 1 and 2. All of the hydrogen bonded pairs comply with the criterion required for hydrogen bonding developed by Koch and Popelier.[26] Estimating the strength of the hydrogen bond based on geometrical terms does not provide satisfactory answers. In Figure 4.3 (a) and (c) the binding energies of the fully optimised complexes evaluated from MP2/6-31G(d)//MP2/6-311++G(d,p) and B3LYP/6-31G(d)//B3LYP/6-311++G(d,p) respectively are plotted against the distance parameter ***D***-$\mathbf{H}_{D}$—***A***. It is clear from these plots that there is not a strong linear correlation between these two parameters which echoes the findings of Kim and Friesner.[6] Graphical plots showing binding energies against electron densities of the complexes is presented in Figure 4.3 (b) and (d). The relationship between binding energies ($\Delta E^{HB}_{MP2/DFT}$) and the

5.2 Computational Details

The program CHARMM[21] was employed for all the work involving empirical force field methods. The Gaussian98[22] suit of programs was used to carry out all electronic structural calculation.

5.2.1 Parameter Development Phase

In order to develop reliable parameter sets, various sets of atomic charge assignments were tested in CHARMM. The atomic charge evaluations were carried out quantum mechanically for each optimised conformers using the Merz-Singh-Kollman (MSK)[23] method (refer to chapter 2) and the B3LYP/6-311++G(d,p).[24] A detailed discussion on the force field development of the NH_2 functional group was included as an example while the parameterisation of other functional groups was performed in the similar manner.

5.2.1.1 Dihedral Angle Parameters

Full geometry optimisations of the monosaccharide mimics were performed under the restricted shell B3LYP level of theory along with the 6-311++G(d,p) basis sets. The same quantum mechanical method was selected to construct potential energy wells that were generated by evaluating the static molecular energy along 12 grid points associated with the 360 degrees rotation about the C2-*X* bond (refer to section 5.1.1.1). The same approach was carried out using CHARMM. Different possibilities of C2-*X* dihedrals parameters were explored in attempt to reproduce energy curves obtained by respective DFT calculations.

5.2.1.2 Charge Assignment

Full geometry optimisations were performed using various conformers of the disaccharide mimics as was mentioned previously in section 5.1.1.2. In additional to this, at each interval all possible hydrogen bonding orientations were selected as the starting molecular conformation. To reduce computational cost, the restricted closed shell ground electronic state B3LYP/6-31G(d) was selected as the method of optimisation. However, all energy evaluation were carried out with a more adequate level of theory, B3LYP/6-311++G(d,p). The energy profiles were constructed by choosing the individual conformer with the most stable molecular energy. The same procedure was carried out in CHARMM for evaluating the change in disaccharide energy with respect to the change of their ϕ,ψ dihedral angles. Different atomic charge parameters were explored in attempt to reproduce energy curves obtained by the respective DFT calculations.

formation of an inter-residue hydrogen bond, the functional groups interact in a more random way. However, these types of conformers are generally located on the higher energy region of the map. In order to successfully guarantee each conformer represents a true local minimum of the adiabatic map, simulated annealing methods were performed in the manner described by Naidoo, *et al.*[19]

5.1.2.2 Vacuum Dynamics

The wells in the adiabatic map indicate the conformations that are expected most frequently in the vacuum simulation. Even though condensed phase dynamics studies include solution entropy effects which alter the favour vacuum dynamics sampled space, but the dynamics sampling region would not radically deviate from this wells. One could perhaps predict regions where the condense phase simulation are likely to sample by examining the scattering of the dihedral angles of the disaccharides during the dynamics.[20]

The fundamental protocol of this chapter is summarised in Figure 5.2 (i.e. the approach used for parameterising the force field and for predicting of the conformation behaviours of the disaccharides).

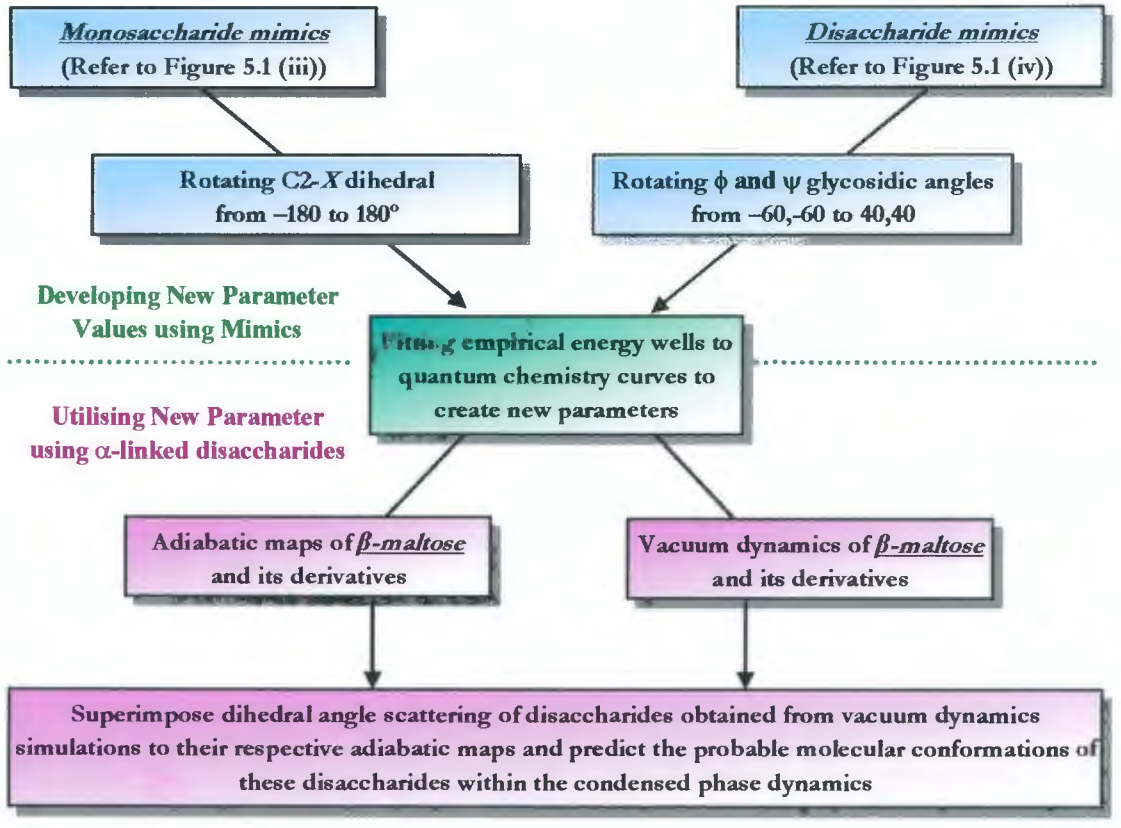


Figure 5.2 **Outline of the force field development procedures and techniques used in conformational analyses.**

5.1.2 Conformational Analyses

Disaccharides are the smallest units possessing an inter-residue glycosidic linkage are the building blocks of the larger oligo-.and polysaccharides. β-maltose (α-D-glucopyrnosyl-(1→4)-β-glucopynose) is perhaps the most studied α(1→4) linked disaccharide both experimentally and theoretically.[17] Studies of maltose have often been employed as the groundwork towards understanding larger polysaccharides. In order to achieve reliable predictions in the conformational analyses of disaccharides, incorporating all the exocylic hydroxyl groups found on the sugar ring is necessary. The reason is these hydroxyl groups often participate in intra-residue hydrogen bonding and thus can strongly influence the degree of the inter-residue hydrogen bondings. Instead of using saccharide mimics as previously used in the parameter development phase, a β-maltose with modified at the C2 and C3' positions of the non-reducing and reducing sugar units respectively was employed.

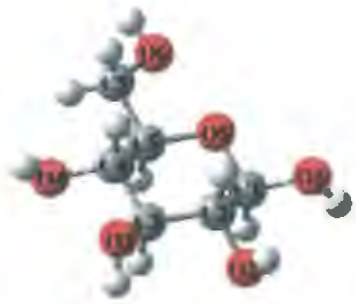
The conformation behaviour of all the disaccharide derivatives was examined. Only the simulations of 2-NH_2-Glc-α-(1→4)-Glc and GlcNAc-α-(1→4)-3-NH_2-Glc are included in this section for comparison with the conformational behaviour of β-maltose. As was shown in the previous chapter, a ranking of the relative hydrogen bond strength across the glycosidic linkage is now possible. The cross glycosidic link hydrogen bonds OH(***D***)----NH_2(***A***) and NAc(***D***)----NH_2(***A***) are known to be more than 3.5 kcal/mol stronger than OH(***D***)----OH(***A***) hydrogen bonds for the α(1→4) linked disaccharides. One of the most useful aspects of using empirical force field calculations is that they allow rapid evaluation of molecular energies. This makes them suitable for problems requiring the understanding of the relationship between changes in molecular conformation and variation in the molecular energy. In the case of disaccharide conformational prediction, one can calculate the complete energy surface results from changing one or more dihedral angles, known as adiabatic map, and subsequently verify the molecule conformations that correspond to the low energy regions.

5.1.2.1 Adiabatic Maps

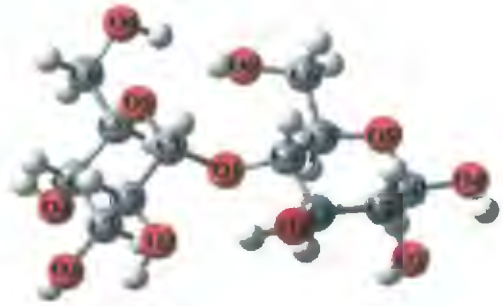
Systematic rotating about the glycosidic C-O-C bonds yields an energy hypersurface, also known as the adiabatic map. Owing to the pervasiveness of hydrogen bonding found in carbohydrates, the energy hypersurface can be critically influenced by the different orientations of any of the functional groups. Each pyranose ring of the carbohydrate is practically rigid since only the $^4C_1$ chair conformation is observed experimentally.[18] The inter-residue non-bonded interactions influence the topology a particular disaccharide adiabatic map. These interactions include steric clashes and hydrogen bonding. The steric interferences between the hydroxyl groups and the glycosidic linkage primarily lead to the high energy regions in the map, whereas the strength of the hydrogen bonding stabilises the disaccharides.

Normally the hydrogen bonding pattern of the disaccharide adapted either the "clockwise" or "anti-clockwise" arrangement. In some cases when the orientations of the glycosidic linkage forbid the

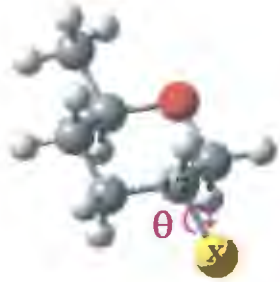
interfere with the energy profiles of all other disaccharide mimic possessing that particular functional group. For example, when changing the atomic charge of the NH_2 group, the energy profiles of disaccharide mimics involving NH_2 ---- OH, NH_2 ---- NH_2 and NAc ---- NH_2 would all be altered. On the other hand, when adjusting the atomic charges of NAc group, the energy profiles of disaccharide mimics involving NAc ---- OH, NAc ---- NH_2 and NAc ---- NAc would be altered. Therefore it is necessary to consider the outcome of each and every potential energy well when adjusting the charge parameter of the atom.



(i)

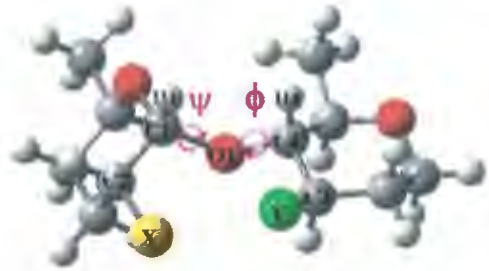


(ii)



-*X* = -NH_2
-NAc
-CH_3
-OCH_3

(iii)



-*X*:-*Y* =

-NH_2	:	-NH_2
-NH_2	:	-OH
-NAc	:	-OH
-NAc	:	-NH_2
-NAc	:	-NAc
-OCH_3	:	-OCH_3
-CH_3	:	-CH_3

(iv)

Figure 5.1 **(i) α-Glucose, (ii) β-maltose, (iii) monosaccharide mimic and (iv) α(1→4) linked disaccharide mimic ("*θ*" represent the systematic rotation for parameterising the C2-*X* dihedral. "ϕ" and "ψ" represent the simultaneous rotation employed for ensuring the charge parameter of the -*X* and -*Y* group are acceptable).**

5.1.1.1 Dihedral Angle Parameters

The development of the dihedral parameters was carried out using 2-hydroxy-5-methyl-pyranose as a monosaccharide mimic (Figures 5.1(i) and (iii) respectively). The site marked *X* (Figure 5.1(iii)), where the O2 hydroxyl group is situated (compare to Figure 5.1(i)), was later substituted with all the above mentioned small functional groups. The topologies of the potential energy wells ware obtained by systematically rotating the C2-*X* dihedral angles (θ) of interest from –180 to 180 in 30 degree increments and using the corresponding DFT and molecular mechanics to evaluate the potential energy at each grid point.

5.1.1.2 Charge Assignments

Momany and Willet performed B3LYP optimisations on maltose with varying ϕ,ψ dihedral angles (Figure 5.1(ii)) and found close similarities between the lowest energy computed structure and crystal structures obtained from the neutron and x-ray diffraction studies.[9a] The objective of this current procedure is not to locate global minimum energy structures for disaccharides, but rather to make sure the charge assignments are adequate. The disaccharide mimics (Figure 5.1(iv)) is employed in this case as they can reduce the complexity of the actual disaccharide but without sacrificing essential conformational features of the sugar complex. This approach has been previously used by Cloran *et al.*[16] but differs from our approach since we include the primary methyl (C6) groups in the disaccharide mimics. This is an important feature for carbohydrate mimics since omitting the primary methyl groups results in inter-atomic distances between C2 to C3' that being longer than is experimentally observed. Here, the functional group were substitute on the *X* and *Y* sites (Figure 5.1(iv)) of the maltose mimics according to the combination scheme previously described in chapter 4.

The dominant feature of an $\alpha(1\rightarrow4)$ linked disaccharide adiabatic map is a broad low energy region situated diagonally across the center of the map starting from (ϕ,ψ) = (-60,-60) to (40,40). Thus, we carried out our energy evaluation in this region by systematically rotating both ϕ,ψ dihedral angles simultaneously with 10 degree increments. (These are the selected grid points: (-60,-60), (-50, -50), (-40,-40), (-30, -30), (-20, -20), (-10, -10), (0, 0), (-10, -10), (-20, -20), (-30, -30) and (40,40)). This results a profile that stretched diagonally across the low energy region of the well of the adiabatic map and such profile should be sufficient to provide an adequate description of glycosidic rotational vs energy topology of the disaccharides.

The process of finding a suitable atomic charge assignment is difficult. It is a tricky process to generating a set of atomic charges which can bring about a good agreement between the empirical energy curves and the quantum mechanical potential energy wells. Further, each atomic charge

using molecular orbital theories to collect a set of potential energy values from an array of conformations with different distorted geometries. Fitting the empirical potential energy function and its derivatives to such energy surfaces can provide more information compared with only using the information obtained from the equilibrium structure. Reiling *et al.*,[12] introduced torional terms by studying various small fragments resembling the saccharides. However their findings rely mostly on lower level quantum chemistry calculations. Tvaroška and Carver had reported the hydroxyl and -*O*-methyl torsional terms by incorporating the gauche effect while using a higher level of *ab initio* theory.[13] Their results seem more valid. However, the high potential energy barriers of their investigation made it somehow impractical for studying carbohydrate conformational properties. Here, we make use of the density functional theory (DFT) calculations for the force field development considering DFT incorporates electron correlation effects. Nonetheless, the main aim of this chapter is to briefly illustrate the procedures employed for extending the existing CSFF which has previously been found to yield dynamics simulation results that are comparable to the experimental findings.[14] The force field parameters of these small functional groups, amine (NH_2), *N*-methylacetamide ($NHCOCH_3$; NAc), permethyl (OCH_3) and methyl (CH_3), were specially developed by utilising the procedures stated in section 5.1.1. This newly developed parameters sets were employed for rapid prediction on the conformational behaviour of maltose, 2-amine-Glc-α-(1→4)-Glc and GlcNAc-α-(1→4)-3-amine-Glc. The procedures are documented in section 5.1.2. Maltose is the natural occurring α-(1→4) linked disaccharide whereas the latter two molecules are chosen due to the higher strength of their inter-glycosidic interactions (established in chapter 4). Furthermore, the force field parameters of these small functional groups were employed for investigating the condensed phase conformational properties of the relevant carbohydrates. This can be examined in chapter 6, 7 and 8.

5.1.1 Parameter Development

An accurate description of the structure energy relationship is considered crucial for the development of any carbohydrates force field. In chapter one, it was mentioned that an inadequate force field set can yield misleading results. The force field parameters must be able to reliably estimate the finer details of the adiabatic map. Here any subtle error in this estimation can lead a large degree of unreliability when predicting the overall chemical-physical properties of the larger α(1→4) linked polysaccharides.[15] When beginning the parameter development, it is usually useful to choose the bond length and bond angle parameters from the similar groups found in other CHARMM carbohydrate force field or experimental data. However, the dihedral angle parameters and the charge assignments for the atoms involved should be carefully assigned. The reason is that the accuracy of the EFF calculation relies to a large extent on the rotational energy barriers which are greatly influenced by these two terms.

5.1 Introduction

Carbohydrates possessing modified functional groups are commonly found in nature. These modified carbohydrates have significant commercial value as they are often the basis of several important industrial materials and may be vital pharmaceutical ingredients.[1] As was mentioned in the previous chapter, every functional group has a different hydrophilicity/hydrophobicity. Hence, each functional group can influence the inter-residue hydrogen bonding strength that consequently affects the overall chemo-physical properties of the complex carbohydrates.[2] The development of computational methods for simulating such modified carbohydrates within the condensed phase has earned considerable awareness in recent years. Empirical Force Field (EFF) methods such as molecular mechanics and molecular dynamics are gaining wide popularity as the preferred techniques for studying structural the details of carbohydrates. The usefulness of EFF methods for predicting the molecular properties of the modified carbohydrates has clearly been established. However, such predictions are not possible unless a set of force field parameters is satisfactory validated.

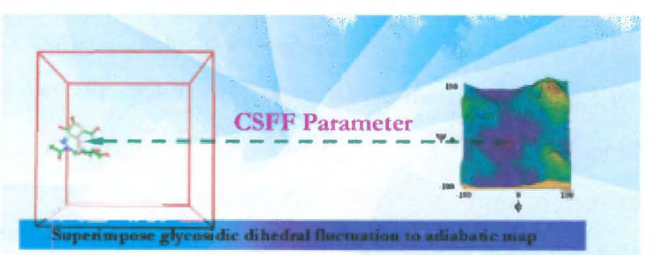
Various different types of Force Fields have been developed in order to describe carbohydrate conformational changes without a solvent environment. Several unique features of carbohydrates such as their highly polar functionalities and the dependence of their conformational behaviour on stereoelectronic effects (*viz.* anomeric, exo-anomeric and gauche effects) have long been recognized and included in the development of various parameterisation schemes.[3] In 1988, I la and co-workers presented one of the early CHARMM-Type carbohydrate parameters. Although this force field does not correctly model the experimental findings of anomer mutarotations, it yields free energies which are consistent with the experimental findings.[4]

During force field development, it is often necessary to employ experimental findings, such as Nuclear Overhuser Effects data, to validate the structural geometries of the predicted conformers. The intricate conformational changes of a carbohydrate molecule in the condensed phase would only quantitatively fit experimental data if the torsional parameters are adequately parameterised. Palma[5] has revised the Brady force field[6] and incorporated some of the Homans[7] parameters in order to produce a primary alcohol dihedral rotational that favours the gg and gt conformations. Based on Homans parameters as well as the charge scheme proposed by Wood *et al.*,[8] Momany and Willet have put forward a force field parameter set which allows an increased in rotation about the C5-C6 bond.[9] Recently, Kuttel *et al.*[10] studied the dihedral angle parameters of both the primary alcohol and the secondary hydroxyl groups by modifying the sugar parameter sets of Palma *et al.* to yield the Carbohydrate Solution Force Field (CSFF). Their CSFF exhibits primary alcohol rotational transitions on a nanosecond time scale for simulating molecular dynamics of sugars in solvent and demonstrated a excellent agreement with NMR data. The studies presented by Corzana *et al.* also confirmed the CSFF to be suitable for simulating carbohydrates in nanosecond scale.[11]

Deriving torsional parameters using the electronic structural approaches (ESA) has been widely accepted as an option when experimental data are scarce. One can generate an energy surface by

Chapter Five

Force Field Parameterisation and Conformational Analyses of α(1→4) Linked Disaccharide Derivatives



C.; Adamo, C.; Clifford, S.; Ochterski, J.; Peterssin, G. A.; Al-Laham, M. A.; Zakrzewski, V. G.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Goresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Baboul, A. G.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Gill, P. M. W.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Johnson, B. G.; Stefanov, B. B.; Gomperts, R.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A.; Gaussian, Inc.: Pittsburgh PA, **1998**.

(31) Møller, C.; Plesset, M. *S. Phys. Rev.* **1934**, 46, 618

(32) Rablen, P. R.; Lockman, J. W.; L., J. W. *J. Phys. Chem. A.* **1998**, *102*, 3782.

(33) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(34) Biegler-König, F.; Schönbohm, J.; Bayles, D. *J. Comput. Chem* **2001**, *22*, 545.

(35) Martínez, A. G.; Vilar, E. T.; Fraile, A. G.; Martínez-Ruiz, P. *J. Phys. Chem. A.* **2002**, *106*, 4942.

(36) Del Bene, J. *J. Chem. Phys.* **1971**, *55*, 4633.

(37) Dixon, J. R.; George, W. O.; Hossain, M. F.; R., L.; Price, J. M. *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 3611.

(38) (a) Karpen, A.; Schuster, P. *Can. J. Chem.* **1985**, *63*, 809. (b) Bakó, I. P., G *J. Mol. Struct. (Theochem)* **2002**, *594*, 197. (c) González, L.; Mó, O.; Yánez, M. *J. Chem. Phys.* **1998**, *109*, 139. (d) Masella, M.; Flament, J. P. *J. Chem. Phys.* **1998**, *108*, 7141. (e) Sum, a. K.; Sandler, S. I. *J. Phys. Chem. A* **2000**, *104*, 1121. (f) Cabalerio-Lago, E. M.; Rodríguez-Otero, J. *Theochem* **2002**, *586*, 225 (g) Kang, Y. K. *J. Phys. Chem. B* **2000**, *104*, 8321 (h) Torii, H.; Tasumi, T.; Kanazawa, T.; Tasumi, M. *J. Phys. Chem. B* **1998**, *102*, 309.

(39) Hagemeister, F. C.; Gruenloh, C. J.; Zwier, T. S. *J. Phys. Chem. A.* **1998**, *102*, 82.

(40) Bizzarri, A.; Stolte, S.; Reusse, J.; Rijdt, J. G. C. M. V. D.-V. D.; Duijneveldt, F. B. V. *Chem. Phys.* **1990**, *143*, 423.

(41) Tatamitani, Y.; Liu, B.; Shimada, J.; Ogata, T.; Ottaviani, P.; Maris, A.; Caminati, W.; Alonso, J. L. *J. Am. Chem. Soc.* **2002**, *124*, 2739.

(42) Cheeseman, J. R.; Carroll, M. T.; Bader, R. F. W. *Chem. Phys. Lett.* **1998**, *143*, 450.

(43) (a) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc* **1983**, *105*, 5761. (b) Taylor, B.; Kennard, O.; Versichel, W. *Acta cryst* **1984**, *B40*, 2808. (c) Murray-Rust, P.; Glusker, J. P. *J. Am. Chem. Soc* **1984**, *106*, 1018.

(44) Adalsteinsson, H.; Maultiz, A. H.; Bruice, T. C. *J. Am. Chem. Soc.* **1996**, *118*, 7689.

(45) (a) Coulson, C. A.; Danielson, U. *Ark. Fys.* **1954**, *8*, 239. (b) Coulson, C. A.; Danielson, U. *Ark. Fys.* **1954**, *8*, 245. (c) Tsubomura, H. *Bull. Chem. Soc. Japan* **1954**, *27*, 445.

(46) Best, R. E.; Jackson, G. E.; Naidoo, K. J. *J. Phys. Chem. B.* **2001**, *105*, 4742.

(47) Klein, R. A. *J. Comput. Chem.* **2002**, *23*, 585.

(5) (a) Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer, H. F. I.; Radom, L. *J. Chem. Phys.* **1990**, *92*, 1240. (b) Gordon, M. S.; Jensen, J. H. *Acc. Chem. Res.* **1996**, *29*, 536.

(6) Kim, K.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 12952.

(7) Vargas, R.; Garza, J.; Friesner, R. A.; Stern, H.; Hay, B. P.; Dixon, D. A. *J. Phys. Chem. A* **2001**, *105*, 4963.

(8) Pacios, L. F.; Gómez, P. C. *J. Comput. Chem* **2001**, *22*, 702.

(9) (a) Brady, J. W.; Schmidt, R. K. *J. Phys. Chem.* **1993**, *97*, 958. (b) Naidoo, K. J.; Denysyk, D.; Brady, J. W. *Protein Eng.* **1997**, *10*, 1249. (c) Clarke, C.; Woods, R. J.; Gluska, J.; Cooper, A.; Nutley, M. A.; Boons, G. *J. Am. Chem. Soc.* **2001**, *123*, 12238.

(10) Naidoo, K. J.; Brady, J. W. *J. Am. Chem. Soc.* **1999**, *121*, 2244.

(11) Naidoo, K. J.; Kuttel, M. M. *J. Comput. Chem.* **2001**, *22*, 445.

(12) Naidoo, K. J.; Chen, Y. J. In *Abstr. Pap. - Am. Chem. Soc*; Witczak, Z. J., Ed.; ACS: Chicago (USA), **2001**; *222*, pCARB-119.

(13) Del Bene, J. E., In *Encyclopedia of computational chemistry*; 1st ed.; John Wiley & Sons, 1998; Vol. 2.

(14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.

(15) Gálvez, O.; GóMez, P. C.; Pacios, L., F *J. Chem. Phys.* **2001**, *115*, 11166.

(16) Latimer, N. M.; Rodebush, W. H. J Am Chem Soc 1964, 42, 1419.

(17) Werner, A. Leibigs Ann 1902, 322(261), 293.

(18) Lii, H. J. In *Encyclopedia of Computational Chemistry*; van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, **1998**, p1271.

(19) Jerziorski, B.; Kolos, W. In *Molecular Interactions*; Ratajczak, H., Orville-Thomas, W. J., Eds.; Wiley: New York, 1982; Vol. 2, p1.

(20) Del Bene, J. E. In *Encyclopedia of Computational Chemistry*; van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, 1998, p1263.

(21) Morokuma, K. *Acc. Chem. Res.* **1977**, *10*, 294.

(22) Scheiner, S. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH publishers, Inc: New York, 1991; Vol. 2, p165.

(23) Morokuma, K.; Kitaura, K. In *Chemical Applications of Atomic and Molecular Electrostatic Potentials*; Politzer, P., Truhlar, D. G., Eds.; Plenum: New York, 1981, p215.

(24) Bader, R. F. W. *Atoms in Molecules- A Quantum Theory*; Oxford Univ. Press: Oxford, 1990.

(25) Popelier, P. L. A. *J. Phys. Chem. A* **1999**, *103*, 2883.

(26) Koch, U.; Popelier, P. L. A. *J. Phys. Chem.* **1995**, 9747.

(27) Lipkowski, P.; Koll, A.; Karpfen, A.; Wolschann, P. *Chem. Phys. Lett.* **2002**, *360*, 256.

(28) Wojtulewskia, S.; Grabowski, S. J. *J. Mol. Struct. (Theochem)* **2003**, *645*, 287.

(29) Grabowski, S. J. *J. Phys. Chem. A* **2001**, *105*, 10739.

(30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli,

conformational energy of biopolymers have previously not been accessible as that energetic contribution had been buried in a large mix of electrostatic, van der Waals and quantum terms. Only a rough approximation could be obtained from the binding energy of model compounds. This however, did not account for the very different conformational, structural and chemical environment found when the functional group pairs were connected in a bonded arrangement.

We have found that intramolecular hydrogen bond strengths can be estimated by establishing a relationship between hydrogen bond energy and the topological properties of electron density between the complex. This relationship allows for the construction of correlation ΔE^{HB} vs Δ_{el+lap} curves for each combination of hydrogen bond pairs. Once the ΔE^{HB} vs Δ_{el+lap} curve is constructed for a pair of hydrogen bonded functional groups it can be used to estimate the intramolecular hydrogen bonds for the same pair in any system. We have used these curves to estimate the cross glycosidic linkage hydrogen bond strength in (1→4) linked disaccharide derivatives and found the relative strengths of these intramolecular hydrogen bonds within their biomolecular conformational context. Thereby demonstrating that a ranking of the relative intramolecular hydrogen bond strength occurring across the (1→4) glycosidic linkage is now possible. We have seen that the strongest intermolecular hydrogen bonds (e.g., Me-NH_2(**A**)---Me-OH(**D**)) do not result in the strongest an intermolecular hydrogen bond cross-glycosidic intramolecular hydrogen bonds when the either of the hydroxyl groups at C(2) and C(3)' are substituted with an amine. This is because the hydrogen bonds between vicinal groups on the reducing and non-reducing sugars affect the formation of the cross glycosidic hydrogen bonds. None the less our estimations of intramolecular cross glycosidic linkage hydrogen bonds showed that the NAc(**D**)----NH_2(**A**) hydrogen bond in 2-amine-Glc-α-(1→4)-Glc and 2-amine-Glc-β-(1→4)-Glc is stronger than OH(**D**)----OH(**A**) found in maltose and cellobiose.

REFERENCES

(1) Baker, E. L.; Hubbard, R. E. *Prog. Biophys. Molec. Biol.* **1984**, *44*, 97.

(2) (a) Rees, D. A. *Polysaccharide Shapes*; Chapman and Hall Ltd.: NY, **1977** (b) Atkins, E. D. T. *Polysaccharides: Topics in structure and morphology*; The Macmillan Press ltd: NY, **1985** (c) Atkins, E. D. T. In *Xylans and Xylanases*; Visser, J., Ed.; Elsevier: London, **1992,** Vol. 7, p39.

(3) (a) Johnson, L. N.; Cheetham, J.; McLaughlin, P. J.; Acharya, K. R. In *Current topics in microbiology and immunology*; Clarke, A. E., Wilson, I. A., Eds.: NY, **1988,** *139*, p81. (b) Sharon, N.; Lis, H. *Scientific American* **1993**, 82. (c) Homans, S. W. In *Molecular glycobiology*; Fakuda, M., Hindsgaul, O., Eds.: NY, **1994**, p231.

(4) (a) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293. (b) Gemmecker, G. *Angew. Chem. Int. Ed.* **2000**, *39*, 1224. (c) Landersjo, C.; Hoog, C.; Maliniak, A.; Widmalm, G. *J. Phys. Chem. B* **2000**, *104*, 5618. (d) Hawley, J.; Bampos, N.; Aboitiz, N.; Jimenez-Barbero, J.; Lopez De la Paz, M.; Sanders, J. K. M.; Carmona, P.; Vicent, C. *Eur. J. Org. Chem.* **2002**, *12*, 1925.

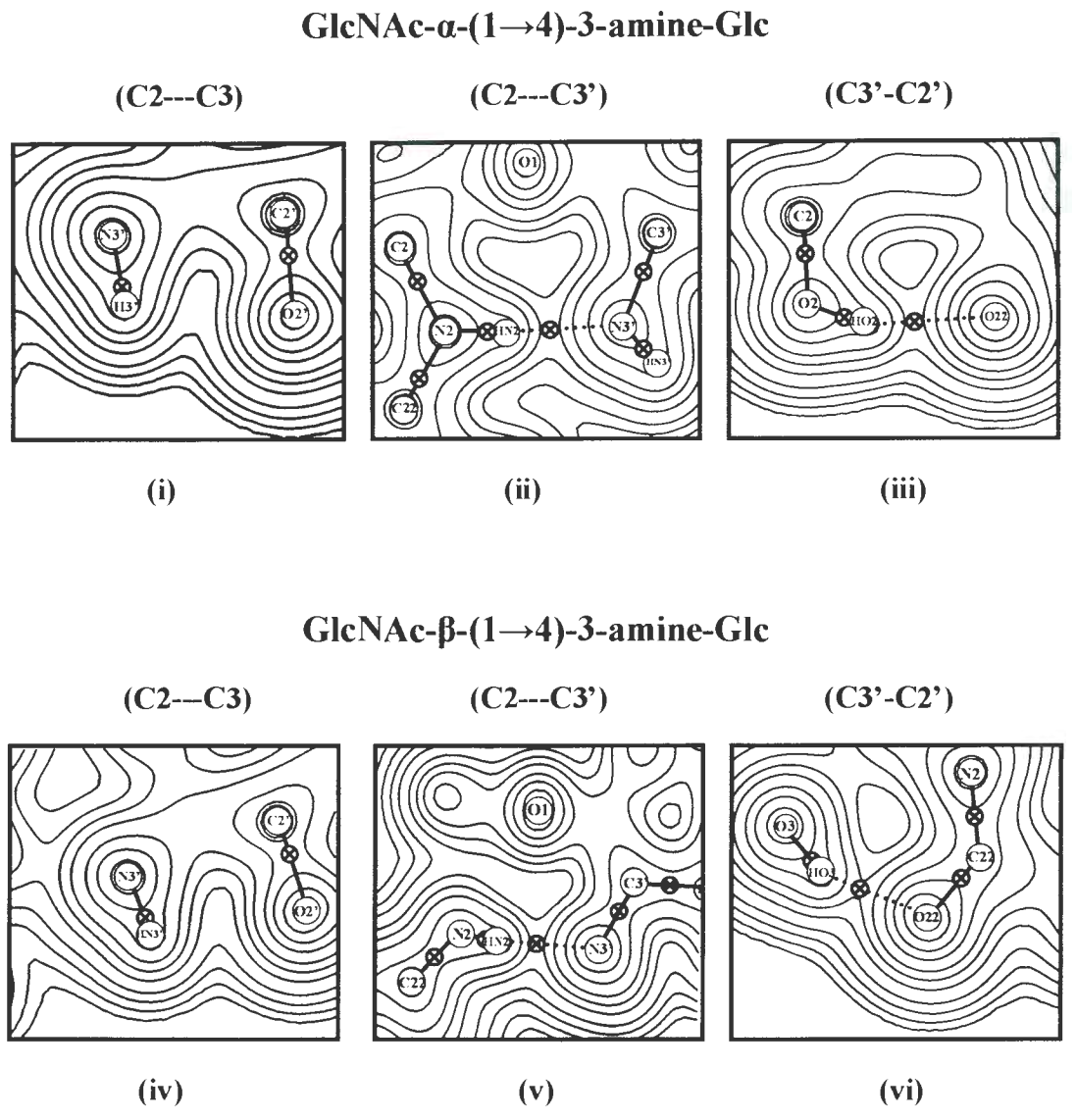


Figure 4.7 The electron density maps of the molecular fragments close to and about the glycosidic linkage of GlcNAc-α-(1→4)-3-amine-Glc *(i)-(iii)* and GlcNAc-β-(1→4)-3-amine-Glc *(iv)-(vi)* The positions of the nuclei (as labelled in Figure 4.1) are marked with an open circle and the positions of the BCPs are marked with a "⊗". Note: O22, N2 and HN2 represent the oxygen, nitrogen and amide hydrogen of the NAc substituent where as N3' and HN3' represent the nitrogen and the amine hydrogen of the amine substituent.

4.4 Conclusion

Evaluating relative intramolecular hydrogen bond strengths in folded saccharides and other biopolymers is an important component in the rational design of conformationally dependent "lock and key" binding. The contribution of intramolecular hydrogen bonds to the

Table 4.3 Geometry parameters and the estimation of cross-glycosidic H-bond strength (ΔE^{HB}) of the B3LYP/6-311++G(d,p) optimised disaccharides.

Rank	-*X*	-*Y*	Estimated ΔE^{HB}	ϕ	Ψ	H---A Distance (Å)		
						HO3---H_X / O3---X	H_X---Y / X---H_Y	H_Y---O2' / Y---HO2'
α(1→4) linked								
1	-NAc	$-NH_2$	-6.03	-14.08	1.55	1.845	1.957	2.475
2	$-OCH_3$	-OH	-5.01	-8.59	3.00	2.469	1.964	2.287
3	-OH	-OH	-4.99	-6.80	7.11	2.358	1.922	2.360
4	-NAc	-OH	-4.76	-13.80	2.77	1.856	1.945	2.286
5	$-NH_2$	-OH	-2.51	-9.79	6.53	2.220	2.127	2.308
6	$-OCH_3$	$-OCH_3$	-1.01	-12.59	2.20	2.302	2.669	2.891
7	-NAc	-NAc	-0.97	-24.86	-38.86	1.941	2.660	2.001
8	$-NH_2$	$-NH_2$	---	-24.91	-12.00	2.246	---	2.267
9	$-CH_3$	$-CH_3$	---	-25.53	-10.39	---	---	---
β(1→4) linked								
1	-NAc	$-NH_2$	-6.08	176.27	7.83	1.760	1.963	2.487
2	-OH	-OH	-5.02	180.93	0.79	2.366	1.889	2.307
3	$-OCH_3$	-OH	-4.89	187.21	7.83	2.366	1.893	2.603
4	-NAc	-OH	-4.62	176.66	8.17	1.767	1.945	2.262
5	$-NH_2$	-OH	-2.47	176.18	2.87	2.785	2.081	2.274
6	$-OCH_3$	$-OCH_3$	-0.98	160.20	4.13	2.451	N/A	3.126
7	$-NH_2$	$-NH_2$	---	188.07	12.40	2.404	2.210	2.195
8	-NAc	-NAc	---	166.76	6.01	1.743	---	2.051
9	$-CH_3$	$-CH_3$	---	162.17	9.62	---	---	---

Inter-glycosidic hydrogen bonds between the hydroxyl groups on C2 and C3'are present in both maltose and cellobiose. Estimating the hydrogen bond strength using the ΔE^{HB} vs Δ_{el+kp} correlation curves gives a value of approximately 5kcal/mol. Interestingly, the intramolecular interaction between the vicinal hydroxyl groups located on the C3 and C2 as well as the C3' and C2' do not satisfy the AIMs criterion for hydrogen bonding. This result confirms the previous observation for 1-2-diols[47] since there is no electron density between these adjacent hydroxyl functional groups.

Ranking the intramolecular hydrogen bond strengths of the (1→4) linked disaccharides derivatives with functional groups substituted at the C2 (X) and C3' (Y) positions (Table 4.3) produced a different trend than that of the corresponding intermolecular hydrogen bonds listed Table 4.2. The strongest intermolecular hydrogen bond is Me-NH_2(**A**)---Me-OH(**D**) which is not the case with this functional group pair substituted about the glycosidic α and β(1→4) linkage. Rather, the functional groups exhibiting the strongest hydrogen bond across both α and β(1→4) glycosidic linkages are the NAc(**D**)----NH_2(**A**) combination. The reason for the differences in the two trends lies with the formation of intramolecular hydrogen bonds between the adjacent hydroxyls and the NH_2 and NAc functional groups on the reducing and non-reducing sugars. An examination of the electron density maps (Figure 4.7) of GlcNAc-α-(1→4)-3-amine-Glc and GlcNAc-β-(1→4)-3-amine-Glc showing fragments of the disaccharides supports the existence of hydrogen bonding between these vicinal groups as seen in the molecular graphs of Figures 4.5(v) and Figure 4.6(v). Unlike the case of the adjacent hydroxyl groups on the reducing and non-reducing sugars of maltose and cellobiose that do not exhibit hydrogen bonding a BCP does occur between the hydroxyl and the NAc of the non-reducing sugar and a further BCP occurs between the hydroxyl and the amine on the reducing sugar both with the correct topology for hydrogen bonding. These hydrogen bonds between the vicinal groups on the two sugar rings optimally direct the NAc toward the amine inducing a strong hydrogen bond of approximately 6kcal/mol for both α and β derivatives compared to approximately 5kcal/mol for the equivalent cross glycosidic hydrogen bond found in maltose and cellobiose.

While the Me-NAc---Me-NAc dimer complex produced a relatively strong intermolecular hydrogen bond when both *X* and *Y* sugar positions were substituted with NAc groups significant steric repulsions between these bulky groups occurs. Consequently no BCP was found between substituents in GlcNAc-α(1→4)-GlcNAc resulting in only a weak cross-glycosidic electrostatic attraction. The GlcNAc-β(1→4)-GlcNAc disaccharide produced a very weak hydrogen bond between the amide hydrogen of the non-reducing sugar and the amide nitrogen of the reducing sugar.

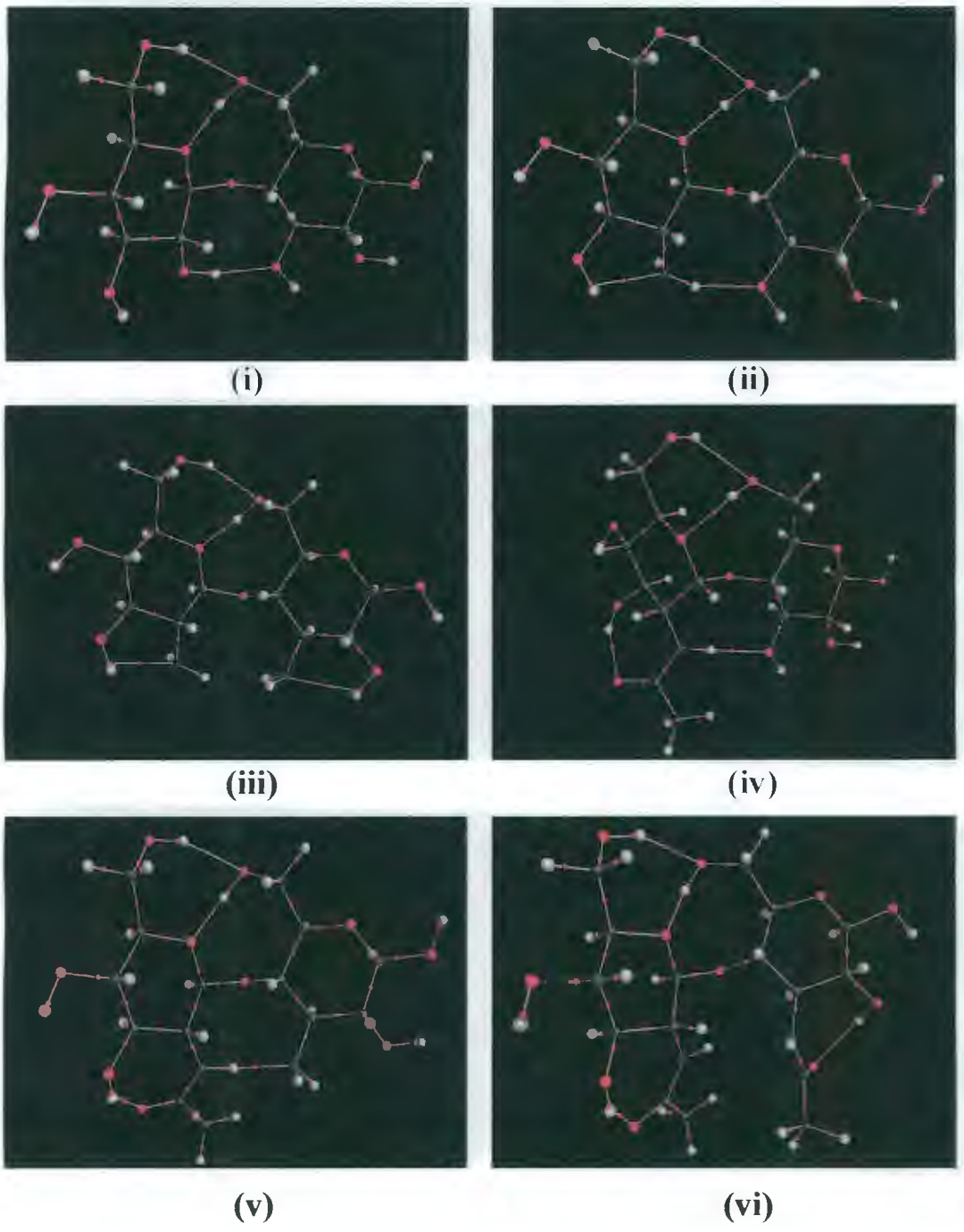


Figure 4.6 **The molecular graphs of the B3LPY/6-311++G(d,p) optimised β(1→4) disaccharides where the small red spheres indicate the positions of the BCPs. The disaccharides shown are (i) cellobiose, (ii) 2-amine-Glc-β-(1→4)-Glc, (iii) 2-amine-Glc-β-(1→4)-3-amine-Glc, (iv) GlcNAc-β-(1→4)-Glc, (v) GlcNAc-β-(1→4)-3-amine-Glc and (vi) GlcNAc-β-(1→4)-GlcNAc.**

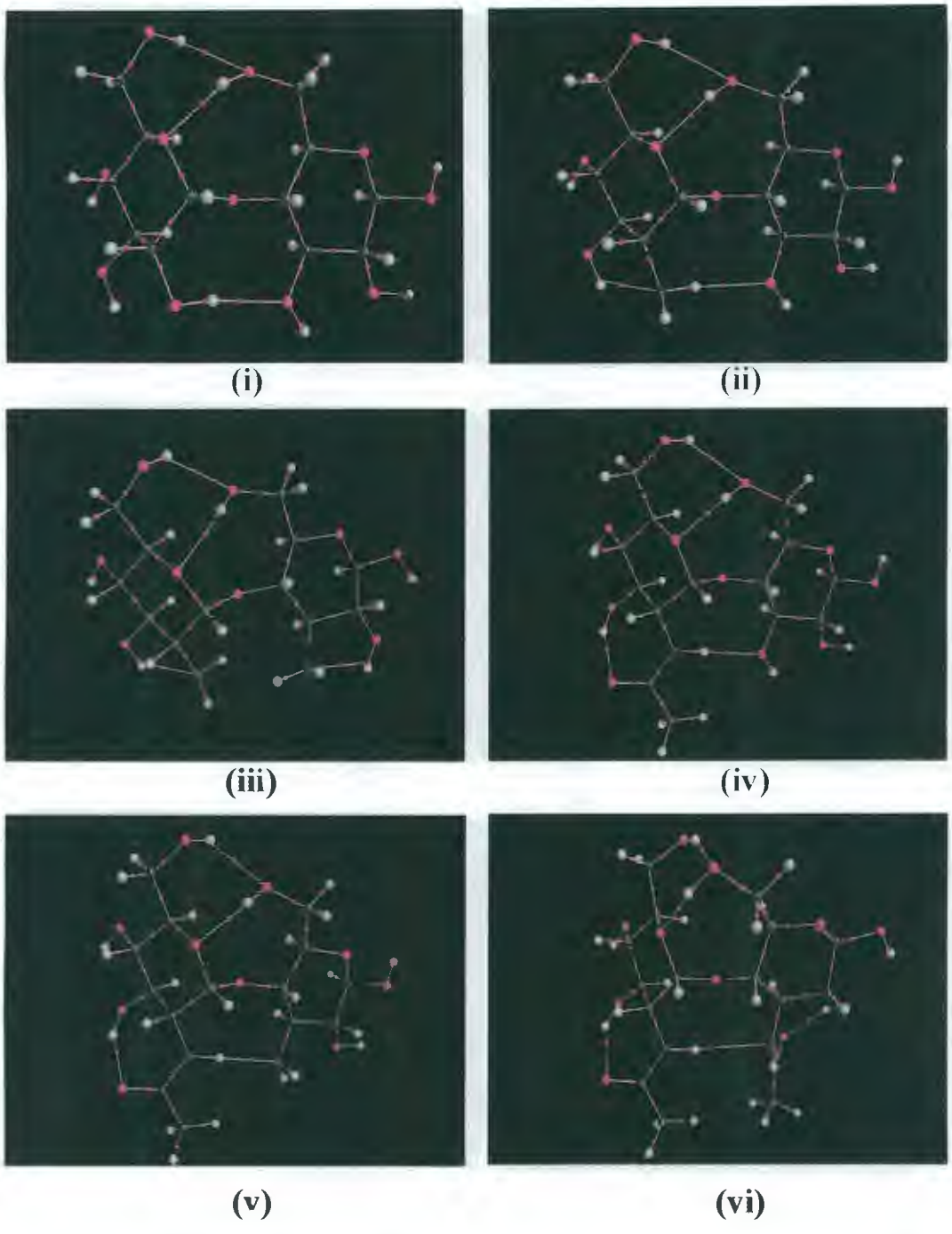


Figure 4.5 The molecular graphs of the B3LPY/6-311++G(d,p) optimised α(1→4) disaccharides where the small red spheres indicate the positions of the BCPs. The disaccharides shown are (i) maltose, (ii) 2-amine-Glc-α-(1→4)-Glc, (iii) 2-amine-Glc-α-(1→4)-3-amine-Glc, (iv) GlcNAc-α-(1→4)-Glc, (v) GlcNAc-α-(1→4)-3-amine-Glc and (vi) GlcNAc-α-(1→4)-GlcNAc.

4.3.4 Intramolecular Hydrogen Bonding: Disaccharides

The binding energy of intermolecular hydrogen bonds can be readily and accurately assessed using model pairs of functional groups such as the ones analysed in the previous section. However, it is nearly impossible to tease out the contribution of individual intramolecular hydrogen bonds to the internal energy of a macromolecule since electrostatic, steric and electronic effects are always present. It is here that the predictive value of the correlation ΔE^{HB} vs Δ_{el+lap} curves shown in Figure 4.4 can be demonstrated. We investigate the strength of hydrogen bonds across glycosidic linkages for both α and β(1→4) linked disaccharides as illustrated in Figure 4.1 using the correlation curves of Figure 4.4.

The conformations of oligosaccharides are dictated by the flexibility of the glycosidic linkages and in turn the hydrogen bonds formed across those linkages. We have previously investigated the nature of the α(1→4) and α(1→6) linkages[46] and found that the flexibility of the α(1→4) linkage in solution is affected by the extent of rapid exchange between bridging intermolecular water hydrogen bonds, formed with the hydroxyls at C2 and C3', and the intramolecular hydrogen bonds usually found between these hydroxyls in the solid state.[10,11] Preliminary investigation into derivatives of α(1→4) linked saccharides was done with a view to the slowing down this exchange of intermolecular and intramolecular hydrogen bonds through increasing the intramolecular hydrogen bond strength by substituting functional groups of varied hydrogen bond accepting and donating properties at the C2 and C3' positions (Figure 4.1).[12]

B3LYP/6-31G(d) optimisations were carried out on all the disaccharides combinations listed in Figure 4.1. Combinations of clockwise and anticlockwise hydroxyl orientations on the reducing and non-reducing sugars were optimised and compared as a way to identify conformations with the lowest energies. The low energy conformations for each of the disaccharides were further optimised with B3LYP/6-311++G(d,p) and the resultant electron density analysed using AIM calculations. The molecular graphs of six global minimum geometry α(1→4) linked disaccharides (maltose, 2-amine-Glc-α-(1→4)-Glc, 2-amine-Glc-α-(1→4)-3-amine-Glc, GlcNAc-α-(1→4)-Glc, GlcNAc-α-(1→4)-3-amine-Glc and GlcNAc-α-(1→4)-GlcNAc) are displayed in Figure 4.5 (i-vi) respectively. Similarly the corresponding six β(1→4) linked disaccharides (cellobiose, 2-amine-Glc-β-(1→4)-Glc, 2-amine-Glc-α-(1→4)-3-amine-Glc, GlcNAc-β-(1→4)-Glc, GlcNAc-β-(1→4)-3-amine-Glc and GlcNAc-β-(1→4)-GlcNAc) are shown in Figure 4.6 (i-vi) respectively.

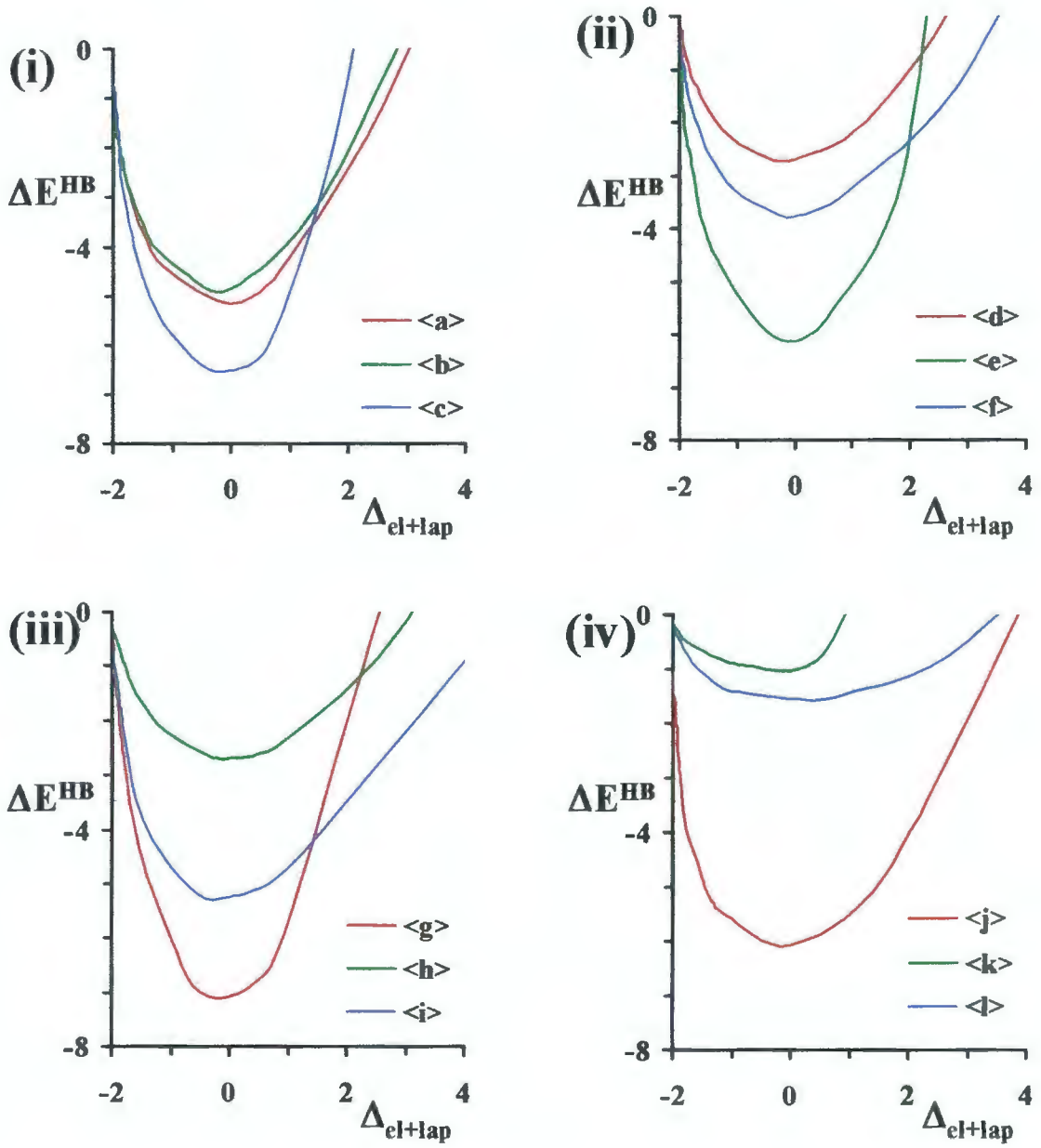


Figure 4.4 Correlation Δ^{HB} vs Δ_{el+lap} curves of the constructed using a B3LYP/6-311++G(d,p) basis set: (i) <a> Me-OH---Me-OH, <b> Me-OH(D)---Me-NAc(A), <c> Me-OH(A)---Me-NAc(D), (ii) <d> Me-NH_2---Me-NH_2, <e> Me-NAc(D)---Me-NH_2(A), <f> Me-NAc(A)-Me-NH_2(D), (iii) <g> Me-NH_2(A)---Me-OH(D), <h> Me-NH_2(D)---Me-OH(A), <i> Me-OH(D)---Me-OCH_3(A) (iv) <j> Me-NAc---Me-NAc, <k> Me-NAc(D(H))---Me-NAc(A(N)), and <l> Me-OCH_3---Me-OCH_3.

4.3.3 Electron Density Hydrogen Bond Correlation Curves

We use the AIM procedure to analyse the electron density in the hydrogen bonds of the pairs of functional groups listed in Table 4.1. Except for the methylated and deoxy-methaylated dimmers, all of the hydrogen bonded pairs comply with the criterion required for hydrogen bonding summarized by Pacios and Gómez.[8] It has been suggested by Cheeseman *et al.* that the near-perfect transferability of electron density of atoms does contribute to the total energy.[42] This underlying relationship is illustrated in Figure 4.3 (iii) and (iv).

Based on this postulation and supported by the results of Grabowski[30] we construct the standardised correlation curves by plotting the binding energies of the fully optimised complexes (ΔE^{HB}) evaluated from B3LYP/6-311++G(d,p)//B3LYP/6-311++G(d,p) against the modified Grabowski parameter Δ_{el+lap}. The results of a survey of crystal structures showed that the directionality of hydrogen bonds were consistent with the diffusive nature and the directional properties of the lone-pair electron densities of acceptors atoms.[43] Furthermore the effect of the $\boldsymbol{D}$-$\mathbf{H}_{\boldsymbol{D}}$---$\boldsymbol{A}$ angle on the binding energy of the Me-NAc hydrogen bonded dimer has been investigated using AM1 semi empirical methods. The results showed a very minor correlation between the variation in angle and the hydrogen bond energy.[44] Therefore, while hydrogen bonds are directional and the overall range of the $\boldsymbol{D}$-$\mathbf{H}_{\boldsymbol{D}}$---$\boldsymbol{A}$ angle is limited between 180° and 140°, for a medium to strong hydrogen bond, the distance between the hydrogen donor and acceptor atoms makes the most significant contribution to the strength of the hydrogen bond. This being the case we kept the $\boldsymbol{D}$-$\mathbf{H}_{\boldsymbol{D}}$---$\boldsymbol{A}$ angle fixed at the value obtained from the optimised geometry (generally not smaller than 155°) and varied the distance between $\boldsymbol{D}$ and $\boldsymbol{A}$ from a minimum distance where the van der Waals radii of the acceptor atom and the donor hydrogen are in contact to a maximum distance where the electron density is zero for each of the fully optimised combinations of functional donor-acceptor pairs. The modified Grabowski parameter and binding energy were calculated for each $\boldsymbol{D}$ and $\boldsymbol{A}$ distance and so a binding energy - Δ_{el+lap} correlation was established for each functional group pair. The Δ_{el+lap} versus the binding energy correlation curves for all the hydrogen donor-acceptor pairs are shown in Figure 4.4. The parabolic nature of these curves allows for comparison between the interacting pairs. It is clear that the electron density in the hydrogen bond is a strong indicator of the hydrogen bond strength, which implies that electron delocalisation, as in an actual chemical bond, is a significant contributor to the character of hydrogen bonds.[45]

electron density in the ***D*-H$_D$---*A*** hydrogen bond is linear for both the MP2 and B3LYP methods. This result is consistent with the suggestion from Cheeseman *et al.* that a near-perfect transferability of electron density of the atoms does contribute to the total energy.[42] This linear relationship between the binding energy strength and electron density clearly illustrates the covalent characteristics of the hydrogen bond.

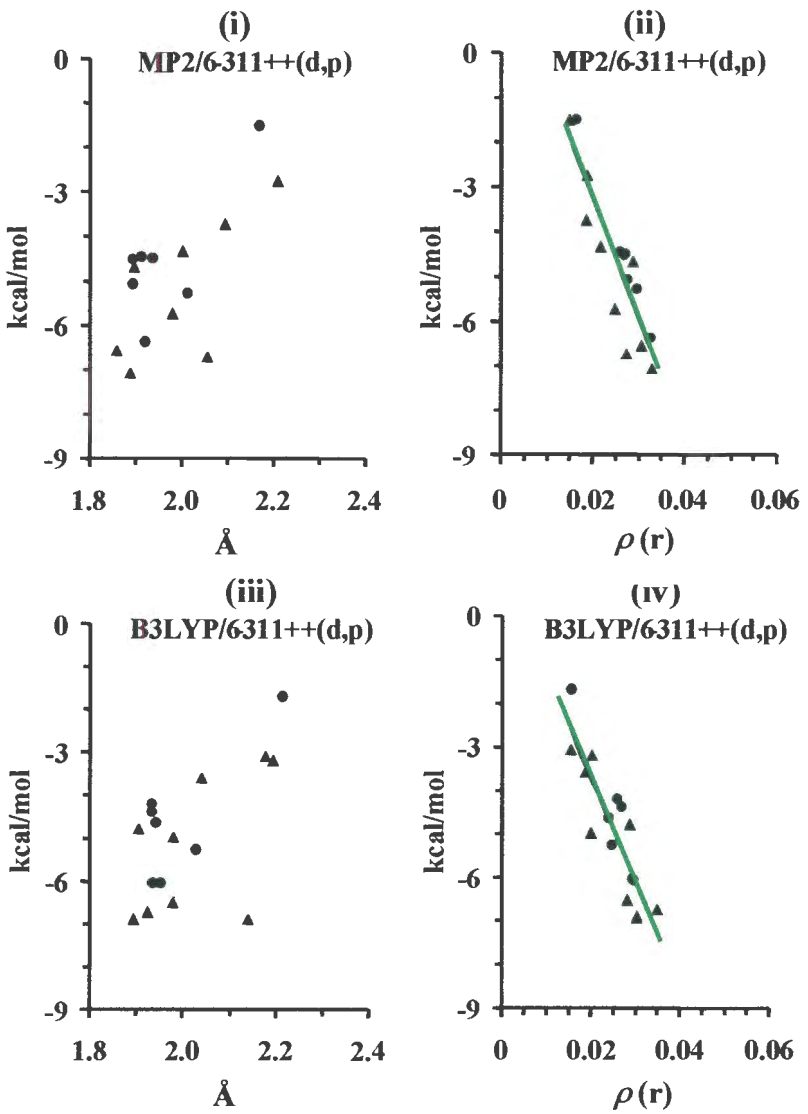


Figure 4.3 Binding energies of various hydrogen bonded dimmer complexes were plotted as a function of D-H·····A distances (Å) (represented by (i) and (iii)) and as electron density ρ(r) found within hydrogen bond (represented by (ii) and (iv)).

5.2.2 Conformational Analyses

5.2.2.1 *Adiabatic Maps*

Simulated annealing involves performing molecular dynamics at a sufficiently high temperature as a means of overcoming the rotational energy barrier of the exocyclic functional groups. This is then followed by a slow cooling process which allows the molecule to gradually reach a conformation of low potential energy. A reliable adiabatic map could be constructed by means of the following systematic procedures:

Initial Grid Search (i) The conformational space of the disaccharides is explored according to the dihedral angles rotation. Harmonic constraints were applied to the endocyclic dihedral angles in order to prevent the monosaccharides flipping out of their natural chair conformation. Glycosidic angles were also harmonically constrained and the ϕ,ψ energy space of the disaccharide was divided in terms of grids by sampling at 20 degree intervals for each glycosidic dihedral. (ii) During the simulated annealing process, each conformer was rapidly heated from 200K to 900K in the increments of 100K per picosecond over 7ps. The dynamics were further equilibrated for another 3ps at 900K heatbath before the cooling process in attempting to explore conformatione of lower energy. (iii) The cooling process was performed from 900K to 300K in 30K per picosecond over 20ps. The rate of the cooling is purposely set at a slower pace to allow the molecule to explore the lower energy regions. The structures obtained from the simulated annealing process were subject to a combined-methods minimisation scheme involving 100 steps of steep descent followed by an additional 100 steps of conjugated gradient. The outcome from the initial grid search provides a rough idea of the topology of the adiabatic map.

Map Refinement: The energy surface generated from the initial grid search does not guarantee finding the conformation representing the lowest energy. In order to clarify this issue, a returning procedure must be performed.

The orientation of the exocylic functional groups found within the different minima on the initial map should practically be very different. Therefore, it is reasonable to assume that for any disaccharide, the orientation of the functional groups yielding the lowest molecular energy would be similar to the exocylic group arrangement found within one of the minima within the adiabatic map. From the initial grid search, 6 different conformers, each with different combinations of ϕ,ψ dihedrals, were taken from 6 very different regions of the map. Each conformer was then employed to conduct further simulated annealing process for 35 cycles as was described previously. This procedure ensures the one to uncover the specific orientations of the exocylic groups which yield the lowest molecular energy.

Finally, the six conformations with the lowest molecular energy from the above mentioned procedure were selected for the map refinement procedure. The map refinement was carried out using the same combined-methods minimisation scheme at 10 degree interval grid search. These

arrangements of exocyclic groups were compare at each grid point as the and only the arrangements yielding the lowest energy structures were taken as a final energy point on the final map.

5.2.2.2 *Vacuum Dynamics*

Vacuum molecular dynamics simulations were implemented by keeping the chemical bonds involving hydrogens fixed via the constraint algorithm SHAKE[25] and employing a 1fs integration time step. The long range interactions were decreased smoothly to zero between 12 Å and 14 Å using the CHARMM switching functions applied on a group-by-group basis. The molecular dynamics simulations were conducted using the microcanonical ensemble (constant N, V, E) and used to explore the phase space of the saccharides (refer to Chapter 3). Configurations of the molecules were stored at intervals of 0.5ps in all simulations and analysed over 6ns long trajectories.

5.3 Results and Discussions

5.3.1 Parameter Developments

As was mentioned in chapter 2, depending on the conformation of the molecule, the MSK charge calculations can yield a variety of atomic charges. One can clearly observe the change in the atomic charge of atoms involving in hydrogen bonding when the geometry of the molecule is altered. This effect was demonstrated by the NH_2 ---- OH disaccharide mimics data as shown in Figure 5.3.

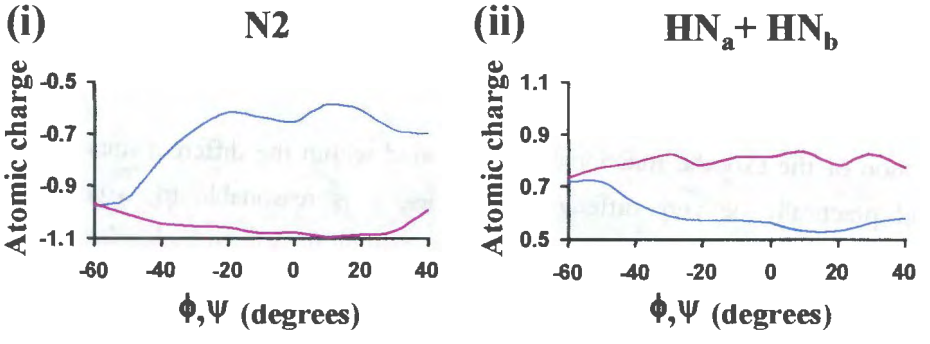


Figure 5.3 The ϕ and ψ dihedrals of the disaccharide mimic influence the atomic charge for (i) N2 and (ii) the net atomic charge of the amine hydrogens. (Obtained from MSK scheme where blue lines resemble N2(A) ---- HO3'(D) and pink lines resemble $HN_{(a\ or\ b)}$ (D) ---- O3'(A)).

The average atomic charge of N2 is -0.72 when it participated as a hydrogen bond acceptor with the HO3-O3 hydrogen bond donor group (shown in the blue curve of Figure 5.3(i)) However, when N2

participates as a hydrogen bond donor itself, the average atomic charge of N2 is -1.05 (shown in the pink curve in Figure 5.3(i)). Similarly, the average atomic charges of the amine hydrogens range from 0.29 to 0.40 depending on the hydrogen bonding orientation. It is interesting to observe the dramatic variation in the atomic charges of these atoms depending on the orientation hydrogen bonding. This confirms that the nitrogen atom is a superior hydrogen bond acceptor but a poor hydrogen bond donor (See chapter 4 for the characteristics of hydrogen bonding).[26]

Unlike the electronic structural approach, the empirical force field method requires each atom can only be assigned with one atomic charge regardless of the hydrogen bonding orientation. Therefore, different combinations of atomic charges were assigned to the N2, NH_a, NH_b, C2 and H2 atoms in an attempt to produce an empirical energy curve that has a comparable topology to the energy profile generated from ESA calculations Some of the atomic charge combinations tested are listed in Table 5.1 ((b) to (j)) and their respective empirical energy profiles are shown in Figure 5.4 and Figure 5.5 and further compared with the energy well obtained from density function theory (Table 5.1 (a)).

Table 5.1. The different combinations of atomic charges employed during the CSFF parameter development of NH2 functional groups

	QM Derived and Empirical Assigned Atomic Charges									
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)
C2	B3LYP/ 6-311++G(d,p)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
H2		0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.05	0.10
N2		-0.40	-0.50	-0.60	-0.70	-0.80	-0.90	-1.00	-0.80	-0.80
NHa		0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.15	0.15
NHb		0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.15	0.15

The internal rotation barriers about single bonds have long been recognized as one of the major factors governing the conformations of a molecule. In Figure 5.4, altering the atomic charge scheme induces only a subtle variation to the energy topology for rotating about the C2-*X* bond of the monosaccharide mimic. However, due to the non-bonded interaction between the functional groups, the slight alteration of the atomic charge on the atoms results in considerable differences on the overall topology of the disaccharide mimic potential energy profiles (Figure 5.5). This intricate relationship between the potential energy profile and atomic charge makes it a difficult task to decide with which atomic charge assignment scheme to incorporate as the finalised parameter. Moreover, modification of the atomic charge of one functional group leads to changes in the potential energy profiles for other disaccharide mimics involving thus particular functional group. For example, fine tuning the functional groups of NAc and NH_2 parameters using the mimics involving -NAc $\cdots$ -NH_2 would affect the outcomes of all of these interactions: -NAc $\cdots$ -OH, -NAc $\cdots$ -NAc, -OH $\cdots$ -NH_2 and -NH_2 $\cdots$ -NH_2.

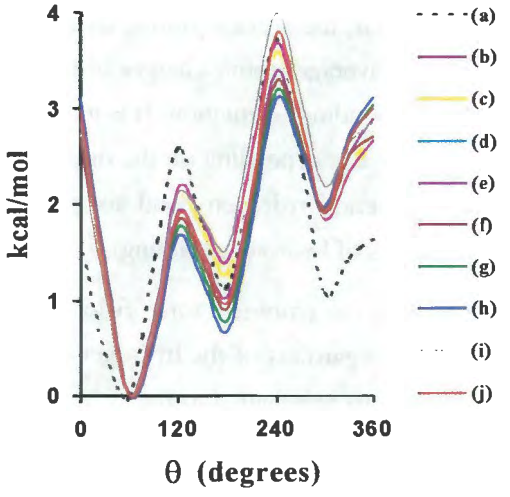


Figure 5.4 **Comparing C2-NH_2 dihedral rotational energy profile of 2-amine-5-methyl-pyranose (fitting different empirical energy profiles (Curves (b) to (j) represent different CHARMM empirical energy profiles whereas curve (a) represents the results obtained from B3LYP calculation).**

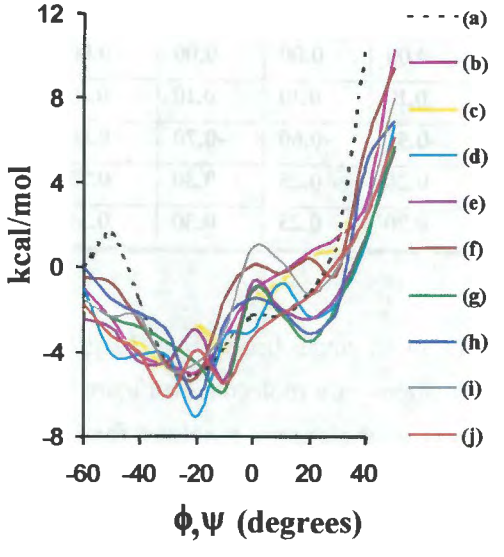


Figure 5.5 **Subtle changes of charge parameters can induce drastic changes in the overall potential energy profiles of an NH_2----OH disaccharide mimic. (Curves (b) to (j) represent different CHARMM empirical energy profiles whereas curve (a) represents the results obtained from B3LYP calculation).**

Due to the space constrain, the details of all the tedious procedure are not include in the rest of the chapter. The results obtained from the final parameters are shown in Figure 5.6 (the dihedral rotation

of C2-*X* bond) and 5.7 (rotating the glycosidic dihedral angels under the influence of the non-bonded interaction of -*X* and –*Y*). These curves were generated after taking all the important factors mentioned previously in to consideration while making some reasonable assumptions of fine turning the parameters. The new force field parameters (red lines) are shown to be comparable with the curves generated from B3LYP calculations (green lines).

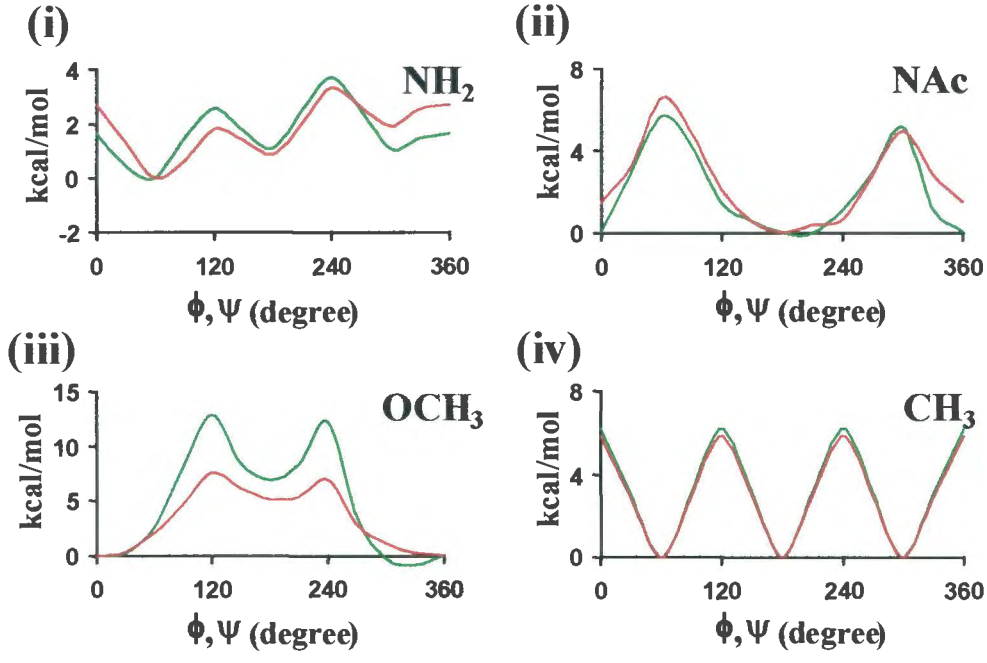


Figure 5.6 **The empirical potential energy profiles for rotating various functional groups on the corresponding monosaccharide mimic.**

Corzana *et al.*[11] performed a series of solution dynamic simulations of methyl-α-D-maltoside and methyl-α-D-isomaltoside and found that rotation of the -*O*-methyl group is somewhat rigid when using the CSFF. Although it is not clear how such a conclusion can be drawn from the CSFF which at the time did not include potential parameters for -*O*-methyl groups, this warning has been taken into consideration. In order to promote free rotation of the -*O*-methyl groups, the necessary potential parameters have therefore been adjusted by lowering the rotational barrier height. Figure 5.6(iii) shows the overall energy profile of the newly adjusted parameters still holds remarkable similarity when compare to the energy curves generated quantum mechanical calculations. Nevertheless, after examining Figure 5.6 and 5.7, it is clear that the overall profiles between the empirical potential energy curves and the corresponding B3LYP generated energy profiles have been shown to be analogous. The force field parameters were hence deemed to be acceptable for the conformational analyses applications.

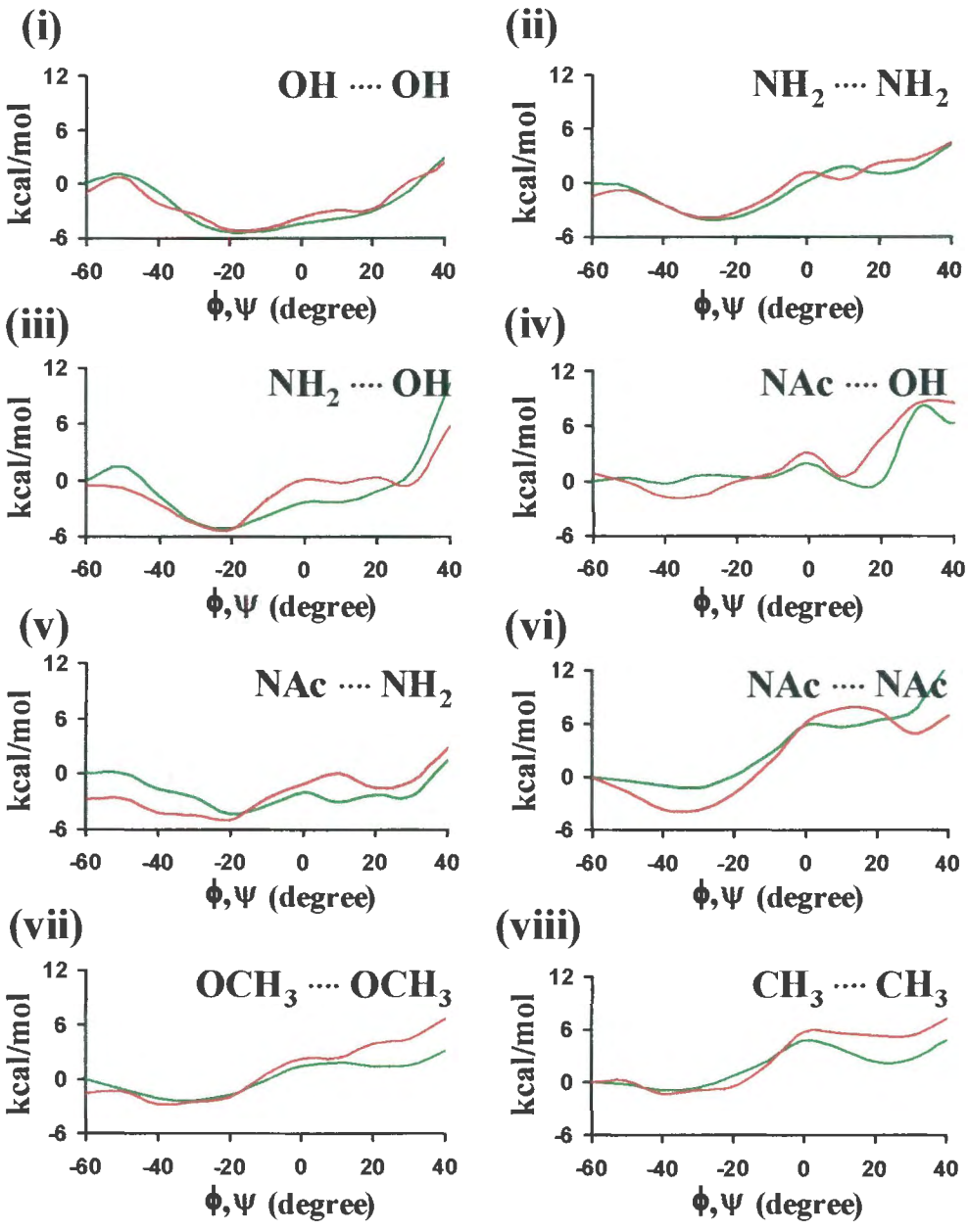


Figure 5.7 **The empirical potential energy profiles for different disaccharide mimic generated by simultaneously rotating both glycosidic dihedrals (ϕ and ψ).**

5.3.2 Conformational Analyses

5.3.2.1 Adiabatic Maps

After modifying the CHARMM-like carbohydrate parameter, we performed conformational analyses on all the α(1→4) linked disaccharides listed previously. However, due to the space constraints, only three disaccharides (maltose, 2-amine-Glc-α-1→4)-Glc and GlcNAc-α-(1→4)-3-amine-Glc) were

incorporated in this section for demonstrating the effect of increasing cross glycosidic hydrogen bonds on the glycosidic flexibility. Adiabatic energy surfaces E(ϕ,ψ) for each of the diasaccharides were produced and are shown in Figure 5.8(i), (ii) and (iii). The molecules were chosen by systematically substituting the hydroxyls at positions C2 and C3' with NH2 and NAc groups so as to maximise the hydrogen bonding across glycosidic linkage based on the ranking from chapter 4.

X-ray[17b] and neutron diffraction[17c] experiments on maltose reveal ϕ,ψ values for the glycosidic linkage that are located in the central energy well of the adiabatic map shown in Figure 5.8(i). In addition, NMR[17d] and optical rotation experiments[17e] performed in aqueous solution predict values around ϕ = -70, ψ = -35 for maltose. The lowest contours represent by the purple regions (9kcal/mol barrier) of the adiabatic maps in the central well for maltose (Figure 5.8(i)) and 2-amine-Glc-a-(1→4)-Glc (Figure 5.8(ii)) are spaced far from each other compared with the GlcNAc-α-(1→4)-3-amine-Glc disaccharide (Figure 5.8(iii)). The purple regions of the adiabatic maps resemble the global minima of 9kcal well-depth. The maps indicated that both the Glc-2-amine-α-(1→4)-Glc and GlcNAc-α-(1→4)-3-amine-Glc have a smaller accessible conformational space than the maltose.

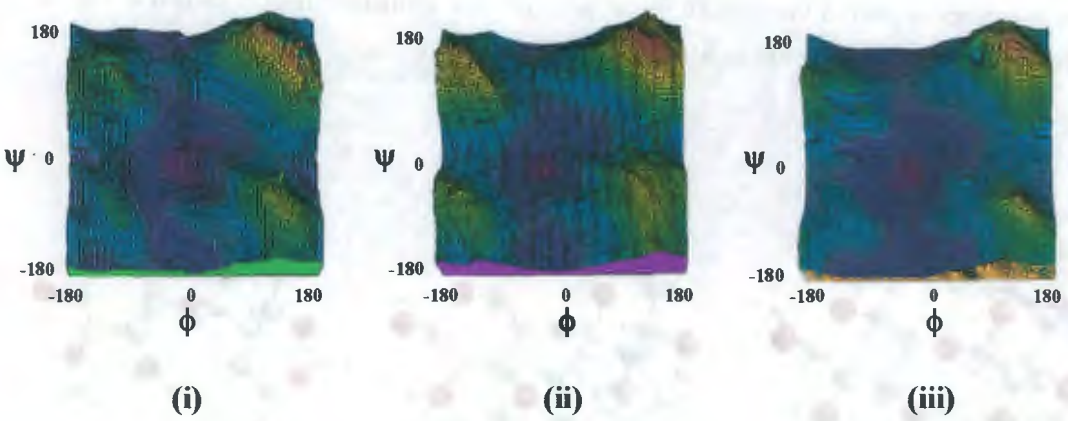


Figure 5.8 **Adiabatic maps for (i) β-maltose, (ii) 2-NH_2-Glc-α-(1→4)-Glc, and (iii) GlcNAc-α-(1→4)-3-NH_2-Glc.**

5.3.2.2 Vacuum Dynamics

The ϕ,ψ time series from a series of 6ns vacuum molecular dynamics simulations of β-maltose, 2-amine-Glc-α-(1→4)-Glc and GlcNAc-α-(1→4)-3-amine-Glc have been extracted and are shown superimposed upon the adiabatic maps. The results can be viewed in Figure 5.9 (i), (ii) and (iii) respectively. The motion of the GlcNAc-α-(1→4)-3-amine-Glc disaccharide is more restricted with respect to rotation about the ϕ,ψ dihedral angles. Even though both Glc-2-amine-α-(1→4)-Glc and GlcNAc-α-(1→4)-3-amine-Glc have a smaller accessible conformational space than maltose, GlcNAc-α-(1→4)-3-amine-Glc shows a distinctive smaller sampled region.

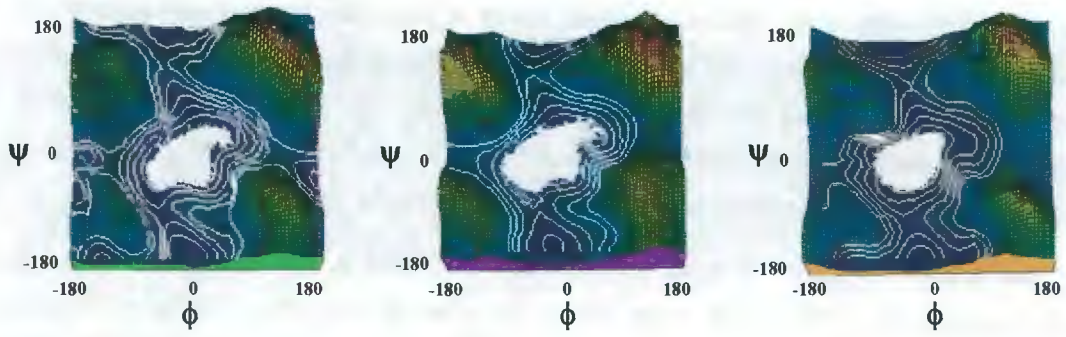


Figure 5.9 **Superimposing the scattering distribution of the dihedral angle obtained from the vacuum molecular dynamics trajectories (represented by the central white regions) onto adiabatic maps ((i)β-maltose, (ii) 2-amine-Glc-α-(1→4)-Glc, and (iii) Glc-NAc-α-(1→4)-3-amine-Glc).**

The differences in dynamics of the three disaccharides can be explained on inspection of the global minima structures shown in Figure 5.10 (i), (ii) and (iii). The numerical figures shown in Figure 5.10 represent the inter-atomic distances (Å).

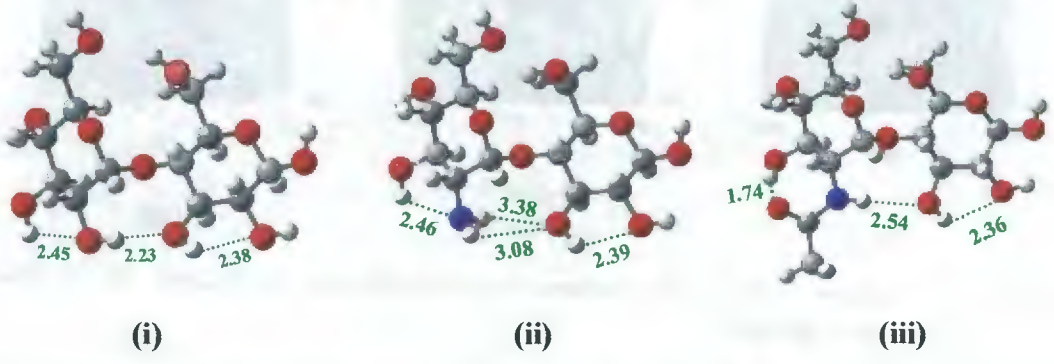


Figure 5.10 **Global minima conformations for maltose (i), 2-amine-Glc-α-(1→4)-Glc (ii), and (iii) Glc-NAc-α-(1→4)-3-amine-Glc. The numerical figures indicate the distance (Å) between the two specified atoms.**

The β-maltose molecule (Figure 5.9(i)) exhibits a series of hydroxyl-hydroxyl hydrogen bonds starting from the non-reducing sugar across the glycosidic linkage to the reducing sugar. This β-maltose structure coincides with the β-maltose crystal structure which was found to adopt the conformation of (ϕ,ψ) = (0,13).[28] The 2-amine-Glc-α-(1→4)-Glc disaccharide (Figure 5.10(ii)) has a stronger hydrogen bond between the hydroxyl and the amine on the non-reducing sugar (i.e., OH(***D***)----

NH_2(*A*)) followed by a weaker NH_2(*D*)—OH(*A*) hydrogen bond across the glycosidic linkage and a moderate hydroxyl-hydroxyl hydrogen bond on the reducing sugar. This weaker cross glycosidic linkage hydrogen bond allows for increased motion about the ϕ,ψ torsion angles. The third disaccharide GlcNAc-α-(1→4)-3-amine-Glc (Figure 5.9(iii)) includes a medium strength OH(*D*)—NAc(*A*) hydrogen bond between the hydroxyl and the carbonyl oxygen of the NAc on the reducing sugar that impedes rotation of the NAc group. This is followed by another medium strength hydrogen bond across the glycosidic linkage (i.e., NAc(*D*)—NH_2(*A*)) through to a relatively weaker hydrogen bond between the amine and the hydroxyl (i.e., NH_2 (*D*)—OH(*A*)) on the reducing sugar. Because of the increased hydrogen bonding between the functional groups adjacent to the glycosidic linkage in GlcNAc-α-(1→4)-3-amine-Glc impedes rotation of the NAc group a more limited volume of conformational space is sampled.

5.4 Conclusion

The lack of experimental data becomes a significant challenge for the development of any new force field parameters. Relying on the electronic structural approach to provide target values have been widely accepted as the alternative method of parametersing the EFFs. The CSFF parameters have been proven to yield simulations results which are comparable with the experimental findings. In this chapter, we extended this force field by introducing several new parameters for modelling these four small functional groups (-NH_2, NAc, -OCH_3 and -CH_3). The empirical energy profiles of these functional groups were further compared with the data obtained from the quantum mechanical calculations as a mean of validation. As was demonstrated, the overall energy profiles generated from CHARMM have to been shown to be fairly comparable with the curves obtained from the B3LYP calculation. The force field parameters were deemed to be acceptable for calculations involving these groups.

Conformational analyses of maltose and its derivatives revealed that the GlcNAc-α-(1→4)-3-amine-Glc sugar has relatively limited motion about the ϕ,ψ torsion angles. In general, although one can not concluded the solvation nature of the molecules by simply relying on the conformation analyses that have been carried out in this chapter, but these analyses do provide a valuable groundwork for designing new type of biopolymer. In addition, they enable one to selectively pre-eliminate the molecules with unwanted characters and focus on the molecules with potential. Nevertheless, in order to confidently deduce the intricate solvation nature of any molecule, the molecule has to be studied within the conditions where the solvent molecules are explicitly incorporated.[27] This issue would be explored in further in the next chapter.

References

(1) Sharon, N.; Lis, H., *Science*, **1989**, *246*, 227.

(2) van Oss, C. J., *Colloids and Surfaces, B: Biointerfaces* **1995**, *5(3/4)*, 91.

(3) (a) Hwang M. J.; Ni X.; Waldman M.; Ewig C. S.; Hagler A., *Biopolymer* **1998**, *45(6)*, 435. (b) Spieser, S. A. H.; Albert van Kuik, J.; Kroon-Batenburg, L. M. J.; Kroon, J., *Carbohydr. Res.* **1999**, *322(3-4)*, 264. (c) Asensio, J. L.; Jimenez-Barbero, J., *Biopolymers* **1995**, *35(1)* 55. (d) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B., *J. Phys. Chem.* **1995** *99(11)*, 3832. (e) Allinger, N. L.; Rahman, M.; Lii, J. H., *J. Am. Chem. Soc.* **1990**, *112(23)*, 8293. (f) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J., *J. Am. Chem. Soc.* **1996**, *118(45)*, 11225. (g) Cramer, C. J., *J. Org. Chem.* **1992**, 57(26), 7034.

(4) Ha, S. N.; Giammona, A.; Field, M.; Brady, J. W., *Carbohydr. Res.* **1988**, *180(2)*, 207.

(5) Palma, R.; Himmel, M.; Liang, G. Brady, J. W., *ACS Symposium Series* **2000** 769(glycosyl hydrolases for biomass conversion)

(6) (a) Brady, J. W.; Schmidt, R. K., *J. Phys. Chem.* **1993**, 97(4), 958. (b) Brady, J. W., *J. Am. Chem. Soc.* **1989**, *111(14)*, 5155.

(7) Homans, S. W., *Biochemistry*, **1990**, *29*, 9110.

(8) Kirschner, K. N.; Woods, R. J., *PNAS* **2001**, 98(19), 10541.

(9) (a) Momany, F. A; Willett, J. L., *Carbohydr. Res.* **2000**, 326(3), 210., (b) Momany, F. A.; Willet, J. L. *J. Comput. Chem.* **2000**, *21*, 1204.

(10) Kuttel, M.; Brady, J. W.; Naidoo, K. J., *J. Comput. Chem.* **2000**, *23(13)*, 1236.

(11) Corzana, F.; Motawia, M. S.; Herve Du Penhoat, C.; Perez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. B., *J. Comput. Chem.* **2004**, *25(4)*, 573.

(12) Reiling, S.; Schlenkrich, M.; Brickmann, J. *Comput. Chem.* **1996**, *17*, 450.

(13) (a) Tvaroska, I.; Carver, J. P., *Carbohydr. Res.* **1988**, 309*(1)*, 1. (b) Tvaroska, I.; Carver, J. P., *J. Phys. Chem.* **1994**, *98(26)*, 6452. (c) Tvaroska, I.; Carver, J. P., *J. Phys. Chem.* **1994**, *98(38)*, 9477. (d) Tvaroska, I.; Carver, J. P., *J. Phys. Chem.* **1995**, *99(16)*, 6234.

(14) Best, R. E.; Jackson, G. E.; Naidoo, K. J., *J. Phys. Chem. B.* **2001**, *105*, 4742-4751.

(15) (a) Tran, V. H.; Brady, J. W., *Biopolymers* **1990**, *29(6-7)*, 961. (b) Ueda, K.; Brady, J. W., *Biopolymers* **1996**, 38*(4)*, 318

(16) Cloran, F.; Carmichael, I.; Serriani, A. S. *J. Am. Chem. Soc*, **1999**, *121*, 9843-9851.

(17) (a) Brady, J. W.; Schmidt, R. K. *J. Phys. Chem.* **1993**, *97(4)*, 958. (b)Takusagawa, F.; Jacobson, R. A. *Acta Cryst.* **1978**, *B34*, 213. (c) Gress, M. E.; Jeffrey, G. A. *Acta Cryst.* **1977**, *B33*, 2490. (d) Shashkov, A. S.; Lipkind, G. M.; Kochetkov, N. K. *Carbohydr. Res.* **1986**, *147*, 175. (e) Stevens, E. S. *J. Am. Chem. Soc.* **1989**, *111*, 4149. (f) Stevens, E. S. *Biopolymers* **1992**, *32*, 1571

(18) Rees, D. A. *Polysaccharide shapes: outline studies in biology* Chapman and Hall Ltd., 11 New Tetter Lane, London EC4P 4EE, **1977**

(19) Naidoo, Kevin J.; Brady, J.W. *Chem. Phys.* **1997**, *224(2,3)*, 263.

(20) (a) Poppe, L., *J. Am. Chem. Soc.* **1998**, *181*, 8421 (b) Marchessault, R., Perez, S. *Biopolymers* **1979**, 18 2369. (c) Lipari, G.; Szabo, A., *J. Am. Chem. Soc.* **1982**, *104*, 4546.

(21) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., *J. Comput. Chem.* **1983**, *4(2)*, 187-217.

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

(23) Singh, U. C.; Kollman, P. A.; Nguyen, D. T.; Case, D. A., *J. Comput. Chem.* **1986**, 5, 129.

(24) Becke, A. D., *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(25) van Gunsteren, W. F.; Berendsen, H. J. C., *Mol. Phys.* **1977**, *34*, 1311-1327.

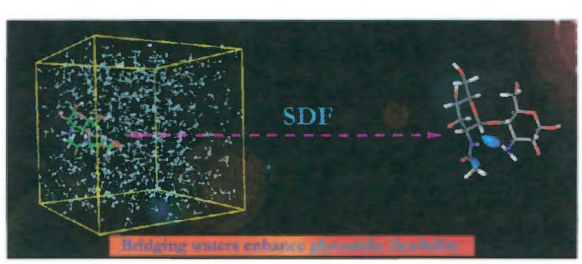
(26) Rablen, P. R.; Lockman, J. W.; L., J. W., *J. Phys. Chem. A.* **1998**, *102*, 3782.

(27) Brady, J. W. Solvation: *Carbohydrates. In Encyclopedia of computational chemistry*, 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, **1998**; p2609.

(28) Jeffery, G. A., *Carbohydr. Res.* **1988**, *222*, 47.

Chapter Six

The Role of Water in the Design of Glycosidic Linkage Flexibility



6.1 Introduction

Glycoproteins are proteins that are covalently linked to an oligosaccharide. The conformational flexibility of the oligosaccharide is an important factor that governs the many biological functions of this class of biopolymer. The monosaccharides that make up the oligosaccharide mostly have no more than one axial group and can therefore be largely approximated as "rigid" rings maintaining a chair conformation. Therefore rotations about the glycosidic linkages between the monosaccharides are the primary determinants of the conformation of oligosaccharides in glycoproteins. Predicting the shape of oligosaccharides by controlling the rotation of the glycosidic linkage is a major objective in the design of saccharide mimics for pharmaceutical applications. However, determining the solution conformation of carbohydrates is a significant challenge to experimental methods. In the case of NMR the long time-scale of the experiment compared with the often rapid internal motion of the saccharide leads to average and mostly virtual structures. Rotation about glycosidic linkages occurs on a comparatively shorter time scale and it has been shown that the resultant structures are best resolved in combination with computer simulations.[1] Molecular dynamics (MD) simulations have been used to interpret information from nuclear Overhauser effect (nOe) and spin-coupling NMR experiments on oligosacchrides in an effort to elucidate their conformations.[2] More recently a promising theoretical method was used to extract glycosidic linkage conformational distributions from NMR dipole-dipole coupling data.[3] However, to our knowledge, a detailed analysis of the solvent-saccharide relationship of disaccharides using NMR has not been performed.

The conformation of oligosaccharides composed of glucopyranosyl monomers is largely determined by rotation about the glycosidic linkage dihedrals. Molecular dynamics (MD) simulation data provide a level of atomistic detail that allows unambiguous characterization of molecular motion. Based on the vacuum dynamics simulation from the previous chapter, when compare the difference of dynamics behaviour between maltose and Glc-2-NAc-α-(1→4)-Glc-3-NH_2, it is apparent that the Glc-2-NAc-α-(1→4)-Glc-3-NH_2 clearly shown more rigidity in their internal motion. The next step of rational designing of new type of carbohydrate is to investigating the influence of water solvent molecule on the behaviour of Glc-2-NAc-α-(1→4)-Glc-3-NH_2.

As with many other biological molecules, carbohydrates undergo a variety of motions when immerse within the aqueous environment. The Voronoi tessellations and spatial distribution functions have been previously employed to analyse the structure of water about maltose and maltohexaose.[4] It was found that the hydrophilic groups (in this case hydroxyls) participate freely in hydrogen bonds adjacent hydrophilic groups and water molecules in the first hydration shell. There is therefore a competition between intramolecular and intermolecular hydrogen bonding. Of particular interest are the intramolecular hydrogen bonds occurring between neighbouring glucopyranose units (i.e., those formed across the glycosidic linkage). These have a significant effect on the rotational flexibility of the torsional angles comprising the glycosidic linkage. This chapter investigates the effect of

increasing the intramolecular hydrogen bonding between glucopyranose sugar residues across the linkage on oligosaccharide solution conformations.

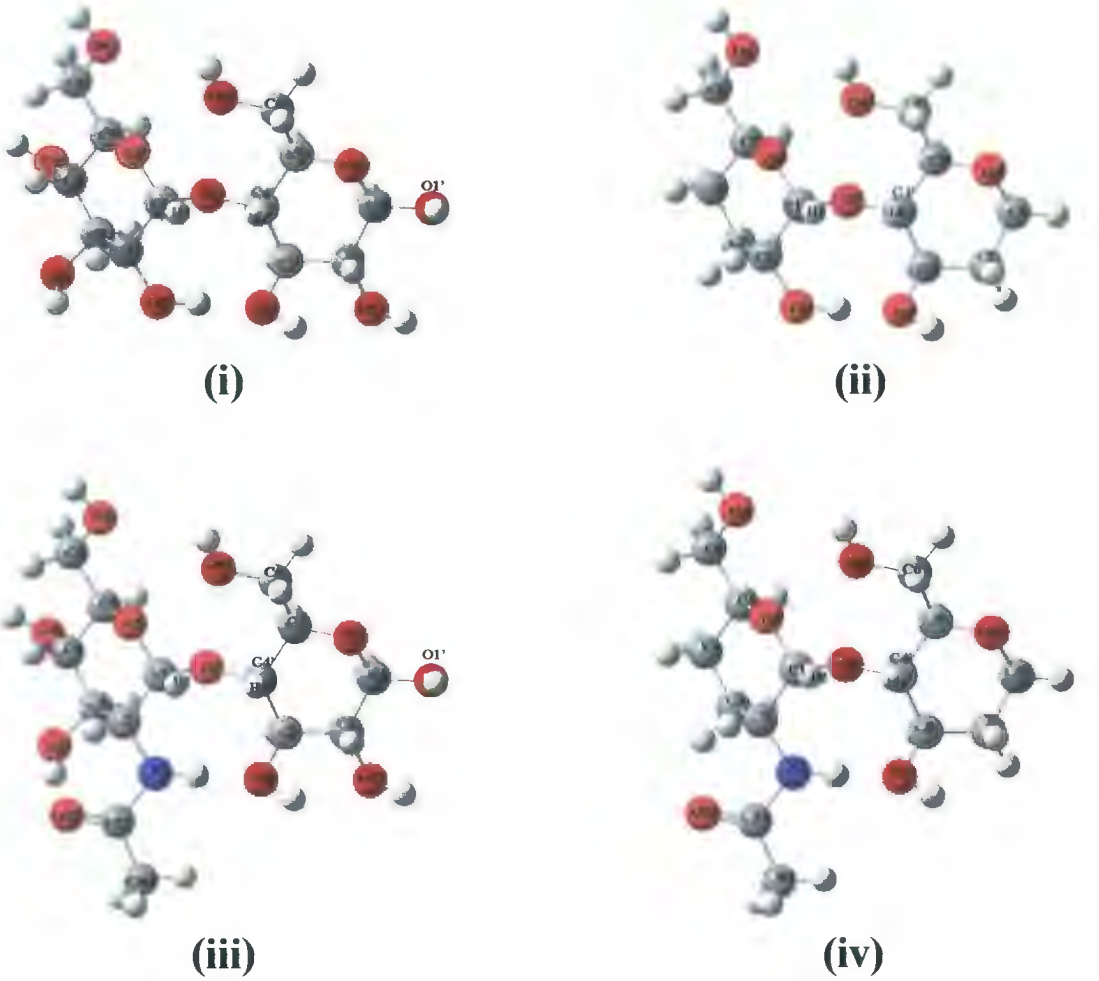


Figure 6.1 **Atom numbering scheme for (i) maltose, (ii) Glc-2-NAc-α-(1→4)-Glc-3-NH_2, (iii) Maltose mimic and (iv) Glc-2-NAc-α-(1→4)-Glc-3-NH_2 mimic.**

6.2 Computational Details

Schematic structures for maltose and GlcNAc-α-(1→4)-3-amine-Glc are depicted in Figure 1 (i) and (ii) respectively, showing the names used for the non-hydrogen atoms and the dihedral angles. The inter-residue dihedrals (ϕ,ψ) for disaccharides are defined as ϕ: H1-C1-O1-C4' and ψ: C1-O1-C4'-H4' according to previous convention.[5]

The program CHARMM has been used for all the molecular mechanics computations reported here.[6] A CHARMM-like force field specifically parameterized for carbohydrates[7] was used and appropriately modified for our computations. The TIP3P water model[8] was used to represent the solvent explicitly. Maltose and GlcNAc-α-(1→4)-3-amine-Glc solution simulations were started from the previously established vacuum global minima (ϕ = -23.77, ψ= -20.56) and (ϕ= -28.68, ψ=-17.39) respectively.[9] MD simulations were run for 6ns keeping chemical bonds involving hydrogens fixed via the constraint algorithm SHAKE[10] with a 1 fs integration time step. The long range interactions were decreased smoothly to zero between 12 Å and 14 Å using the CHARMM switching functions applied on a group-by-group basis.[11] All solution simulations were run using a isothermal-isobaric ensemble (NPT) where the pressure and temperature were kept constant (P=1bar and T=300K) using the Langevin Piston method as implemented in CHARMM.[12] In this scheme the pressure is controlled using a Langevin equation of the type:

$$\frac{d^2V}{dt^2} = \frac{1}{W}\left[P(t) - P_{ext}\right] - \gamma\frac{dV}{dt} + R(t) \qquad \text{eq 6.1}$$

where the volume, V, is treated as a dynamical variable, W=500 is the "mass" of the piston with a collision frequency of γ=5ps^{-1} and applied with a random force R(t) that is determined from a Gaussian distribution. Configurations of the molecules were stored at intervals of 0.5 ps in all simulations and analyzed over the entire 6ns trajectory.

Full geometry optimisations were performed using Gaussian 98[13] and AIM2000.[14] on disaccharide mimics of maltose and GlcNAc-α-(1→4)-3-amine-Glc (Figure 6.1 (iii) and (iv) respectively) each complexed with a single water molecule located at the glycosidic linkage. Density functional Becke-Three Lee-Yang-Parr (B3LYP) [15] level of theory was used in the optimisation along with the 6-31G(d) basis set. Due to the shallow potential curves of these systems, the keyword OPT=Tight was included to increase the convergence criteria. This ensures a better reliability of the final geometries. Basis set superposition errors (BSSE) were corrected with a Boys-Benardi functional counterpoise scheme [16] and energy evaluations where done with a 6-311++G(d,p) basis set.

6.3 Results and Discussions

6.3.1 Glycosidic Linkage Dynamics

We analyzed the motion of the glycosidic torsional angles from the 6ns solution MD simulations of the two disaccharides. The time series of the ϕ and ψ torsional angles are shown in Figure 6.2. The glycosidic torsional angles fluctuate about ϕ≈-40, ψ≈-20 in the case of both saccharides. However, the maltose glycosidic torsional angles (Figure 6.2 (i) and (ii)) appear to fluctuate less than the

GlcNAc-α-(1→4)-3-amine-Glc torsional angles (Figure 2 (iii) and (iv)). X-ray[17] and neutron diffraction[18] experiments on maltose reveal ϕ,ψ values for the glycosidic linkage that are located at ϕ≈-60, ψ≈-40. In addition, NMR[19] and optical rotation experiments[20] performed in aqueous solution predict values around ϕ = -70, ψ = -35 for maltose. The glycosidic dynamics observed here is therefore consistent with experimental findings for maltose and related disaccharides.

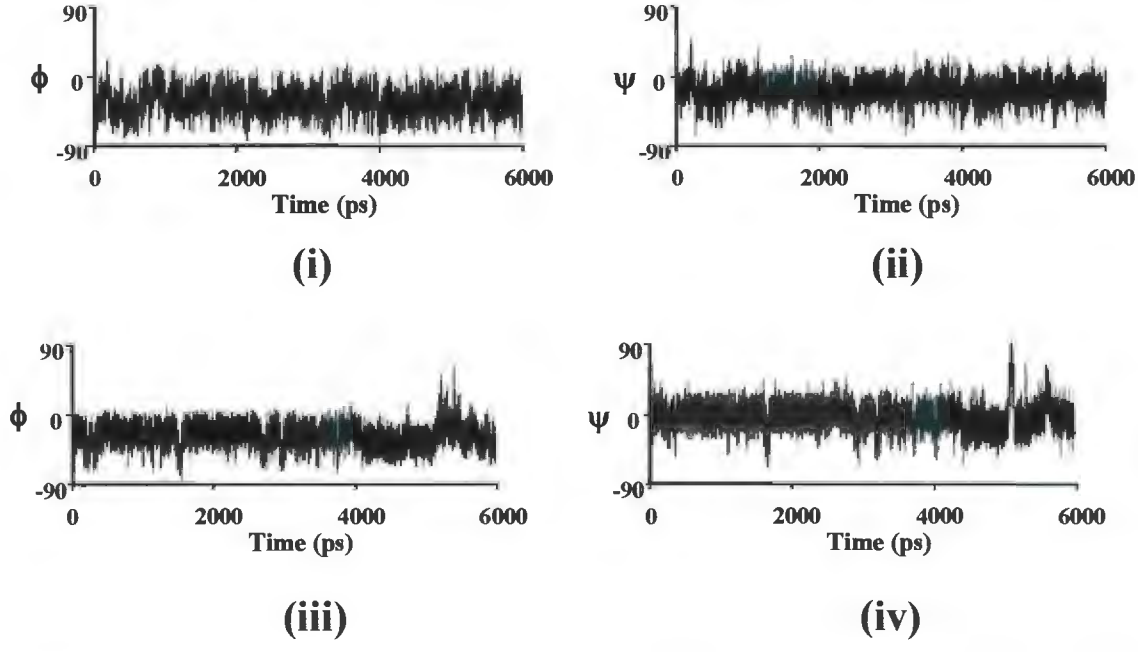


Figure 6.2 **ϕ and ψ time series from the 6ns MD simulations of maltose ((i) and (ii)) and Glc-2-NAc-α-(1→4)- Glc-2-NH_2 ((iii) and (iv)).**

Time correlation analysis of the fluctuations in a dihedral angle (e.g., Φ) indicates the extent to which the motion is correlated with that at a time *t* later. The *time correlation function* (TCF) for the fluctuation of a dihedral angle Φ is defined in chapter 3, where the average is over all the time origins possible for each value of t. Thus the correlation of Φ with the value at a later time may be examined. This gives an indication of how quickly Φ becomes uncorrelated (the *correlation time*) as well as any regular motion (manifested as later peaks in the function). The rate of TCF decay will give an indication of the rate of fluctuation. The solution $C_\phi(t)$ and $C_\psi(t)$ was calculated for both disaccharides and the results are shown in Figure 3. The TCF's for both dihedrals decay slightly faster for the GlcNAc-α-(1→4)-3-amine-Glc disaccharide than for maltose where the correlation time for maltose is 17ps and the correlation time for GlcNAc-α-(1→4)-3-amine-Glc is 19ps. The effect of water damping on the

glycosidic motion of the two disaccharides is similar despite the cross glycosidic linkage intramolecular hydrogen bonds being stronger in the former case.[9]

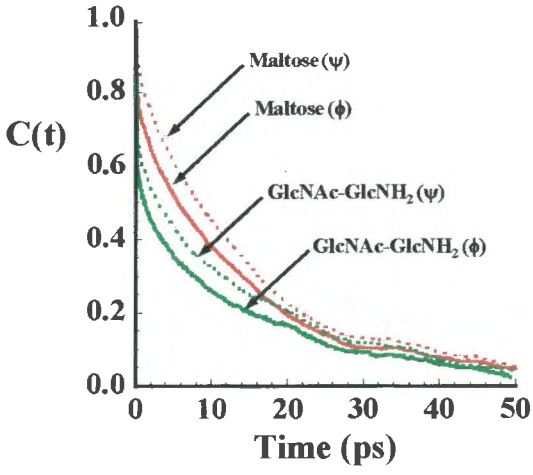


Figure 6.3 Time correlation analysis of the fluctuations in a dihedral angles ϕ and ψ comprising the glycosidic linkage in the disaccharides.

6.3.2 Solution Structures

The hydrophilic interactions between water and the sugar solute induce a specific distribution of solvent about the solute where water molecules that are closer to the solute tend to be localized in certain high probability regions compared with the normal "bulk water" solvent probability density further away.[4] Radially averaged pair distribution functions (PDF) between sugar oxygens or nitrogens and the water oxygens, $g(O_w)$,[21] provide approximate pictures of how the water is structures about each functional group. In maltose for example, the O3' oxygen is typical of most by having a strong first peak at 2.82Å, which on integration over a sphere of radius 3.5Å about the oxygen, results in a first shell of approximately 3.2 waters for the hydroxyl. This is much larger than the number of hydrogen bonded interactions calculated previously, so there is clearly significant local water density that is not participating in hydrogen bonds to the hydroxyls.[22]

The pair distribution function gives only approximate information about the nature of solvent structure about each disaccharide, as it is radially averaged: the position of the maximum is the same for all hydroxyls, although the maximum peak height may vary. Specific solute-solvent interactions, however, would be expected to give a complex directional distribution of water about each saccharide, which may not be revealed in the PDF. A more detailed picture may be obtained by calculating the full three-dimensional spatial distribution function (SDF).[4, 23] This is done by averaging the water distribution about each disaccharide over the whole trajectory, after correcting for translational and rotational solute motion. The PDF's have been calculated for the O2 and O3'

atoms of maltose (Figure 6.4 (i) and (ii)) and the N2, O22 and N3' atoms of GlcNAc-α-(1→4)-3-amine-Glc (Figure 6.4 (iii) and (iv)). The SDF's were calculated at 90% above bulk water probability for both sugars as a function of the linkage atoms C1-O1-C4' and are shown together with the PDF's in Figures 6.5 (i) and (ii).

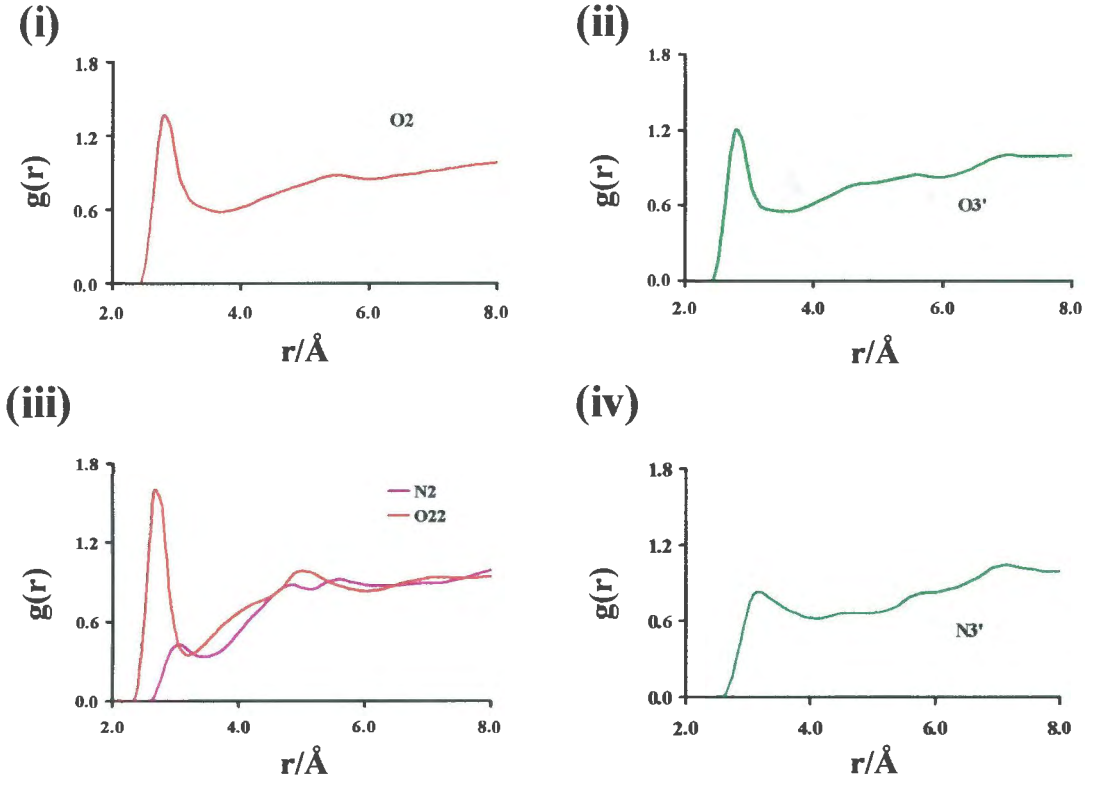


Figure 6.4 **Pair distribution function (PDF) for maltose ((i) O2 to water oxygen atoms, (ii) O3' to water oxygen atoms) and Glc-2-NAc-α-(1→4)-Glc-3-NH_2 ((iii) N2 and O22 to water oxygen atoms, (iv) N3' to water oxygen atoms).**

The combination of the PDF and SDF data allows us to estimate the location of the water probability density about the glycosidic linkages of the two disaccharides relative to the functional groups adjacent to these linkages. This analysis is summarized in Table 1 for both disaccharides. There is a 90% probability of finding a water molecule 3.14Å from the O2 and 3.42Å from the O3' atoms in the maltose solution. The equivalent analysis for GlcNAc-α-(1→4)-3-amine-Glc reveals that a water molecule is most likely to be located 3.19Å from N2 and 3.40Å from N3' and a second water located at 2.79Å from O22 and 3.54Å from N3'. There are therefore two bridging waters in the case of GlcNAc-α-(1→4)-3-amine-Glc with one participating in hydrogen bonds with the N2 (NAc) and N3'(amine) atoms and a second participating in hydrogen bonds with O22 (NAc) and N3'(amine) atoms.

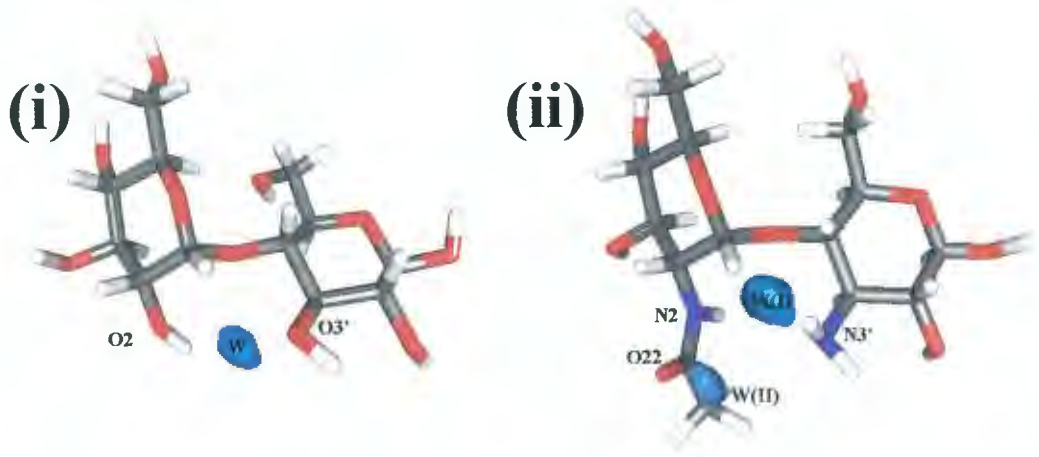


Figure 6.5 Pictorial representation of spatial distribution function (SDF) at 90 % above bulk water probability density for solution containing (i) maltose and (ii)Glc-2-NAc-α-(1→4)-Glc-3-NH_2.

6.3.2.1 Water Survival Rate

The value of function Pj(t,t$_i$;t*) was defined as 1 if the center of mass of a water molecule j is at least within 4.5 Å of each of the oxygen or nitrogen atoms on the functional groups on either side of the glycosidic linkage at both time frames t_i and t_i+t and in the interim does not break the hydrogen bond for any continuous period, longer than t*. Otherwise it takes on a value of 0. These waters are participating in weak to moderate bridging hydrogen bonds between the two functional groups. The parameter t* is introduced to ignore cases when a water molecule only temporarily breaks the bridging hydrogen bond and reforms it without properly entering the bulk solution. A value of t*=2ps, which has been used in studies of hydration shell sizes around ions.[24] Consequently a function that gives a probability for a water molecule participating in a bridging hydrogen bond for time t is defined in equation 6.2 below and plots of p(t) vs. time for both sugar molecules are shown in Figure 6.6.

$$p(t) = \frac{\langle P_j(t, t_i; t^*) \rangle_{i,j}}{\langle P_j(0, t_i; t^*) \rangle_{i,j}} \qquad \text{eq 6.2}$$

After a fast initial fall the p(t) curves decay almost exponentially as exp(-t/T) where T is the residence time. The residence time (T=8.36ps) for water molecules representing a solvent bridge between the amine and NAc groups of the GlcNAc-α-(1→4)-3-amine-Glc disaccharide corresponds to the 90% probability volumes shown in Figure 6.5(ii). The residence time for bridging waters about maltose corresponding to the 90% probability volume shown in Figure 6.5(i) is a factor of 2 lower than in the case of the NAc-amine derivative with T= 3.66ps.

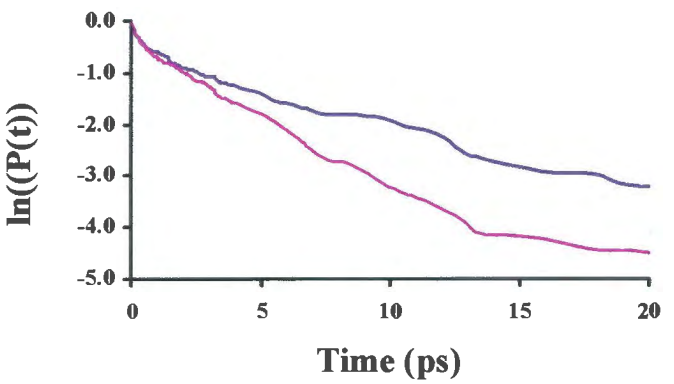


Figure 6.6 **Semi-natural log plot of the survival function (p(t)) of hydrogen bonded bridging waters located between O2 and O3' in maltose and N2/O22 and N3' in Glc-2-NAc-α-(1→4)-Glc-3-NH_2.**

6.3.2.2 Competing Hydrogen Bonds

In MD simulations a frequent exchange between interactions that are less than $5k_bT$ (i.e., 3kcal/mol at T=300K) can be expected if there are no kinetic "bottlenecks". The geometry of the mimic disaccharides were optimised using a density Functional B3LYP/6-31G(d) method. These structures were then compared with a set of micro solvated mimic disaccharides, where single water molecules in close proximity to the functional groups straddling the glycosidic linkage were geometrically optimised at the same level of theory. The structures of disaccharide mimics are shown in Figure 6.7 (i) and (ii) and the mimics complexed with a water molecule are shown in Figure 6.7 (iii)-(vi).

Using a method developed previously[9] the intramolecular hydrogen bond strengths between the functional groups and the multiple intermolecular hydrogen bond strengths between the functional groups and the water molecule were calculated by way of B3LYP/6-31G(d)//6-311++G(d,p). These results are shown in Table 6.1 along with the corresponding hydrogen bond distances and glycosidic dihedral angles.

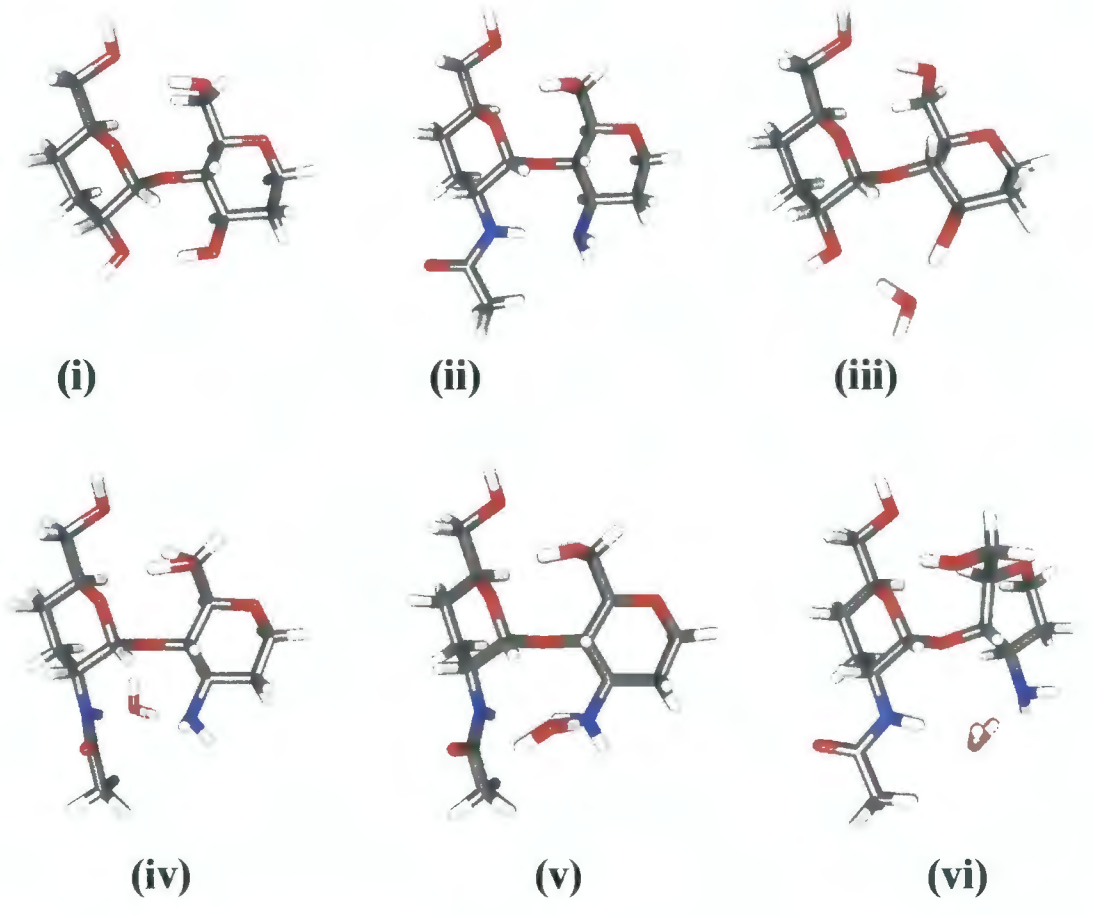


Figure 6.7 B3LYP optimized structures of (i) maltose and (ii) Glc-2-NAc-α-(1→4)-Glc-3-NH_2 mimics and their water complexes. ((iii) water ··· maltose mimic, (iv) water ··· Glc-2-NAc-α-(1→4)-Glc-3-NH_2 mimics (orientation A) (v) water ··· Glc-2-NAc-α-(1→4)-Glc-3-NH_2 mimics (orientation B) and (iv) water ··· Glc-2-NAc-α-(1→4)-Glc-3-NH_2 mimics (orientation C)

It is important to point out that that the sum of the hydrogen bonds for the individual interactions with water produces a greater binding strength compared with the binding energy calculated including BSSE. The lower value calculated from the binding energy is due to the inclusion of steric repulsions occurring in the water-saccharide complex. The magnitude of the hydrogen bonds between water and the hydroxyls of maltose are of the order of 4.5 kcal/mol; this is only marginally weaker than the hydroxyl-hydroxyl hydrogen bond strength that was previously estimated to be 4.99 kcal/mol[9] and is shown in Table 6.1. Similarly the water -NAc and -NH_2 hydrogen bonds are around 5-6.4 kcal/mol which is only slightly weaker than the -NAc ··· -NH_2 hydrogen bonds that have a magnitude of 6.78 kcal/mol. The continual exchange of hydrogen bonds between the saccharide functional groups and water molecules in the first hydration shell can therefore be expected.

Table 6.1 **Geometries, binding energies and hydrogen bond strengths for Maltose and Glc-2-NAc-α-(1→4)-Glc-3-NH_2 mimics and the related micro solvated structures shown in Figure 6.7.**

Structure	6.7(i)	6.7(ii)	6.7(iii)	6.7(iv)	6.7(v)	6.7(vi)
ϕ	-18.03	-17.87	-33.23	-35.55	-22.88	-10.63
ψ	-20.62	0.79	-32.90	43.97	8.30	34.68
Inter-residue hydrogen bond distance (Å)						
O2 to O3'	2.9172	–	3.7301	–	–	–
N2 to N3'	–	2.0388	–	4.3943	3.4240	3.4762
O22 toN3'	–	5.1170	–	6.6160	4.0772	3.9149
Bridging water distance (Å)						
O2(N2) to O_w	–	–	2.7753	2.8706	2.8808	4.3040
O22 to O_w	–	–	–	4.8276	4.8160	3.0563
O3'(N3') to O_w	–	–	2.7655	2.7917	2.8233	2.9393
Inter-residue hydrogen bond strength (kcal/mol)						
Inter-residue	-4.99	-6.78	–	–	–	–
	Bridging water hydrogen bond strength (kcal/mol)					
O2 to O_w	–	–	-4.50	–	–	–
O3' to O_w	–	–	-4.46	–	–	–
N2 to O_w	–	–	–	-5.86	-6.14	-6.38
O22 to O_w	–	–	–	–	–	-3.76
N3' to O_w	–	–	–	-5.03	-5.29	–
Total energy	–	–	-8.96	-10.89	-11.43	-10.14
Binding energy (kcal/mol)						
Include BSSE	–	–	-8.4769	-9.2662	-10.0452	-8.8728

6.4 Conclusion

Medium strength hydrogen bonds formed between uncharged functional groups such as hydroxyls, amines and NAc are prevalent in many biopolymers and particularly in the oligosaccharides of glycoproteins. The strongest attractive interaction amongst these is between the NAc and amine groups. These groups when positioned around the glycosidic linkage in oligosaccharides effectively determine the conformational space of the biopolymer. It is therefore conceivable that oligosaccharide shape can be designed and controlled by tuning the hydrogen bond strength of these functional groups.

The studies here tested this hypothesis by comparing solution conformations of maltose and a NAc-amine derivative where the inter functional group hydrogen bond strength had been maximised. The difference between the internal motions of the two disaccharides in the presence of an aqueous

solvent is not significant. Both disaccharides deviate from their vacuum global minima conformations. This is due to the competition between the intermolecular hydrogen bonds formed between water molecules and functional groups on saccharides and the intramolecular hydrogen bonds formed between these functional groups.

The molecular similarity theory states that "a molecule acts like it acts, because it is what it is".[25] Contextually this theory can be applied to this study. Consider the chemical characteristics of hydroxyl groups of water molecules does not vary much from the characteristics of exocyclic hydroxyl groups found within sugars and since water is a small molecule that can readily reorientate itself around large compounds, the water-bridging phenomenon thus occurs instantaneously. Due to the fact that the functional groups substituted were only able to participate in weak inter-residue hydrogen bond, a regular interchange between solvent-solute and solute-solute hydrogen bonds was observed thus resulting in a similar frequency of fluctuation in maltose and its NAc-amine derivative albeit about different conformations. This chapter fully illustrated that increase in inter-residue hydrogen bonding strength does not necessary reduces in motion above the glycosidic linkages.

REFERENCE

(1) (a) Engelsen, S. B.; du Penhoat, C. H.; Perez, S. *J. Phys. Chem.*, **1995**, *99*, 13334. (b) Engelsen, S. B.; Perez, S.; Braccini, I.; Du Penhoat, C. H. *J. Comput. Chem.*, **1995**, *16*, 10969. (c) Hardy, B. J.; Egan, W.; Widmalm, G. *Int. J. Biol. Macromol.*, **1995**, *17*, 149.

(2) (a) Rutherford, T. J.; Homans, S. W. *Biochemistry*, **1994**, *33*, 9606. (b) Naidoo, K. J.; Denysyk, D.; Brady, J. W. *Protein Eng.*, **1997**, *10*, 1249.

(3) Stevensson, B.; Landersjo, C., Widmalm, G.; Maliniak, A. *J. Am. Chem. Soc.*, **2002**, *124*, 5946.

(4) Naidoo, K. J.; Kuttel, M. M. **2001**, *J. Comput. Chem.*, *22*, 445.

(5) (a) Brady, J. W.; Schmidt, R. K. *J. Phys. Chem.*, **1993**, *97*, 958. (b) Ha, S. N.; Madsen, L. J.; Brady, J. W. *Biopolymers*, **1988**, *27*, 1927.

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

(7) Kuttel, M. M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.*, **2002**, *23*, 1236.

(8) (a) Jorgensen, W. L. *J. Phys. Chem.*, **1982**, *77*, 4156. (b) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.*, **1983**, *79*, 926.

(9) Chen, J. Y.-J.; Naidoo, K. J. *J. Phys. Chem. B*, **2002**, 107, 9558.

(10) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.*, **1977**, *34*, 1311.

(11) Tasaki, K.; McDonald, S.; Brady, J. W. *J. Comput. Chem.*, **1993**, *14*, 278.

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

(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.;

Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Peterssin, G. A.; Al-Laham, M. A.; Zakrzewski, V. G.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Goresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Baboul, A. G.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Gill, P. M. W.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Johnson, B. G.; Stefanov, B. B.; Gomperts, R.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A.; Gaussian, Inc.: Pittsburgh PA, **1998**.

(14) Biegler-König, F.; Schönbohm, J.; Bayles, D. *J. Comput. Chem.* **2001**, *22*, 545.

(15) Becke, A. D. *J. Chem. Phys.*, **1993**, *98*, 5648.

(16) Boys, S. F.; Bernardi, F. *Mol. Phys.*, **1970**, *19*, 553.

(17) Takusagawa, F.; Jacobson, R. A. *Acta Cryst.*, **1978**, *B34*, 213.

(18) Gress, M. E.; Jeffrey, G. A. *Acta Cryst.*, **1977**, *B33*, 2490.

(19) Shashkov, A. S.; Lipkind, G. M.; Kochetkov, N. K *Carbohydr. Res.*, . **1986**, *147*, 175.

(20) (a) Stevens, E. S. *J. Am. Chem. Soc.*, **1989**, *111*, 4149. (b) Stevens, E. S. *Biopolymers*, **1992**, *32*, 1571.

(21) (a) Brady, J. W. *J. Am. Chem. Soc.*, **1989**, *111*, 5155. (b) Engelsen, S. B.; Monteirob, C.; Herve de Penhoatb, C.; Perez, S. *Biophysical Chem.*, **2001**, *93*, 103.

(22) (a) Best, R. E.; Jackson, G. E.; Naidoo, K. J. *J. Phys. Chem. B.*, **2001**, *105*, 4742. (b) Best, R. E.; Jackson, G. E.; Naidoo, K. J. *J. Phys. Chem. B.*, **2002**, *106*, *106*, 5091.

(23) (a) Svishchev, I. M.; Kusalik, P. G. *J. Chem. Phys.*, **1993**, *99*, 3049. (b) Kusalik, P. G.; Svishchev, I. M. *Science*, **1994**, *265*, 1219. (c) Lui, Q.; Brady, J. W. *J. Am. Chem. Soc.*, **1996**, *118*, 12276. (d) Vishnyakov, A.; Widmalm, G.; Kowalewski, J.; Laaksonen, A. *J. Am. Chem. Soc.*, **1999**, *121*, 5403.

(24) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Chem. Phys.*, **1983**, *87*, 5071.

(25) Popelier, P. L .A. In *Molecular Similarity in Drug Design*, Dean, P. M. Ed.; Chapman & Hall: London **1995**; p174

Chapter Seven

Molecular Properties Related to the Anomalous Solubility of α-, β- and γ-Cyclodextrin

7.1 Introduction

Cyclodextrins (CDs) are cyclic α(1→4) linked carbohydrate oligomers constructed from glucose units. They are produced by treating starch with an enzyme, cyclodextrin glycosyltransferase (CGTase). CDs have commercial application in many areas, including pharmaceuticals, agricultural chemicals, cosmetics and foods. Almost all applications involve the ability of the cyclodextrin to alter the physical, chemical and biological properties of guest molecules, through the formation of inclusion complexes.

Three of the most important cyclodextrins are α-CD, β-CD, and γ-CD in order of increased size. These cyclic oligosaccharides have cone shaped rings with primary alcohol functional groups at the small base and secondary alcohol functional groups at the large base. The interior cavity is covered with hydrogens linked to C-3 and C-5 and glycosidic oxygens and is therefore largely hydrophobic. The external part of the cone is mostly hydrophilic so that the CD's protect chemically unstable hydrophobic guest molecules from reacting with their environment while being transported through a biologically aqueous media. Because of its cavity size and relative ease of production derivates of β-CD have been used commercially as a drug carriers. An inhibiting property of β-CD has been its relatively low solubility in water compared with α-, and γ-CDs that are eight and twelve times more soluble, respectively. Of the nearly twenty eight thousand papers published (to date) on the chemistry and properties of CDs the overwhelming majority focuses on the carrier properties of this class of molecule.[1] An electronic search in the literature revealed only a few publications devoted to the uncorrelated solubility to mass relationship of α-, β- and γ-CDs.[2,3] However, a thorough understanding of the deviation of β-CD from expected mass predicted solubility in aqueous solutions and the gain in solubility of derivatives such as O-methylated β-CDs is important to the design of experiments for drug preparation.[4]

The solubility of a solute in a solvent depends fundamentally on the attraction and association of the solute molecules with that of the solvent molecules. In general at the molecular level the CD's dissolve in water and in so doing they spread out and become surrounded my water molecules.

The thermodynamic process for dissolving a solid in a solvent requires a release of energy from the crystal lattice in which the individual CDs are bound. The attraction between the various CDs in the crystal represents the lattice energy. The degree to which the CDs are soluble in water depends on the competition between the CD-CD interaction energy, the CD-water interaction energy and the water-water interaction energy. For a solution with mole fraction x of CDs, the CDs can aggregate and thus change their local concentration. The entropy of mixing for this dissolved solute is given as:

$$\Delta S = -k[x \ln x + (1 - x) \ln(1 - x)] \qquad \text{eq 7.1}$$

and the internal energy is

$$\Delta U = \chi x(1 - x) \qquad \text{eq 7.2}$$

Furthermore, we know that ideal and the non-ideal saturated solubility is directly related to the enthalpy of melting,

$$[d \ln m / d (1/T)]_{sat} = -\Delta H_{melting} / R \qquad \text{eq 7.3}$$

$$[d \ln m / d (1/T)]_{sat} [1 + d \ln \gamma_m / d (1/T)]_{sat} = -(\Delta H_{melting} + \Delta H_{mixing})_{sat} / R \qquad \text{eq 7.4}$$

where T and m represent time and molality respectively. The activity coefficient of the non-ideal mixing is represented by γ_m.

These factors affect solubility and solvation of α-CD, β-CD and γ-CD. However, in this thesis we compare the relative solubility and so look for factors that may cause the deviation from linearity of in the solubility vs mw plot. Here a possibility may be an unusual increase in the aggregation of β-CD at similar saturation points in the solubility curve. This would imply that the free energy of attraction between two solvated β-CDs may be greater than the counterpart of α-CD and γ-CDs. It has been previously shown that the free energy of binding between dimers of α-, β- and γ-CDs were calculated from 5ns MD simulations using a Poisson-Boltzmann procedure. There it was shown that the free energy of dimer binding decreases with increasing CD size. It may therefore appear that the molecular origins of the uncorrelated solubility to mass relationship of α-, β- and γ-CDs are more intimately related to the individual CD-solvent relationship. Therefore our present investigation attempts to resolve the unusual CD solvation pattern by addressing only the comparative relation of these cycloamyloses with the surrounding waters as they compete with the water-water interactions.

We performed NMR experiments and MD computer simulations to investigate the dynamics of α-, β- and γ-CDs in aqueous solutions. In particular we measured the translational diffusion coefficients (D_i) of the three CDs using the pulse-field-gradient spin-echo nuclear magnetic resonance (PGSE NMR) experiments. These values were subsequently compared with the D_is calculated from the MD trajectories. Detailed analyses of the translational diffusion provide important information about the motion of CDs in water, which in turn, reflects the complex intermolecular interactions. These interactions are also manifested in the water structure around the CD molecules. Molecular dynamics simulations provide a unique possibility for investigations of different properties at the molecular level related to water structuring. In addition, the internal motion of the CDs can be monitored using time correlation functions. These dynamical processes are strongly affected by the solute-solvent interactions. Finally, the assumptions about cyclodextrin-water interactions based on the analyses of the MD trajectories are used to reveal the reason for the anomalous solubility properties of the three investigated cyclodextrins.

7.2 Simulation and Experimental Conditions

The program CHARMM has been used for all the molecular mechanics computations reported here.[6] A CHARMM-like force field specifically parameterized for carbohydrates[7] was used to model the CDs. The SPCE water model[8] implemented with periodic boundary conditions was used to simulate the cyclodextrins in explicit bulk water. More precisely the α, β and γ-CD's were immersed in solvent boxes of 4040 SPC/E water molecules of average dimensions 49.61 Å, 49.65 Å and 49.69 Å respectively. This corresponds to solution densities of 1.013 g/cm^3 in each case. The MD simulations were run for 500 ps to equilibrate the system and 5ns for data collection using a 1fs integration time step with the chemical bonds involving hydrogens fixed via the constraint algorithm SHAKE[9]. The long range interactions were decreased smoothly to zero between 12 Å and 14 Å employing the CHARMM switching functions applied on a group-by-group basis.[10] All solution simulations were run using a isothermal-isobaric ensemble (NPT) where the pressure and temperature were kept constant (P=1bar and T=300K) using the Langevin Piston method as implemented in CHARMM[11] (refer to chapter 3) In this scheme the pressure is controlled using a Langevin equation of the type where the volume, V, is treated as a dynamical variable, W=500 is the "mass" of the piston with a collision frequency of γ=5ps^{-1} and applied with a random force R(t) that is determined from a Gaussian distribution about a zero mean and variance. Configurations of the molecules were stored at intervals of 0.5 ps in all simulations and analysed over the entire 5ns trajectory.

Translational diffusion measurements were performed at 298 K using the pulse-field-gradient spin-echo nuclear magnetic resonance (PGSE NMR) experiment. The conventional Stejskal-Tanner experiment[12] was extended with an additional gradient pre-pulse. The latter purges all transverse magnetization from the previous FID and thereby equalized the two diffusion encoding gradient pulses. Furthermore, it is followed by an additional weak gradient during the acquisition period that enables spatial resolution along the z-axis. The deviation from linearity of the pulsed-field-gradient (PFG) pulses over the sample volume was characterized as described by Damberg *et al.*[13] The distribution of the gradient strengths was modeled by a simple truncated linear gradient resulting in calibrated conditions for a sample with a known translational diffusion coefficient (1% H_2O in D_2O + 1 mg mL^{-1} $GdCl_3$; D_t = 1.90 × 10^{-9} m^2 s^{-1} at 298 K).[14] The gradient strengths were varied between 0.5 and 14.6 G cm^{-1} in 30 steps. The duration of the PFG pulses was 5 ms and the refocusing delay was 50 ms. Diffusion coefficients of the cyclodextrins (1.0 mM in D_2O) were obtained by the Longitudinal Eddy-Delayed experiment using 10 consecutive measurements for each molecule.[15] The gradient strengths were in the range 0.5 to 28.8 G cm^{-1}. The duration of the PFG pulses were 5 ms and the refocusing delay was 100 ms. In addition, dephasing gradients were applied after the second and fourth 90° pulses as well as after the acquisition period. Subsequently, a non-linear two-parameter fit (A_0 and D_t) was performed of the attenuation of signal intensities. The average HDO

D_t value in the CD samples were 1.85 × 10^{-9} m^2 s^{-1} at 298 K compared with the calibrated HDO D_t value of 1.90 × 10^{-9} m^2 s^{-1} at 298 K reflecting the accuracy of the D_t measurements.

7.3 Results and Discussion

7.3.1 CD Diffusion Properties and Solvent Behaviours

7.3.1.1 Diffusion of Cyclodextrins: Comparison Between NMR and MD

The diffusion rates for each of the three CDs were measured experimentally as described in the previous section and the results are listed in Table 1. The trend shows a clear molecular mass dependency with the heaviest molecule, γ-CD, diffusing the slowest in water and the lightest molecule, α-CD, diffusing the fastest. The diffusion rates obtained from the MD simulations were calculated from the mean square displacement using the Einstein relation (refer to chapter 3).

The simulation values compared well with the experimental results but are approximately 10% greater in magnitude. This difference originates most likely from the water model used in the simulation. The experimental diffusion constant determined for pure water[16] is 2.3 × 10^{-9} m^2 s^{-1} at 298 K while the simulated counterpart obtained using the SPCE force field[8] is 2.5 × 10^{-9} m^2 s^{-1} at 298 K. As with the experimental measurements the translational motion is molar mass dependent and indicates that the CD's are diffusing relatively freely in the solvent. However, on closer inspection a small but significant deviation from this trend can be observed originating from the β-CD that is diffusing slightly slower than predicted from a linear interpolation for the three CD set. The β-CD is therefore experiencing an additional drag as it diffuses through the solvent.

Table 7.1: Diffusion coefficients (x10^{10} m^2/s) of α-, β- and γ-CD measured from NMR and calculated from MD simulations within 95% confidence limits.

Method	α-CD	β-CD	γ -CD
NMR	2.292±0.025	2.168±0.015	2.097±0.017
MD	2.535±0.035	2.400±0.060	2.375±0.070

We analyzed the spatial distribution of the individually diffusing water molecules in the three CD solutions by applying a method detailed by us previously.[17] Using this procedure we are able to display the distribution of solvent diffusion rates found in the α-, β-, and γ-CD solutions (Figure 7.1). The diffusion rates are contoured at 50% and 10% of the diffusion rate of bulk SPCE water. It is clear that water molecules diffuse significantly slower when close to the cyclodextrins. This

corresponds to our earlier findings for other saccharides where the diffusion rates of the water in the first hydration shell are seen to be significantly slower than that of bulk water.[17] Moreover the solvent in the central cavities of both the α- and β-CD are diffusing slower than bulk water.

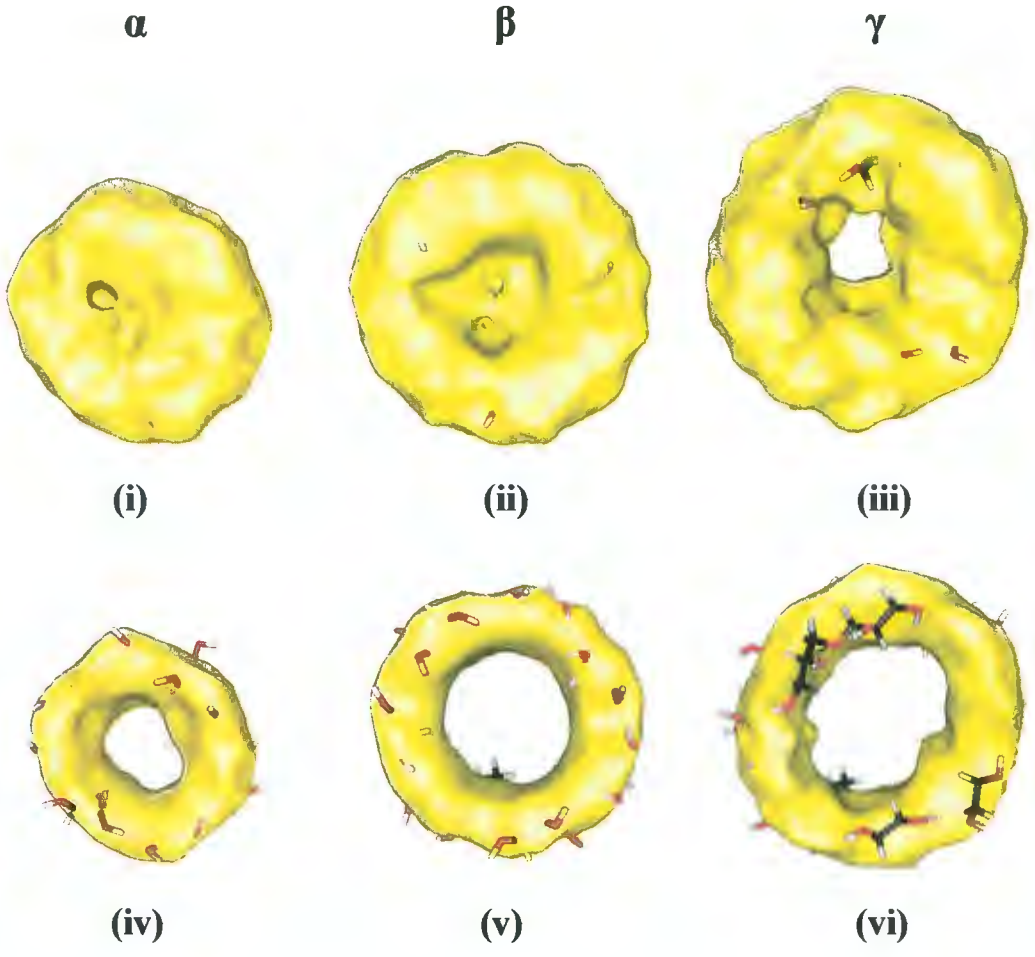


Figure 7.1 Diffusion rate of water around Cyclodextrins α- ((i) and (iv)), β- ((ii) and (v)) and γ-CD ((iii) and (vi)) at 50 % and 10% of the normal diffusion rate (Top view).

7.3.1.2 Solvent Molecules Trapped within a CD Cavity

The cavity volumes of α-, β-, and γ-CDs are known to be 174, 262 and 427 Å^3 respectively.[18] Assuming that the CD cavity volume is spherical, we can determine an effective cavity radius for each CD (Table 2). The centres of mass (CM) of the CDs are located in the middle of their cavity volumes. We studied the water survival rate by defining the $P_j(t,t_i;t^*)$ function (refer to chapter 3) to take the value 1 if the CM of a water molecule j is within the spherical volume of the CD at two time frames t_i and t_i+t and does not leave the cavity in the interim for any continuous period, longer than t^*. If this condition is not met $P_j(t,t_i;t^*)$ takes on a value of 0. The parameter t^* is introduced to

ignore cases when a water molecule only temporarily leaves the CD cavity and returns to it without properly entering the bulk solution. We use a value of $t^*=2$ps, which has been used in studies of hydration shell sizes around ions[19] and in line with our previous calculations involving saccharide solutions.[20] The τ was found to be 1.46ps, 3.85ps and 1.14ps for water in the cavities of α-, β- and γ-CD's respectively. Therefore the 5.35 waters in the β-CD cavity are trapped roughly 3 times longer than the 3.09 and 8.42 waters in the α- and γ-CD cavities. The number of cavity waters obtained for β-CD from these MD simulations corresponds to the number observed in solid state NMR experiments of hydrated β-CDs.[21]

7.3.1.3 Solvent Structure and CD Dynamics

We investigated the structural properties of the waters making up the first hydration shell about the CDs and as a consequence those waters that are located in the CD cavities. The size of the hydration shells was calculated using a Voronoi procedure previously reported by us, where a volume of space is allocated to each water molecule in a saccharide solution.[22] In this method the volume of space is represented by a Voronoi Polyhedra (VP), which is constructed around each solvent molecule. This was done by considering water oxygens within a radius of 4 Å from the hydroxyl oxygens and 4 Å from the methylene hydrogens. The VPs of the waters that are immediate neighbours of the CD Voronoi volume are added up for each frame and then averaged over the entire trajectory to yield the mean number of water molecules that are in the first hydration shell for α-, β- and γ-CDs. The hydration shell analyses were performed over the entire 5ns trajectory (for each CD) with intervals of 0.5 ps between successive frames analyzed. From these calculations we found that α-, β- and γ-CD each had 84, 99, and 113 first hydration shell waters respectively which amounts to approximately 14 water molecules for every glucose monomer in the cycloamylose rings.

The nature of the first shell waters can be analyzed by using the Voronoi Polyhedra (VP) analyses (described in chapter 3) to determine if their collective structure is different from that observed for bulk water. An accurate measure of local water structure and hence cyclodextrin solution structure can be found by calculating an asphericity parameter η for each of the water VPs. We calculated asphericity values for the first hydration shells water VPs of α-, β- and γ-CD solutions to assess the structure of the surrounding water. All three CD solutions have similar asphericity profiles. The relationship between the waters and the hydroxyl (hydrophillic) groups are similar to that found in bulk water with η distributions (mode≅1.64) tailing off at 2.1 in all three cyclodextrin solutions. However, analysis of the VP of the water molecules that are neighboring the hydrophobic CH groups (including the ones found in the ring cavity) reveal that these waters have a significantly anisotropic relationship with their water neighbors. This is evidence by the shift in the η distribution mode≅1.74 toward one that is more characteristic of an anistropic overall structure which tails off at 2.4.The water pockets in the ring cavities of all three cyclodextrin solutions are therefore structurally

different from that of bulk water and tend to be structured more like the solvent does on crystallization in ice I_h.

The three-dimensional water distributions, as measured by the spatial distribution function (SDF), about the cyclodextrins was calculated by a similar method to that used previously.[22,24] All frames of the three trajectories were used in turn for solvent probability density analysis. The volume around the solute was divided into a three-dimensional Cartesian grid with the width of each division about 0.5 Å. Each set of coordinates from the trajectory was translated to the solute centre of mass and rotated by least squares fitting to a reference structure.[25] The value of a normalized gaussian function centred on each water oxygen was used for binning, to achieve a smoother density. Finally, the total oxygen density was normalized by dividing by the number of frames binned and the expected oxygen density per bin for a uniform distribution, such that a density of 1.0 corresponds to the bulk probability density. Contours of 1.5, 1.7 and 1.9 times larger than bulk water probability density were used to illustrate the solvent configurational probability about the three cyclodextrins that is 50% and 90% greater than that observed in the pure solvent. These contour plots showing the cyclodextrins limiting the configurational entropy of the water solvent from that observed in the bulk solvent are shown in Figure 7.2.

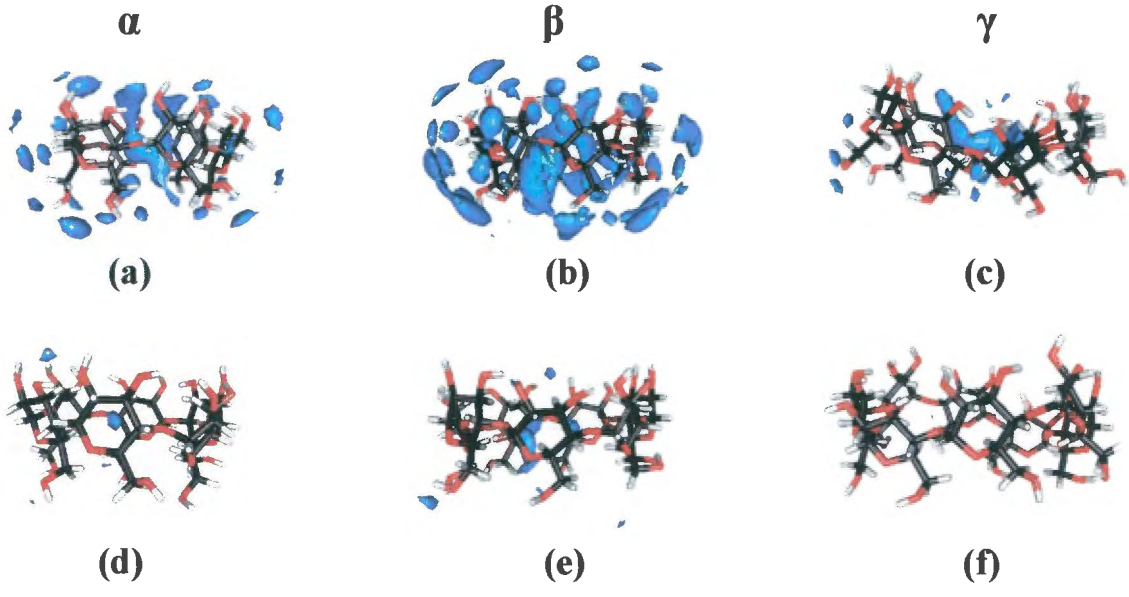


Figure 7.2 Contour plots of water spatial distribution functions corresponding to the probability density 50% and (a, b and c) 90% (d, e and f) above bulk water as calculated from the MD trajectories.

Visual inspection of the SDF's reveal that the probability of water molecules locating themselves around β-CD (Figure 7.2 (c) and (d)) is greater than that for α-CD (Figure 7.2 (a) and (b)) and γ-CD (Figure 7.2 (e) and (f)). The decrease of the β-CD solubility is due to the marked water structuring initiated from water-β-CD interaction. This reasoning has been previously put forward by various researchers.[26] However, their studies involve using insufficient solvent numbers or/and unsatisfactory cutoff distance (see chapter 3 for graphical representation of problems associate with cutoff distances). Moreover, not only didn't they provide comprehensible representations of reasons regarding to water-CD solvent orderliness, some of these studies dealt exclusively with β-CD only and neglected any comparisons with other cyclodextrins.

Calculating the volume of the contours associated with above bulk water probability densities for the three cyclodextrin solutions gives a more accurate assessment of the total fraction of waters that are localised about the three cyclodextrins compared with waters that are located in the bulk solvent. The results of these calculations are graphically illustrated in Figure 7.3. Here a strong correlation between solubility and the extent to which individual CD's interfere with the water molecule locations in solution emerges. The result is an alteration of the water structuring as seen in the bulk solvent. The least soluble compound, β-CD, induces water to locate around it the most, this is followed by α-CD which has a slightly lower structuring effect on the solvent and finally γ-CD which structures the water significantly lower than either of the other two CD's and is by far the most soluble cyclodextrin in aqueous solution. This result is consistent with the number of waters found on crystallization of β-CD (8.3-9.14/CD), α-CD (5.5/CD) and γ-CD (5.86-12.75/CD).[18]

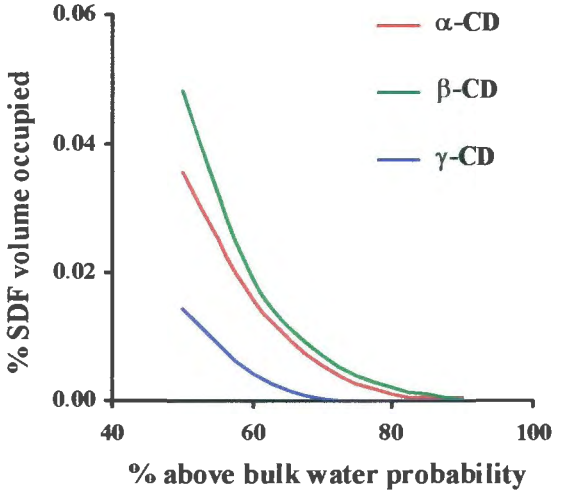


Figure 7.3 **The volume fraction (x10⁴) occupied by the structured water, derived from the SDFs for the CDs.**

7.3.2 Conformational properties of individual CD

Our simulations account for the properties of CDs in dilute solutions and so far the solvation of the CDs appear to be more related to the way in which they affect the water structure rather than their relative molecular weight and macromolecular ring size. The rational for the lowest solubility observed for β-CD is due to the ice-like water structure that is trapped within its central cavity for a significantly longer period than that occurring for the other two carbohydrates. The water structure about β-CD is therefore disturbed to a greater extent than in the case of the other two cycloamyloses. In addition the solvent spatial probability density calculations reveal that the first shell waters surrounding this CD are more highly localized than their counterparts in α- and γ-CD solutions. Since the relative chemical structure and compositions of the three CDs are no different from each other with each displaying a similar ratio of hydrophobic (methylene groups) to hydrophilic (hydroxyl groups) the differential solute solvent interactions of these three cyclic sugars must be due to differences in their conformational behaviour.

7.3.2.1 Internal Motion of the CDs

Although on the basis of X-ray early studies assumed cyclodextrins to be rigid molecules this notion has been discounted and CDs are accepted to be flexible[3] displaying significant rotational motion of the individual glucose rings as well as undergoing distortions of the macrocyclic rings.[27] We probe this further, in relation to the unusual effect that β-CD has compared with α-, and γ-CD on the surrounding water, by calculating the relative internal (conformational) motion of the cyclodextrins in solution. The orientational and translational motions of the CDs were removed leaving only the conformational motion of each CD over each of the 5ns data sets. The contribution of internal motion toward molecular relaxation was then estimated by calculating the root mean square (RMS) fit autocorrelation function (Equation 5), where $RMS(t_a,t_b)$ is the best RMS fit obtained between structures at times t_a and t_b using a standard fitting procedure.[25] (refer to Chapter 3)

Of particular interest to us is the ring motion of α-, β-, and γ-CD in water. This was calculated from the relative distortion of the cycloamylose ring as observed from the glycosidic oxygens between different time frames (t_a and t_b). An interval of 0.05 ps was used between time origins with each frame separated by 0.05 ps therefore C_{RMS} values for consecutive frames will result in very small RMS differences and the $C_{RMS}(t)$ curve is bounded by the maximum RMSD of the glycosidic oxygens. The plots for all three CDs are shown in shown in Figure 7.4. The relaxation rate of the macrocyclic ring motion is taken from the point where the functions at plateau to the maximum RMSD. This analysis shows that β-CD undergoes the least macrocyclic ring motion and relaxes the fastest (0.5 Å, 210 ps) followed by α-CD (0.6 Å, 502 ps) with γ-CD (0.9 Å, 840 ps) undergoing the greatest amount of ring distortion and relaxing the slowest. Although the difference in the extent of RMSD macrocyclic ring motion (0.5 Å for β-CD c.f. 0.9 Å for γ-CD) between the three CDs the

hydrophobic cavity and first hydration shell compared with the comparatively flexible γ-CD, which interferes the least with the surrounding solvent structure and diffusion.

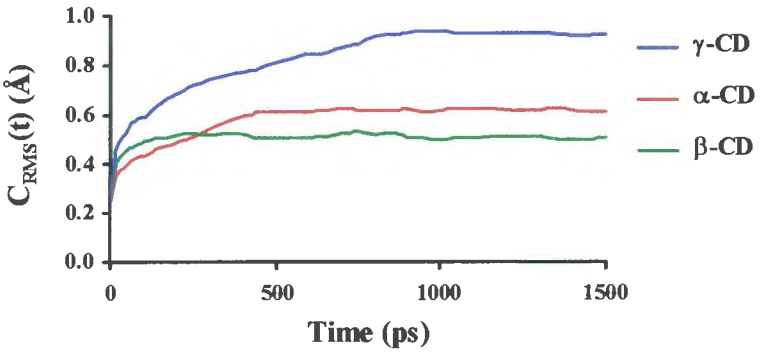


Figure 7.4 Auto-correlation functions of the RMS fit calculated for the three CDs.

Over the years, the conformational flexibility of cycldextrins has been the subject of intensive discussion without establishing a conclusive verification on the origin of the β-CD's anomalous aqueous solubility.[26,28-31] The comprehensive assessments of CDs' dynamical behaviours have appeared in a series of publications documented by Koehler, Saenger and van Gunsteren dated as early as in the late 1980s.[32] These publications addressed the characteristics of asymmetry motion of the macrocyles, mobility of the water molecules and the distribution of a variety of CD structural geometries. Recently, other publications concerning the how the hydration effects on the conformational properties of the CDs can also be found. However, it is to our knowledge, the fundamental molecular motion of the CDs has not been properly described. Furthermore, some of these publications attributed the solubility trend of these three CDs as a direct result of the differences between their inter-glycosidic hydrogen bonding networks participated between their C2-C3' hydroxyl groups. This phenomenon is referred as hydrogen bonding belt. For the remaining part of this chapter, we aim to carry out a detailed investigation of the factors influencing the contrasts between the internal motions of three cyclodextrins.

Traditionally studies are directed towards examining the fluctuation of the ϕ, ψ glycosidic angle and torsional angle pairs on the cyclodextrins.[30] However, such type of analyses can not successfully distinguish the intricate motional fluctuation behaviour within a cyclodextrin. Moreover, the scatter plots for the different ϕ, ψ glycosidic torsional angle pair in α-, β-, and γ-CD have been examined by Kuttel. But no definite comparative trend was established due to the fact that all three scatter plots show analogous conformational sample space.[33]

After performing the Pucker analysis, it was found that the glucopyranose monomer units essentially remain in their preferred $^4C_1$ chair conformations. Thus the factors influencing the internal motion of a cyclodextrin was deemed to be essentially dependent on three different types of internal

After performing the Pucker analysis, it was found that the glucopyranose monomer units essentially remain in their preferred $^4C_1$ chair conformations. Thus the factors influencing the internal motion of a cyclodextrin was deemed to be essentially dependent on three different types of internal motions: (1) rotation of the exo-cyclic groups, (2) macro-cyclic ring distortion and (3) glucopyranose monomer tilting. Since the exo-cyclic groups found on each cyclodextrin are identical, they should essentially rotate in the similar way. Therefore, the drastic contrasts between the RMS fit amongst the three cyclodextrins should be due to their differences in the latter two contributing factors. We found that probing underlying molecular motion of the CDs could be more adequately achieved by describing the motions according to the schematic representations found in Figure 7.5.

The macro-cyclic ring distortion of the cyclodextrin was inspected by exploring behaviour of the "macro-cyclic ring pseudo-dihedral angles" (denoted as ω) over the whole dynamics trajectory. This pseudo-dihedral term is made of 4 consecutive glycosidic oxygen atoms found on the cyclodextrin ring. The graphical illustration of ω is shown in Figure 7.5(i) (where O_a, O_b O_c, and O_d, are symbols representing the label of 4 successive glycosidic oxygen atoms). According to such a scheme, if a cyclic molecule resembles a perfectly planar polygon ring, then all of its pseudo-dihedral angles should be equal or close to zero and the degree of deviation indicates the level of ring distortion. Each ω of the CD was analysed and binned with 5 degree increment interval. Summing up all possible binned grids over the 5ns MD trajectory, the degree of macro-cyclic distortion of each individual CD was revealed and the differences between the ring distortions amongst the three CDs can be examined in Figure 7.6.

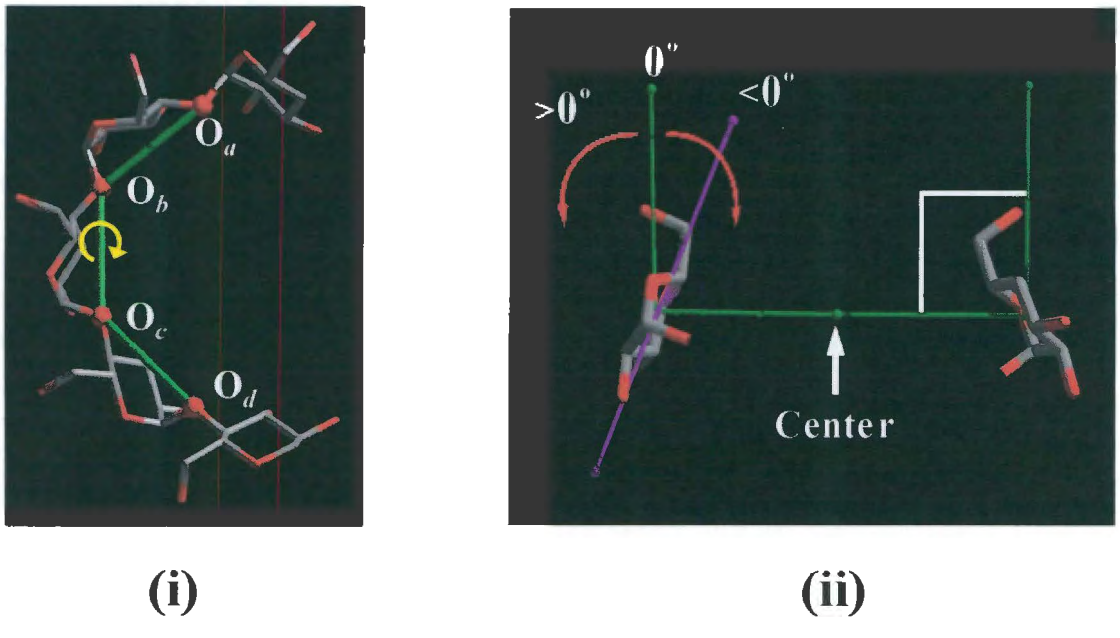


Figure 7.5 Defining the two major modes of internal motions found within a cyclodextrin: (i) macro-cyclic pseudo-dihedrals distortion and (ii) glucopyranose monomer tilting

For the investigation of the glucopyranose monomer tilting behaviour, the tilted angle (τ) of each glucopyranose monomer was analysed and binned with 5 degree increment interval according to the scheme defined in Figure 7.5(ii). The titling angle distribution of every glucopyranose unit of the CDs can be examined in Figure 7.7. The curves denoted A1-6, B1-7 and G1-8, illustrated the tilting behaviour of each monomer for α-, β- and γ-CD respectively. If the cavity of the cyclodextrin is cylindrical, then according to this scheme, the tilt angle should be equal to 0 degree. Should the cavity of the cyclodextrin adapt the torus shape with the primary hydroxymethyl group pointing into the cavity, the tilt angle would be negative. On the contrary, a positive tilt angle reveals the ring opening behaviour. The fluctuation of the glucopyranose monomer tilting should display the flexible opening/closing motion of the cavity.

7.3.2.2 Macro-cyclic Ring Distortion

The molecular mechanics investigation of Lipkowitz *et al.* suggested that all three CDs are somewhat rigid in gas phase but solvation can induce ring certain level of distortion.[29] Similarly, the recent review article by Saenger and co-workers stated that the larger CD has more distorted cycloamylose rings.[30] However, according to the computational work performed by Dauchez *et al.*, only the cycloamylose ring of β-CD remains rigid and upholds a perfect symmetry when immersed in solution while the cyclo-amylose ring α- and γ-CD become more distorted.[31] They claimed that this macro-cylic ring distortion phenomenon contributes to the solubility difference amongst these three CDs. Nevertheless, it is to our knowledge that these references do not draw any conclusive remarks concerning the direct comparison between the ring distortion differences between these three CDs.

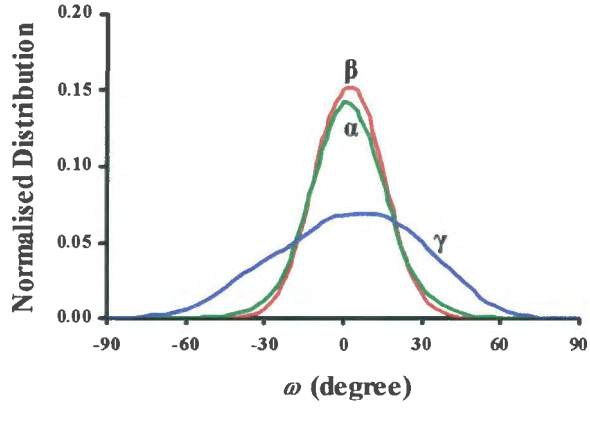


Figure 7.6 The normalised distribution of macro-cyclic ring pseudo-dihedral angles (ω) of α-, β- and γ-CD.

From the above figure, it was found that α- and β-CD cycloamylose rings exhibit similar flexibility to one another. Even though both α- and β-CD rings reveal a fair degree of ring flexibility, the distortion of their cycloamylose rings is much more restricted in comparison to γ-CD. The mean value, the standard deviation fluctuations ($SD_{(Fluc)}$) and the time correlation function of the distortion motion of each macro-cyclic ring pseudo-dihedral angle were calculated and listed in table 7.2.

Table 7.2 The ring distortion behaviour of α-, β- and γ-CDs

	Macro-ring dihedral angles								
	Average	1	2	3	4	5	6	7	8
α-CD									
Mean / o	-0.12	-2.86	4.02	-3.36	0.81	0.29	0.38	--	--
SD(Fluc) / o	13.82	14.21	14.44	12.22	13.76	13.56	14.71	--	--
TCF / ps	0.0199	0.0139	0.0147	0.0149	0.0133	0.0227	0.0401	--	--
β-CD									
Mean / o	-0.04	2.17	-3.15	3.68	-1.55	-2.93	3.25	-1.75	--
SD(Fluc) / o	14.83	13.10	15.92	14.34	16.50	13.16	16.07	14.70	--
TCF / ps	0.0197	0.0286	0.0111	0.0476	0.0291	0.0149	0.0152	0.0414	--
γ-CD									
Mean / o	3.72	13.68	6.37	6.93	-8.19	8.83	-6.80	9.16	-0.26
SD(Fluc) / o	26.48	19.37	28.11	20.28	32.50	27.06	30.02	24.60	29.87
TCF / ps	0.0033	0.0123	0.0036	0.0020	0.0026	0.0041	0.0031	0.0070	0.0030

Base on the results stated in Table 7.2, it is apparent that both α- and β-CD rings exhibit a higher coplanarity since the mean values of their pseudo-dihedral angles do not deviate far from zero. On the contrary, the wide distribution of γ-CD's pseudo-dihedral angles (-8.19 to 13.68) clearly indicates that the eight-membered cycloamylose ring has a high degree of flopping tendency. In the condensed phase simulation, all three cyclodextrins are not as rigid as it has been previously claimed to be. The extent of the $SD_{(Fluc)}$ is related to the magnitude of the flexibility of the macro-ring. The γ-CD ring clearly reveal a higher degree of flexibility capability since its $SD_{(Fluc)}$ values are about twice as large in comparison with both α- and β-CD. The time correlation functions of each macro-cyclic ring pseudo-dihedral angles support the above statements as it is apparent that two of the ωs of β-CD showed a relatively higher degree of rigidity. Furthermore, the distortion motion of γ-CD of the ωs are far more rapid than the other two cyclodextrins

7.3.2.3 Monomer Tilting

According to the x-ray structure of α-CD hexahydrate, the results indicated that two glucopyranose units have their hydroxymethyl side inclined towards the cycloamylose cavity. The average tilting angle of the glucopyranose monomer unit for α-, β- and γ-CD are -13±10, -14±10 and -19±9 degrees respectively.[33] Rees, as well as, Schonberger *et al.* had previously discussed such tilting motion by analysing the inter-residue hydrogen bond belt of the secondary alcohol groups.[34,35] They argued that since one of the glucopyranose units preferred to tilt in a distorted position, the hydrogen bond belt is broken. Thus under such circumstance only four out of all possible six hydrogen bonds can be established for α-CD. They believed such phenomenon in α-CD enhanced its solubility. Similarly, Gillet *et al.* studied the β-CD and regarded the low aqueous solubility of β-CD as a result of a belt of hydrogen bond occurring in the low rim of cyclodextrin. However, these findings are based either solely on the static cyclodextrin geometries or using hydrogen bonding phenomenon to predict the rigidity differences between the cyclodextrins, none of the above studies have addressed the issue adequately by showing the significant differences between the vibrant tilting motion of the glucopyranose units amongst different cyclodextrins. From the dynamics simulations, it was found that in contrast to γ-CD which relies on distorting the coplanar symmetry of the cycloamylose ring as a means of increasing its solubility, α-CD relies on the rotational motion of its glucopyranose units to enhance its solubility.

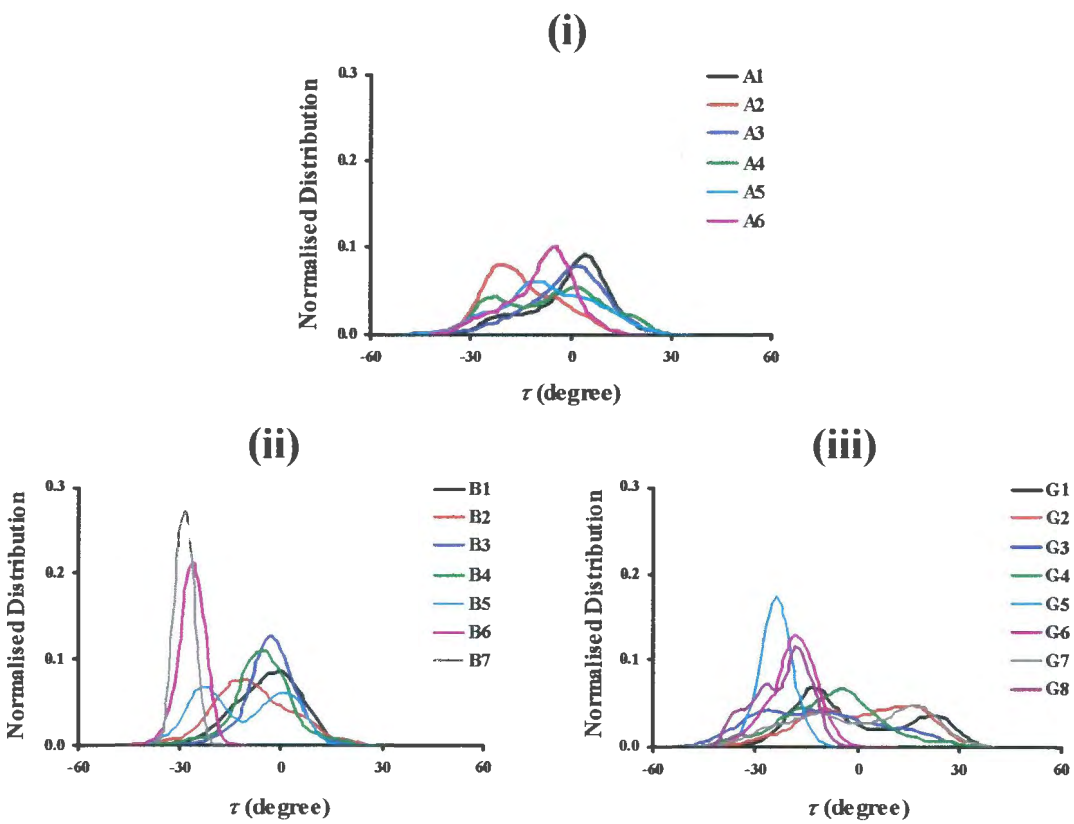


Figure 7.7 The normalised distribution of each individual glucopyranose monomer tilting angle ((i) α-, (ii) β- and (iii) γ-CD)

Figure 7.7 clearly reveals that all three cyclodextrins adapt the torus shape considering the majority of the τ populated over the negative sector. In additional to this, Figure 7.7(i) shows each glucopyranose units of the α-CD can undergo a wide tilting motion. Likewise, most of the glucopyranose of γ-CD exhibits a fairly broad range titling motion (Figure 7.9(iii)). In contrast, evidently, the monomer unit 6 and 7 of β-CD tend to interfere with each other and bring about more hindrance in their each other's tilting motion (Figure 7.9(ii)).

Using the potential mean force calculations, Kuttel *et al.* demonstrated that there is a significant shift of the **ϕ,ψ** glycosidic dihedral angles when comparing the global minimum conformations of maltose found in vacuum (**ϕ,ψ**; -22.5, -25.0) and in aqueous condense phase (**ϕ,ψ**; -47.5, -35.0).[36] The corresponding energy minimum structures are shown in Figure 7.8((i), (ii) (iv) and (v)). The effect of the bridging water phenomenon is noticeable from their study. The schematics (Figure 7.8(iii) and (vi)) illustrate the contrast of the individual glucopyranose position in relation with each other as the result of different environmental influences. It is apparent that the two glucopyranose residues of maltose, an α-(1→4)-linked disaccharides, preferred to position themselves in a slant angle with respect to each other. We postulate that since both α- and γ-CD have even number of monomer residues, this inter-residue arrangement is geometrically favourable. However, β-CD has an odd amount of glucopyranose monomers thus two of its monomer residues generally ended up hindering each other's and thus can not easily adapt this favourable conformation. (These findings can be further extrapolated to explain the solubility of δ-CD (not shown here). The solubility of the δ-CD has been reported to be higher than β-CD but lower than both α- and γ-CD.[37] Since δ-CD comprises of 9 individual glucopyranose monomers, once again, two of the residues would experience the same problem as was observed in β-CD. However, since a larger CD ring induces a higher degree of macro-cyclic twisting motion, δ-CD is therefore more soluble than β-CD).

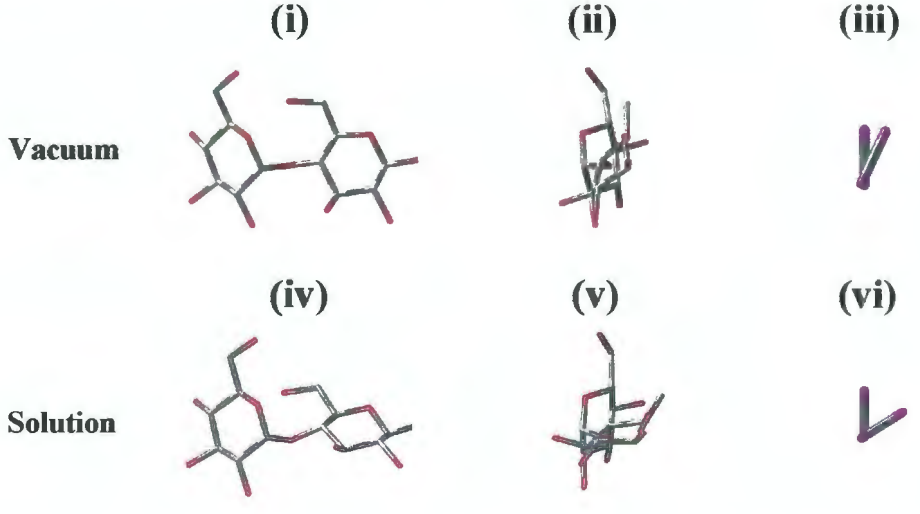


Figure 7.8 Comparing the minimum energy conformations of maltose ((i) and (ii) represent the vacuum global minimum and (iv) and (v) represent the solvated global minimum. (iii) and (vi) illustrated the contrast between their relative glucopyranose positions.

It is important to mention that even though it can be argued that since all the glucopyranose units should experience the exact identical condition within a particular cyclodextrin, hence the titling motion of each monomer unit found within the ring should be indistinguishable to one another. The inequality of their tilting motion suggested that the dynamics trajectory is not sufficiently long enough. However, alternatively it can also be rationalized that after 5ns of simulation time, β-CD remains less flexible when compared with both α- and γ-CD and this argument corresponds to the anomalous low aqueous solubility of the β-CD.

The x-ray studies of the inter-residue distances between the O2(n) to O3(n-1) of all three cyclodextrins for α-, β- and γ-CD are as following, 3.0265±0.1235Å, 2.8995±0.0885Å and 2.8380±0.0730Å.[32] It is worth-noting that α-CD holds a much larger fluctuation range of the estimated inter-glycosidic oxygen atom distance. Since the bond length of each glycosidic bonds of α-CD does not unexpectedly elongate, this unexpected wide range distribution of α-CD's O2(n) to O3(n-1) distancing should primarily brought about by the tilting motion of the monomer. This coincides with the above investigation which clearly indicates that the glucopyranose residues of the α-CD can easily undertake this type of motion.

7.3.2.4 Influence of Hydrogen Bonding on Ring Flexibility

From the geometrical stand-point, after examined the x-ray findings of the inter-residue distances between the O2(n) to O3(n-1) of all three cyclodextrins mentioned in the above paragraph, it is apparent that γ-CD is more suitable for forming cross-glycosidic hydrogen bonding. However, this is not the case when one investigate the hydrogen-deuterium-exchange rate of these CDs within the DMSO solvent. The 1H NMR chemical shifts of the secondary hydroxyl protons of α, β and γ-CD have been compared. It was found that the hydrogen bonding phenomenon predominates on secondary hydroxyl groups of the CDs and the interaction is particularly stronger within the β-CD than other α-CD and amylose.[37].

Based on the findings observed from chapter 6, an increase in hydrogen bonding strength does not bring about a decrease of cross-glycosidic motion. Moreover, it has been documented that most of the cyclodextrin crystalline structures are often trapped in a higher energy state and the hydration tends to distort these relatively flexible cyclic-amyloses into a more stable conformation.[27] Therefore, we suspect that the stronger hydrogen bonding of β-CD detected by NMR experiment is the result of β-CD's macro-cyclic inflexibility rather than being the factor which causes β-CD to be anomalously rigid. This issue will be investigated in detail in the next chapter.

7.4 Conclusion

The diffusion rates of the CDs obtained from the pulse-field-gradient spin-echo nuclear magnetic resonance experiments show good agreement with molecular dynamics computer simulations diffusion rate. On closer examination it is seen that β-CD diffuses slower than expected from a linear trend for the three CDs. Analyses of the hydration shells indicate that water in the first hydration shell is more spatially localized. Furthermore the waters within the ring cavity of β-CD diffuse more slowly than the cavity waters of α- and γ-CD and are trapped for periods of up to 3.5ps. The result is that β-CD drags five or six cavity waters along with it in solution and therefore diffuses slightly slower than expected for its size and weight. The relatively constrained macrocyclic ring motions of β-CD compared with that of the other two CDs is the reason for its greater interference in the water structure. This relatively "rigid" sugar ring is therefore prevented from being as easily accommodated into the overall water structure compared with α-, and γ-CDs. Remarkably the increase in flexibility from β-CD to α-CD and finally to γ-CD is consistent with the solubility pattern of the three cyclic saccharides. After further investigation of the differences of the internal motions between the three CDs, it was found that γ-CD relies on distorting its cyclo-amylose rings for relieving its internal strain whereas α-CD relies on tilting the glucose monomers motion. The β-CD shows the least degree of flexibility in both types of internal motions. Hence results in its low aqueous solubility.

REFERENCES

(1) D'Souza, V. T.; Lipkowitz, K. B., Eds. *Special Thematic Issue: Cyclodextrins*; American Chemical Society: Washington, 1998; Vol. 98.

(2) (a) Jozwiakowski, M. J.; Connors, K. A. *Carbohydr. Res* **1985**, *143*, 51; (b) Starikov, E. B.; Brasicke, K.; Knapp, E. W.; Saenger, W. *Chem. Phys. Lett.* **2001**, *336*, 504; (c) Manunzaa, B.; Deianaa, S.; Pintorea, M.; Gessab, C. *J. Mol. Struct. (Theochem)* **1997**, *419*, 133.

(3) Dodziuk, H. *J. Mol. Struct.* **2002**, *614*, 33.

(4) Stezowski, J. J.; Parker, W.; Hilgenkamp, S.; Gdaniec, M. *J. Am. Chem. Soc.* **2001**, *123*, 3919.

(5) Bonnet, P.; Jaime, C.; Morin-Allory, L. *J. Org. Chem.* **2002**, *67*, 8602.

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

(7) Kuttel, M. M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236.

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

(9) van Gunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34*, 1311.

(10) Tasaki, K.; McDonald, S.; Brady, J. W. *J. Comput. Chem.* **1993**, *14*, 278.

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

(12) Stejskal, E. O.; Tanner, J. E. *J. Phys. Chem.* **1965**, *42*, 288.

(13) Damberg, P.; Jarvet, J.; Gräslund, A. *J. Magn. Reson.* **2001**, *148*, 343.

(14) Mills, R. *J. Phys. Chem* **1973**, *77*, 685.

(15) Gibbs, S. J.; Johnson, C. S. *J. Magn. Reson.* **1991**, *93*, 395.

(16) Krynicki, K.; Green, C. D.; Sawyer, D. W. *Faraday Discuss. Chem. Soc.* **1978**, *66*, 199.

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

(18) Szejtli, J. *Chem. Rev.* **1998**, *98*, 1743.

(19) Impey, R. W.; Madden, P. A.; McDonald, I. R. *J. Chem. Phys.* **1983**, *87*, 5071.

(20) Vishnyakov, A.; Laaksonen, A.; Widmalm, G. *J. Mol. Graphic and Modell* **2001**, *19*, 338; (b)Naidoo, K. J.; Chen, J. Y.-J. *Mol. Phys.* **2003**, *101*, 2687.

(21) Usha, M. G.; Wittebort, R. J. *J. Am. Chem. Soc* **1992**, *114*, 1541.

(22) Naidoo, K. J.; Kuttel, M. M. *J. Comput. Chem.* **2001**, *22*, 445.

(23) Ruocco, G.; Sampoli, M.; Vallauri, R. *J. Chem. Phys.* **1992**, *96*, 3857.

(24) Lui, Q.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 12276.

(25) Kabsch, W. *Acta Cryst.* **1976**, *A32*, 922; (b)Kabsch, W. *Acta Cryst.* **1978**, *A34*, 827.

(26) (a) Linert, W.; Margl, P.; Renz, F.; *J. Chem. Phys.*, **1992**, *161*, 327-338 (b) Momany, F. A.; Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 210

(27) Lipkowitz, K. B. *J. Org. Chem.* **1991**, *56*, 6357.

(28) Jaime, C. Cyclodextrins. In *Encyclopedia of Computational Chemistry*, 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, 1998; p644.

(29) Lipkowitz, K. B. *Chem. Rev.*, **1998**, 98, 1829

(30) (a) Saenger, W.; Jacob, J.; Steiner, T.; Hoffmann, D.; Sabe, H.; Koizumi, K.; Simth, S. M.; Takaha, T. *Chem. Rev.* **1998**, 1787. (b) Harata, K, *Chem. Rev.* **1998**, 98, 1803.

(31) Dauchez, D.; Vergoten, G. In *Minutes of the 5th Interantional Symposium on Cyclodextrins. Paris*, **1990**, Ed. D. Duchene, Ed. De Santé, Paris, **1990**, 101

(32) (a) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *Eur. Biophys. J.* **1987**, 15, 197. (b) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *Eur. Biophys. J.* **1987**, 15, 211. (c) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *Eur. Biophys. J.* **1988**, 16, 153. (d) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *Eur. Biophys. J.* **1988**, 20, 241. (e) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *Biomol. Struc. Dynam.* **1988**, 6, 181.

(33) Kuttel, M. M. *Developing Analytical Tools for Saccharides in Condensed Phases, MSc thesis University of Cape Town* **1999**

(34) (a) Rees, D. A. *J. Chem. Soc.* **1970**, 877. (b) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *J. Mol. Biol.*, **1988**, 203, 241.

(35) Schanberger, B. P.; Jansen, A. C. A.; Janssen, L. H. M. In *Minutes of the 4th Interantional Symposium on Cyclodextrins. Paris, 1988* Ed. Huber, O.; Szejtli, J. Kluwer, Dordrecht, **1988**, 61.

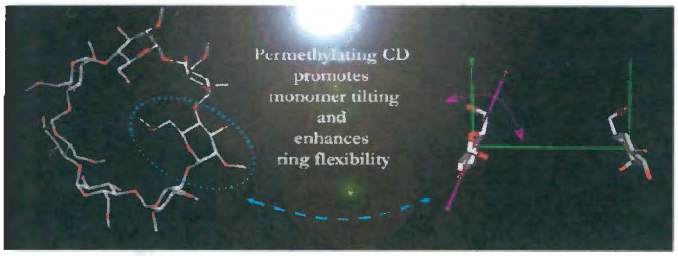
(36) M. Kuttel, Ph.D thesis, University of Cape Town, **2003**

(37) Miyazawa, I.; Endo, T.; Udea, H.; Nagai, T. *In Proceedings of the 7th Interantional Symposium on Cyclodextrins, Tokyo, 1994*, ed. T. Osa, Publ. Office Acad. Soc. Japan, Tokyo, **1994**, p214

(38) Casu, B.; Gallo, G. G.; Reggiani, M.; Vigevani, A. *J. Chem. Soc Spec. Publ.*, **1968**, *23*, 4015.

Chapter Eight

Non-polar substituents derivatisations inducing molecular flexibility of β-cyclodextrins in aqueous phase



8.1 Introduction

A major field of application of cyclodextrins (CDs) is controlled drug delivery. Pharmaceutical drugs are relatively small in size generally and β-CD is often found to be a suitable complex host due to the optimal size of its cavity. The solubility of numerous poorly soluble drugs can be enhanced several fold once complexed with β-CD. However, one of the key factors that presides over the applicability of CDs relies upon their individual intrinsic aqueous solubility and since the aqueous solubility of β-CD itself is inefficiently low, thus it would be advantageous to improve the solubility of β-CD by modifying the necessary functional groups. Through simple chemical modification, the natural CDs are useful templates for producing a wide range of cyclic-amylose derivatives. CD derivatives can be tailored to meet specific requirements of various molecular recognition properties and these molecules often have lower toxicity affinities. In general, incorporating an extended range of functional groups can greatly expand the utility of CDs and subsequently open the gateway to diverse new areas of chemical applications.[1]

Up to this point, this work has illustrated that a reduction of the internal motion of carbohydrates would unquestionably lead to a decrease in their aqueous solubility. Moreover, when the saccharides are chemically modified for the sake of increasing their intermolecular hydrogen bonding strength, these procedures do not necessarily bring about a reduction in their internal motions. There are several other ways one can tamper with the inter-glycosidic motion of a carbohydrate, one of which is to increase the hydrophobicity and the steric effect of the inter-glycosidic groups. This chapter aims to address the influence of these two factors on the overall conformational behaviour of the carbohydrate. Furthermore, this chapter is also dedicated to propose rational explanations regarding the effect of the non-polar inter-glycosidic interaction on the solubilities of CDs.

Four chemically modified β-CD derivatives were selected for aqueous condensed phase simulation. For convenience, these four derivatives, heptakis(2,6-di-*O*-methyl)-, heptakis(2,3,6-tri-*O*-methyl)-, heptakis(2-deoxy-2-hydro)- heptakis(2-deoxy-2-methyl)-β-CD, are denoted as DIMEB, TRIMEB, 2D2HB and 2D2MB respectively. Their motional properties were examined and compared with their seven-membered native precursor. Amongst these four β-CD derivatives, DIMEB and TRIMEB are commercially available and are both frequently used in pharmaceutical applications, while 2D2HB and 2D2MB were selected due to the differences in their steric hindrance ability. The differences between the monomers structures of these CDs are illustrated in Figure 8.1. The structural differences of the monomer units between these derivatives (Figure 8.1(ii-v) and the β-CD (Figure 8.1(i)) are highlighted with blue circles.

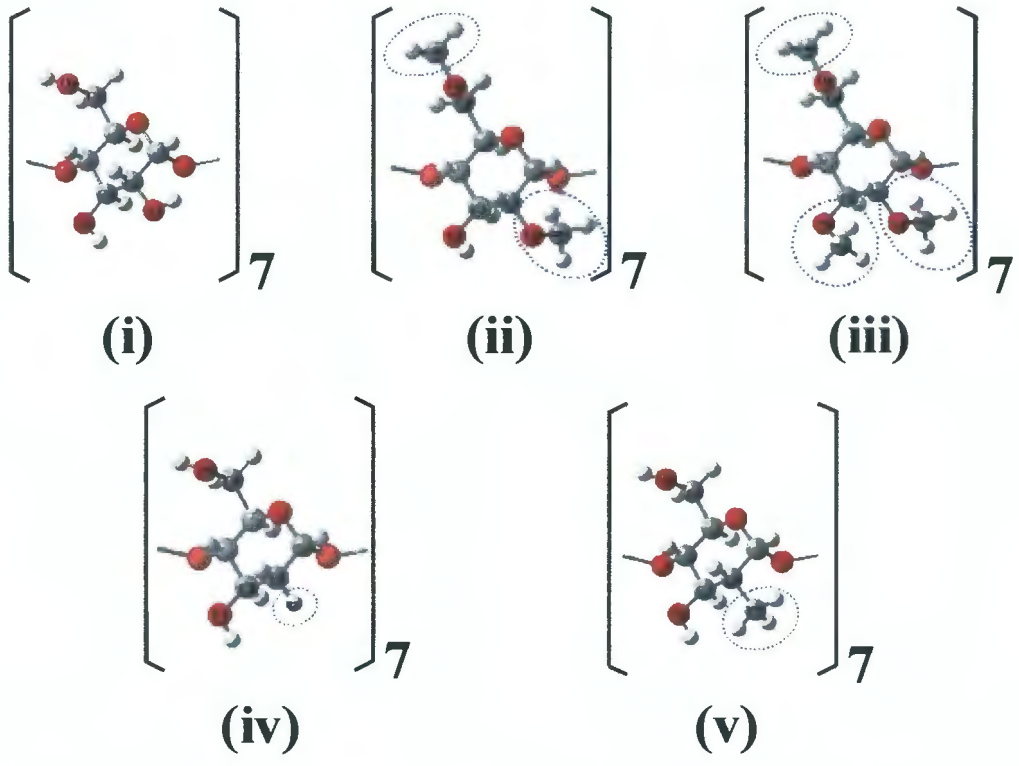


Figure 8.1 The individual monomer structures of (i) β-CD and its derivatives ((ii) DIMEB, (iii) TRIMEB, (iv) 2D2HB and (v) 2D2MB)

Permethylating the β-CD has become a widely used procedure owing to the straightforward modification processes that can subsequently yield two versatile complexing agents known as DIMEB and TRIMEB.[2] The host-guest complexation abilities of DIMEB and TRIMEB in aqueous solution are usually more stable than the corresponding complexes of un-substituted natural cycloamyloses.[3] The differences in the solubilities between the derivatives and their parental β-CD are astonishing. The trend and the respective approximated aqueous solubilities are as follows: DIMEB (33mg/100ml) > TRIMEB (25mg/100ml) > BCD (1.85mg/100ml).[4] At a glance, it is somewhat preposterous considering the hydrophilicity of the β-CD is reduced with the addition of the hydrophobic groups and the aqueous solubility of the CD should be reduced. The reasoning regarding the solubility enhancement of methylating the CDs has not yet been rationalised. Most of theoretical and experimental studies involving DIMEBs and TRIMEBs tend to focus on their pharmaceutical applications than to uncover their fundamental solvation properties.[5]

As was mentioned in the previous chapter, NMR spectroscopy has been applied extensively for identifying the hydrogen bonding network in cyclodextrin. The secondary hydroxyl groups at the wider rim of the cyclodextrin were found to take part in the inter-residue hydrogen bonding phenomena.[6] Some researchers suggested that this is the primary cause of the abnormal solubility of β-CD. Based on the same reason, most researchers consider the rigidity of the cyclodextrins to be

primarily due to the inter-residue hydrogen bonding formed between the C2 and C3' hydroxyl group.[7] Some researchers even postulate the increase in solubility of DIMEB and TRIMEB is simply due to a reduction of their molecular rigidity, caused by breaking the completeness of inter-residue hydrogen bonds on the lower rim of the cyclo-amylose.[8] From the condensed phase simulation standpoint, one has to consider that not only would the bridging water effect be constantly interfering with the inter-glycosidic hydrogen bonding, but the flexible distortion motions of the CDs would certainly also be hindering the inter-glycosidic hydrogen bonding.[9] therefore we suspected that the participation of the hydrogen bonding should be strong only if that particular cyclodextrin exhibits high internal rigidity. In addition, the hydrogen bonding belt argument fails to account for the reason why TRIMEB (with no possibility of inter-residue hydrogen bonding occurrence) has a lower aqueous solubility compare to DIMEB (with allows a partial hydrogen bonding belt). Thus, this led us to suspect that the hydrogen bond belt would not play such a decisive role in altering the flexibilities of these CDs. In order to fully justify our suspicion, two additional CDs, 2D2HB and 2D2MB, were also incorporated. Both of them do not participate in hydrogen bonding but have clear differences in their steric hindering nature.

8.2 Simulation and Experimental Conditions

The program CHARMM[10] has been used for all the computations reported here and the simulation conditions were kept essentially the same as listed in section 7.2 in the previous chapter, except for the size of the solvent box. In order to maintain the same solution densities of 1.013 g/cm^3 in each case, the solvent box sizes were adjusted to maintain the solution densities of 1.013 g/cm^3 in each case. The DIMEB, TRIMEB, 2D2HB and 2D2MB were immersed in solvent boxes of 4040 SPC/E water molecules of average dimensions 49.69Å, 49.71Å, 49.67Å,and 49.65 Å respectively.

Similarly, the translational diffusion measurements were performed using the pulse-field-gradient spin-echo nuclear magnetic resonance (PGSE NMR) experiment under the same conditions stated in section 7.2 in the previous chapter.

8.3 Results and Discussion

8.3.1 CD Diffusion Properties and Solvent Behaviours

8.3.1.1 Diffusion of Cyclodextrins: Comparison Between NMR and MD

The experimental diffusion rate measurements for each of the permethylated CDs are listed in Table 8.1. Within the same table, the diffusion rates obtained from the MD simulations are also listed. These values were calculated from the mean square displacement using the Einstein relation (refer to chapter 3). It is necessary to point out the slight irregularity of the NMR results. The diffusion rates

obtained the pulse-field-gradient spin-echo nuclear magnetic resonance (PGSE NMR) experiments from do not seem to follow a trend of solubility nor a dependency on molecular weight. This disagreement should be dealt with and resolved in the future work.

Table 8.1: Diffusion coefficients (x10^{10} m^2/s) of β-CD and its derivatives measured from NMR and calculated from MD simulations within 95% confidence limits

Method	β-CD	DIMEB	TRIMEB	2D2HB	2D2MB
NMR	2.168±0.015	2.014±0.014	2.218±0.003	--	--
MD	2.400±0.060	2.525±0.066	2.507±0.064	2.890±0.072	1.821±0.031

The MD results showed that both DIMEB and TRIMEB exhibit higher diffusion rates than that of the β-CD. This could be due to the fact that both these methylated derivatives are substantially more soluble than β-CD hence they tend to move about with less obstruction within the bulk solvent. Similarly, one would estimate the aqueous solubility 2D2HB to be higher than the β-CD also. Moreover, since 2D2MB reveals a diffusion rate that is much lower than the rest CDs, this implies that 2D2MB should be less soluble within the aqueous medium. The hydrophobic effect initiated by the stiffed methyl groups disrupt the hydrogen bonding network of the water solvent molecules hence could induce an additional drag of salvation shells.

8.3.1.2 Three-dimensional Water Distributions about the CDs

The three-dimensional water distributions above each β-cyclodextrin derivative was analysed using the spatial distribution function (SDF) method described in chapter 7. Contour probability densities of 50% and 90% greater than that observed in the pure solvent were selected for graphical illustration of the solvent configurational probability about the three cyclodextrins. These contour plots are shown in Figure 8.4.

Visual assessment of the SDF's indicates that the probability of water molecules locating themselves around β-CD (Figure 8.2 (i) and (ii)) is greater than that for DIMEB (Figure 8.2 (iii) and (iv)) and TRIMEB (Figure 8.2 (v) and (vi)). The SDFs of TRIMEB seems to bring about a higher degree of solvent orderliness when compare with the SDFs of the DIMEB. This coincides with the trend of their aqueous solubility differences. Based on the same argument, we estimated that 2D2HB (Figure 8.2 (vii) and (viii)) should ideally have a higher aqueous solubility whereas 2D2MB (Figure 8.2 (ix)

and (x)) would be poorly solvated when immersed within an aqueous medium. It is apparent that there is a substantial degree of marked water solvent orderliness around the methyl groups of the 2D2MB.

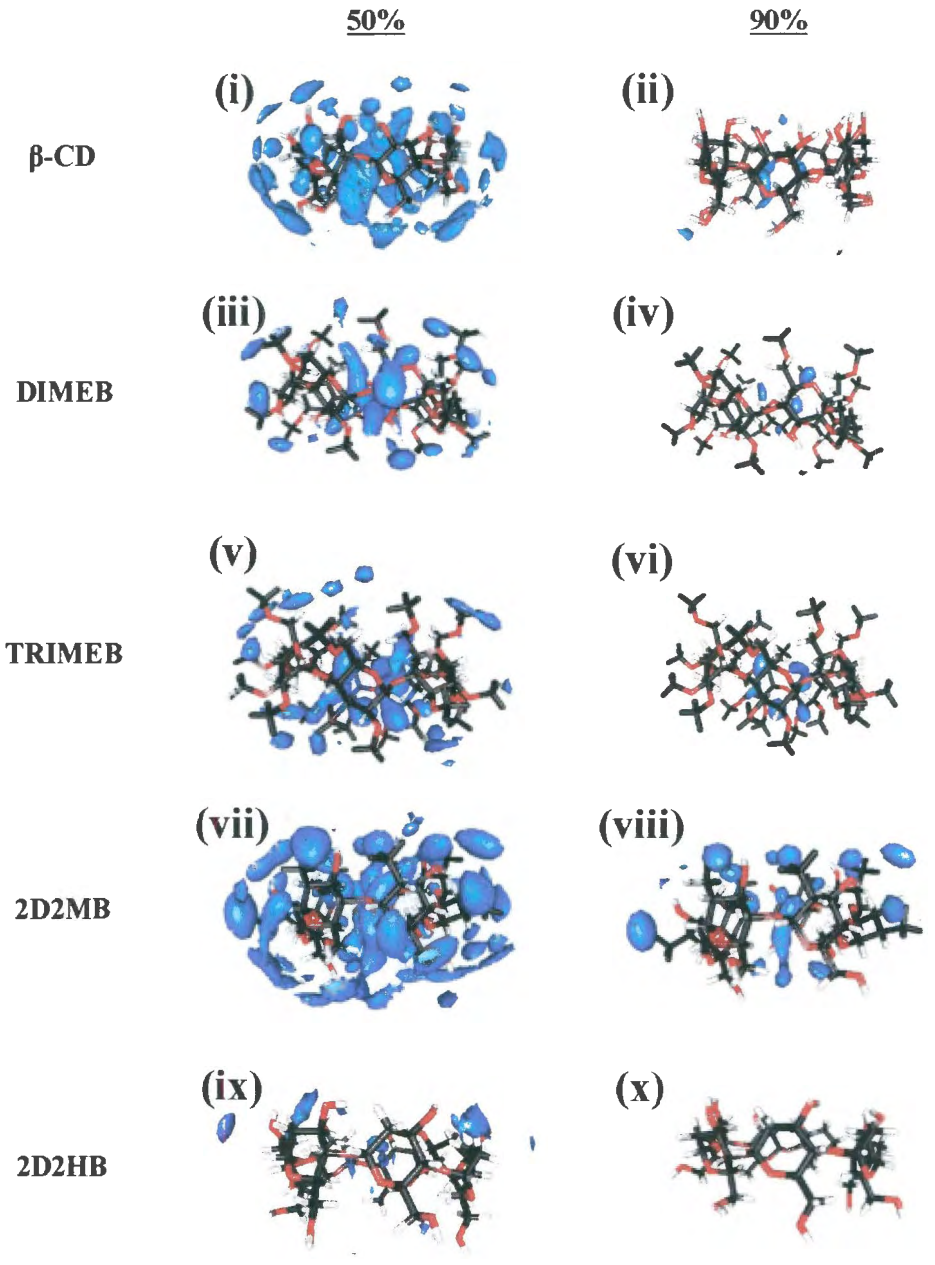


Figure 8.2 **The average structure of CDs along with the contour plots of water SDFs corresponding to the probability densities 50%, and 90% greater than bulk water as calculated from the MD trajectories.**

The volume of the contours associated with the above bulk water probability densities of the four CD solutions are calculated and compared with their native form. This gives a more accurate assessment of the total fraction of waters that are localised about these CDs in comparison with the normal waters that are located in the bulk solvent. The results of these calculations are illustrated in Figure 8.3.

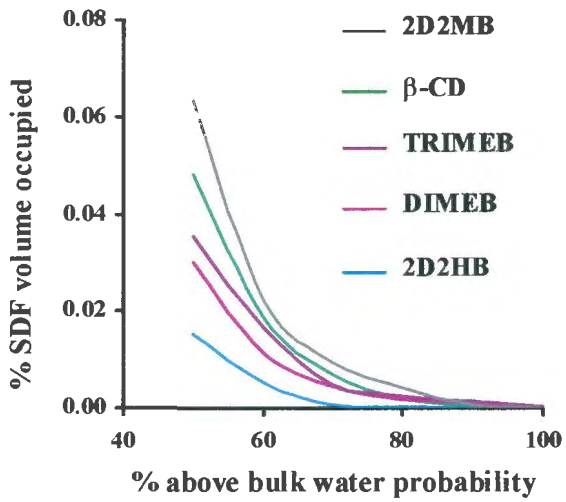


Figure 8.3 The volume fraction (x10⁴) occupied by the structured water, derived from the SDFs for the CDs.

Once again a strong correlation between solubility and the extent to which individual CD's interfere with the water molecule locations in solution emerges. It is clear that the less soluble β-CD induces a far more substantially occupied volume of the localisation of water solvent molecules. This is followed by TRIMEB and then DIMEB. The trend coincides with the differences in their solubilities.

When examining the other two derivatives, the water solvation mapping around the 2D2MB is notably more orderly than that around D2HB which seems to have disturbed a bulk water pattern only slightly. This is due to the fact that the water molecules around the bulky hydrophobic functional groups require the adaptation of physical changes associated with increased hydrogen bond organization.

8.3.1.3 Solvent Molecules Trapped within a CD Cavity

It has been widely accepted that the permethylated groups cause a subtle increase of the cavity size of both the DIMEB and TRIMEB torus and as a result of this, the amount of water molecule

entrapped within the cavity of individual CDs was found to differ slightly. Table 8.2 listed the estimated volume of the CDs, the amount of water found within the cavity and the water survival rate (T) obtained from the Pj(t,ti;t*) calculation.

Table 8.2 Predicted cavity volume, quantity and the survival rate (T) of the water molecules within respective CDs

	β-CD	DIMEB	TRIMEB	2D2MB	2D2HB
Entrapped water molecule	5.35	5.89	5.03	5.41	5.32
T (ps)	3.85	0.86	0.98	4.35	1.03

The calculation reveals that the survival rates of DIMEB and TRIMEB are estimated to be shorter than that of their parent β-CD. Furthermore, the calculations also indicate 2D2HB exhibits a similar T in comparison with those of permethylated CDs. Conversely, 2D2MB showed the tendency of water molecules being trapped in its cavity for a much longer period of time than the rest of the CDs does. In chapter 7, a connection was established between the affinity of an individual CD to entrap water molecules within its cavity and the internal motion of that CD. The degree of solvation of the CDs is related to the extent of disturbance exerted by the CDs on the free moving bulk water rather than being related to their molecular weight. The conformational analyses uncover a direct correlation between the flexibility of the CDs and the extent of localised water solvents surrounding the cyclo-amylose ring. The differences between solute solvent interactions for these CDs must be a result of their conformational behaviour. Once again, similar investigations involving the conformational properties of the CDs were performed in order to rationalise the above findings for these chemically modified CDs.

8.3.2 Conformational properties of the modified CDs

8.3.2.1 Internal Motion of the CDs

The amount that the internal motion of each CD affects its molecular relaxation was estimated by the root mean square (RMS) fit autocorrelation calculation. The evaluations were carried out after affects of orientational and translational motions of the CDs were eliminated. Hence, only the conformational motions of each CD over their 5ns data sets were taken into consideration. The internal motions of the β-CD are compared to these other chemically modified CDs in Figure 8.4.

There are noticeable differences in the magnitude of RMSD macrocyclic ring motion between the CDs (0.5, 2.2, 3.1, 0.9 and 0.3Å for β-CD, DIMEB, TRIMEB, 2D2HB and 2D2MB, respectively). The estimated relaxation times have shown significantly differences between them (approximately

210, 320, 590, 290 and 170ps, in the same order mentioned respectively). Interestingly, even though the RMS deviation of the permethylated CDs is far larger than that of β-CD, but their relaxation rate plateau is much slower than that of β-CD. This analysis shows that the permethylated CDs takes on various extensively distorted macrocyclic ring motion rapidly. Figure 8.4 also reveals that the 2D2MB exhibits a smaller range of internal motion and is thus more likely to trap water molecules within its hydrophobic cavity and its first hydration shell compared with the relatively more flexible 2D2HB, which hardly interferes the surrounding solvent molecule.

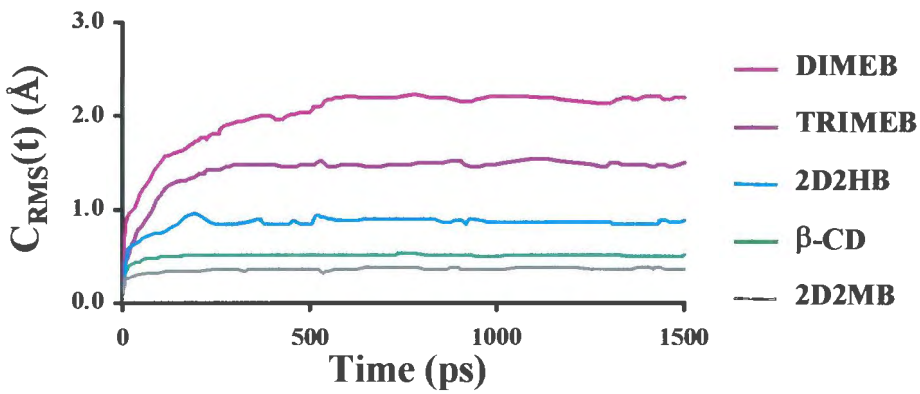


Figure 8.4 Auto-correlation functions of the RMS fit calculated for the β-CDs and its chemically modified derivatives.

The X-ray study has shown that within the crystalline state, one of the monomer units of TRIMEB prefers the $^1C_4$ conformation.[11] However, it has been reported that solvation induces a flexibility freedom upon CDs and further reduces their internal strain.[12] After performing the Pucker analysis, the results indicate that all of the monomer units practically remain in their favoured $^4C_1$ chair conformations. Therefore, we believe that these β-CD derivates also rely on the macro-cyclic ring distortion or the monomer tilting for relieving the internal strain.

8.3.2.2 Macro-cyclic Ring Distortion

The macro-cyclic ring distortion and monomer tilting behaviour of the CDs were carefully analysed over the 5ns trajectory. The procedures are in accordance with those previously defined in chapter 7. It was found that the macro-cyclic ring distortion behaviours of β-CD and its permethylated derivatives are fairly similar to one another. The latter, however, show a slight increase in the degree of flexibility in comparison with β-CD.

From the X-ray crystallographic study, it was found that the trend of planarity of the macro-cyclic ring is in the order of DIMEB (0.09Å), β-CD (0.14Å) and TRIMEB (0.44Å). It is evidently that such results do not correlated with the trend of aqueous solubility of these CDs. Thus, since molecules are often trapped in a higher strained conformation within the crystalline state due to the lattices energy, we believe that one cannot rely on the above crystallographic data to explain the ring flexibility tendency in the case of these methylated β-CD derivatives.

Unlike in the crystalline state, CDs in the condensed phase can undergo a variety of conformational changes and resulting in a range of different molecular conformations.[12,13] The assessment of macro-cyclic ring distortion revealed that both methylated β-CD derivatives possess a slightly highly degree of macro-ring flexibility compared to their native β-CD. Although this investigation can not sufficiently account for the difference between the aqueous solubility of the three CDs (specially since both DIMEB and TRIMEB both show a similar degree of ring flexibility), but result does imply that both DIMEB and TRIMEB are more flexible than the β-CD. This finding is in agreement with the trend of their aqueous solubilities. Figure 8.5 also shows that the macro-cyclic ring of the 2D2HB is much more flexible than the corresponding rings of the other CDs. The wide range in the value of 2D2HB's pseudo-dihedral clearly indicates that its cycloamylose ring possesses a high degree of flexibility. On the contrary, 2D2MB demonstrates a relatively restricted macro-cyclic ring distortion.

At this point, we would like to emphasise that TRIMEB, 2D2HB and 2D2MB are all incapable of initiating inter-residue hydrogen bonding involving the functional groups situated on the C2 and C3' position. However, the contrast of their ring distortions is clearly observed. Thus, it is believed that one can not assume the solubility of the CDs is directly linked to the concept of hydrogen bonding belt as was suggested by various researchers.

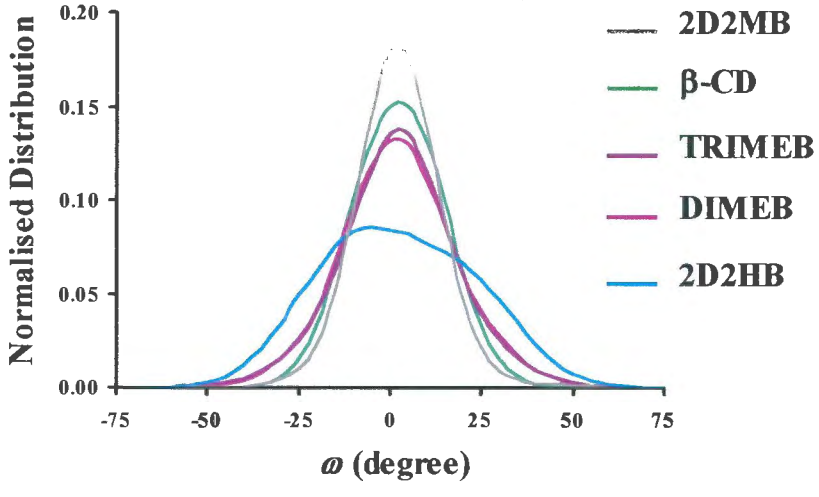


Figure 8.5 The normalised distribution of macro-cyclic ring pseudo-dihedral angles of the CDs

8.3.2.3 Monomer Tilting

According to the x-ray structure, the tilting angles of the glucopyranose monomer unit for β-CD, DIMEB and TRIMEB have averages and standard deviation of -14±10, -14±8 and -20±25 degrees respectively.[14] The relatively large standard deviation of the TRIMEB tilting angle indicate there is a high degree of steric hindrance between the monomers and the monomers are forced away from each other. However, the since these monomers can be affected by crystal packing forces and therefore the relative molecular geometry of the TRIMEB may be trapped in the unfavourable conformation.[12]

The tilting motion of each individual monomer of the different CDs are analysed over their dynamics trajectories and the results are shown in Figure 8.6. These different monomers of the DIMEB, TRIMEB, 2D2HB, 2D2MB are denoted as B1-7, D1-7, T1-7, H1-7 and M1-7, respectively. The plots clearly indicated that both DIMEB and TRIMEB exhibit a relatively high degree of monomer tilting when compared with the same glucopyranose motion of β-CD. Moreover, it is apparent that the monomer units of TRIMEB showed a more limited tilting range in comparison to the dynamics of their dimethylated derivatives. The trend of the motional freedom for these three CDs is consistent with the trend of their aqueous solubilities. The results present here is also in good agreement with the computational work performed by Reinhardt *et al.*[15] However, without performing potential mean force calculations, one can not propose a more accurate reasoning regarding as why DIMEB exhibit a higher degree of monomer titling than as of TRIMBE. One can only speculate that the inter-glycosidic steric hindrance is higher in TRIMEB than of DIMBE due to the additional replacement of a hydroxyl functional group by an *–O*-methyl substituent.

The limited range of tilting motion exhibited by 2D2MB (Figure 8.6(v)) demonstrates a distinct dissimilarity to the tilting motion of the glucopyranose units found within other derivatives (Figure 8.6(i)-(vi)). We believe that the restricted motion of the 2D2MB monomer is owing to the steric interference caused by rigid but bulky methyl substituents that constantly impede with the rotational motion of the neighbouring monomer residues. In the case of the 2D2HB, the hydroxyl group on the C2 position of each glucopyranose residue is been replaced by a small hydrogen atom, the high magnitude of tilting motion was observed. It is believed that since hydrogen atoms have smaller van der Waals radii, they tend to cause less inter-residue interferences, such substitution presides an increase of the glycosidic rotational freedom of 2D2HB.

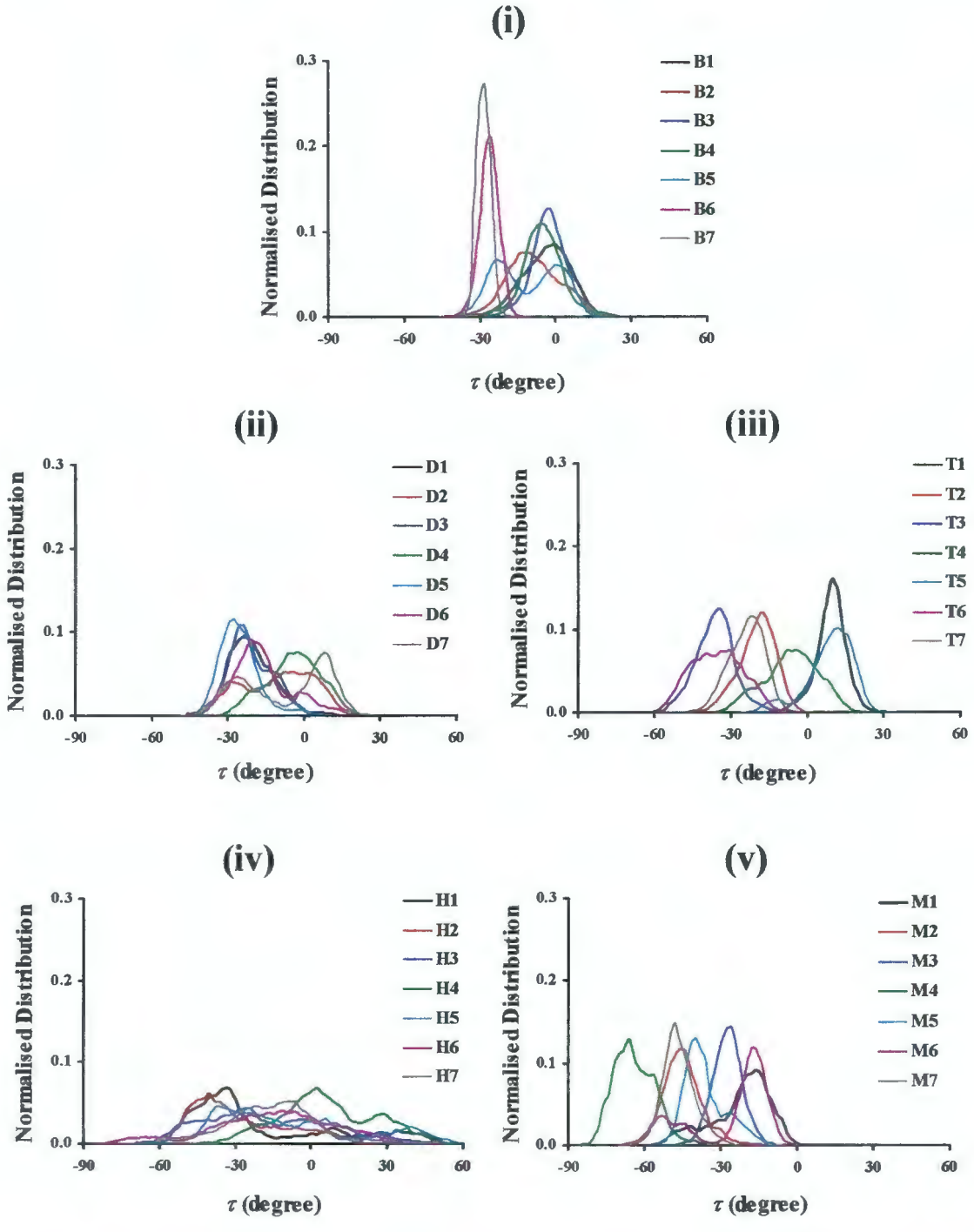


Figure 8.6 The distribution and the normalised ratio of each individual glucopyranose monomer titling angle ((i) β-CD, (ii) DIMEB, (iii) TRIMEB, (iv) 2D2HB and (v) 2D2MB).

8.4 Conclusion

A number of structural modifications to the native cyclodextrins have been made in an attempt to improve their physical properties. Methylation is one of the simplest modifications and this synthetic route is commonly used to yield the heptakis(2,6-di-*O*-methyl)- and heptakis(2,3,6-tri-*O*-methyl)-CD derivatives. The substituting the hydroxyl groups with the permethyl groups can drastically enhance the aqueous solubility of the CDs. Most computational studies of methylated CDs have involved host-guest complexes and have neglected the fundamental problem of the preposterous solubility enhancing nature. In this chapter, we carried out that a combination of some straight-forward studies to rationalise the puzzling solvating nature of these methylated CDs.

Traditionally, it is believed that the anomalous aqueous solubility of β-CD is as a result of the C2 and C3' inter-residue hydrogen bonding network. Two additional β-CD derivatives, 2D2HB and 2D2MB, were employed to justify this theory. Even though both CDs derivatives are incapable of participating inter-residue hydrogen bonding, the solvent ordering around 2D2HB is noticeably lower than that around 2D2MB. It has already been previously established that an increase in intra-residue hydrogen bonding strength does not necessary induce a reduced glycosidic rotation. The results obtained from this chapter further invalidate the conventional assumption on the significance of these belts of hydrogen bonding towards the CD aqueous solubility. However, in order to consolidate the results presented in this thesis, the potential mean force calculations should be included in the future studies.

REFERENCE

(1) (a) Szejtli, J. *Chem. Rev.* **1998**, *98*, 1743. (b) Hedges, A. R. *Chem. Rev.* **1998**, *98*, 2035. (c) Jones, S. P.; Grant, D. J.W.; Hadgraft, J.; Parr, G. *Acta Pharm. Tech.* **1984**, *30(4)*, 263.

(2) (a) Gattuso, G.; Nepogodiev, S.; .A Stoddart, J. F. *Chem. Rev.* **1998**, *98*, 1919. (b) Khan, A. R.; Forgo, P.; Stine, K. J. D'Souza, V. T. *Chem. Rev.* **1998**, *98*, 1977.

(3) (a) Uekama, K.; Hirayama, F.; Irie, T. *Chem. Rev.* **1998**, *98*, 2045. (b) Kata, M.; Selmeczi, B. *J. Incl. Phenomen.* **1987**, *5(1)*, 39-43. (c) Mueller, B. W.; Brauns, U. Int. J. Pharm. **1985**, *26(1-2)*, 77. (d) Szejtli, J. *Starch/Staerke 1984, 36(12), 429-32.*

(4) (a) CycloLab, Cyclodextrin Research and Development Laboratory Ltd. *Sample Specification Product Information Sheet* Budapest, H-1525 Hungary **2001** (b) MATERIAL SAFETY DATA SHEET based on Directive 67/548/EEC (as amended) or Directive 1999/45/EC of the Commission of the European Communities. **1999**

(5) (a) Linder, K. and Saenger, W. *Carbohydr. Res.* **1982**. *99*, 103. (b) Kuan, Fu Hua; Inoue, Yoshio; Chujo, Riichiro. J. Incl. Phenom. **1986**, *4(3)*, 281. (c) Suzuki, M.; Szejtli, J.; Szente, L. Carbohydr. Res. **1989**, *192*, 61. **(d)** Green, A. R.; Miller, E. S.; Guillory, J. K. *J. Pharm. Sci.* **1991**, *80(2)*, 186.

(6) (a) Miyazawa, I.; Endo, T.; Udea, H.; Nagai, T. *In Proceedings of the 7th Interantional Symposium on Cyclodextrins, Tokyo, 1994,* ed. T. Osa, Publ. Office Acad. Soc. Japan, Tokyo, **1994,** p214 (b) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *Eur. Biophys. J.* **1988,** 20, 241. (c)

(7) (a) Gillet, B.; Nicole, D. J.; Delpueche, J. J. *Tetrahedron Lett.* **1982,** *23*, 65. (b) Rees, D. A. *J. Chem. Soc.* **1970**, 877. (c) Koehler, J. E. H.; Saenger, W. van Gunsteren, W. F., *J. Mol. Biol.,* **1988,** 203, 241.

(8) Schanberger, B. P.; Jansen, A. C. A.; Janssen, L. H. M. In *Minutes of the 4th Interantional Symposium on Cyclodextrins. Paris, 1988* Ed. Huber, O.; Szejtli, J. Kluwer, Dordrecht, **1988,** 61.

(9) (a) Naidoo, K. J.; Kuttel, M. *J. Comput. Chem.* **2001**, *22*, 445. (b) Brady, J. W. Solvation: *Carbohydrates. In Encyclopedia of computational chemistry*, 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons: Chichester, **1998**; p2609.

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

(11) Caira, M. R.; Griffith, V. J., Nassimbeni, L.R.; van Oudtshoorn, B. *J Chem. Soc. Perkin Trans. 2*, **1994,** 2071.

(12) Lipkowitz, K. B. *Chem. Rev.* **1998**, *98*, 1829

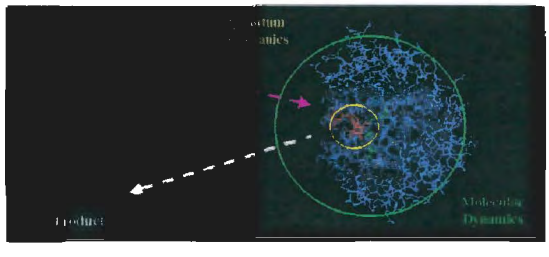
(13) (a) Linert, W.; Margl, P.; Renz, F.; *J. Chem. Phys.*, **1992**, *161*, 327-338 (b) Momany, F. A.; Willett, J. L. *Carbohydr. Res.* **2000**, *326*, 210 (c) Jaime, C. Cyclodextrins. In *Encyclopedia of Computational Chemistry*, 1st ed.; van Ragué Schleyer, P., Ed.; John Wiley & Sons Ltd.: Chichester, 1998; p644. (d) Saenger, W.; Jacob, J.; Steiner, T.; Hoffmann, D.; Sabe, H.; Koizumi, K.; Simth, S. M.; Takaha, T. *Chem. Rev.* **1998**, 1787. (e) Harata, K, *Chem. Rev.* **1998**, 98, 1803. (f) Dauchez, D.; Vergoten, G. In *Minutes of the 5th Interantional Symposium on Cyclodextrins. Paris,* **1990**, Ed. D. Duchene, Ed. De Santé, Paris, **1990**, 101 (g) Maunza, B.; Deiana, S.; Pintore, M.; Gessa, C. *J. Mol. Struc.* **1997**, *419*, 133.

(14) Harata, K. *Chem. Rev.* **1998**, *98*, 1803

(15) Reinhardt, R.; Richter, M.; Mager, P. P. *Carbohydr. Res.* **1996**, *291*, 1.

Chapter Nine

Conclusions and Future Studies



9.1 Conclusions

Carbohydrates such as α(1→4) linked saccharides have unique physiochemical properties which make them invaluable in the industrial applications. The flexibility of the α(1→4) linked glycosidic bond in the aqueous condensed phase can be attributed to the type of atoms involved in the glycosidic linkage and the cross-glycosidic non-bonded interactions, *viz.* hydrogen bonding and hydrophobic effects. This thesis investigate influence of these interactions which initiated from the inter-residue functional groups attached on the C2 and C3' position of the α(1→4) saccharides.

The greater part of a saccharide is considered to be hydrophilic as carbohydrates normally contain a large amount of polar hydroxyl groups. These hydroxyl groups are capable of forming both inter- or intra-molecular hydrogen bonding. While the magnitude of the inter-molecular hydrogen bonds can be readily and accurately assessed both computationally and experimentally, the energy of intra-molecular hydrogen bonds has not been easily accessible in the past. An efficient procedure for evaluating the intra-molecular hydrogen bond strength of different saccharides has been specially developed. Using Bader's Atoms in Molecule Theory (AIM), this thesis presents an effective way for evaluating the various important topological properties of electron density between the complex. The ΔE^{HB} vs Δ_{el+lap} relationship allows the construction of separate correlation curves for each combination of hydrogen bond donor/acceptor pairs. Using the appropriate ΔE^{HB} vs Δ_{el+lap} curve associate with a pair of hydrogen bonded functional groups, the strength of the intra-molecular hydrogen bonds for the same type of interaction in biopolymer system can be predicted. In order to find the strongest cross-glycosidic interactions, the hydrogen bond strengths between a set of chemically modified saccharides were evaluated and ranked using this method.

In continuation with the above findings, the adiabatic maps and vacuum dynamic calculations were performed to examine the flexibility of these modified disaccharides. Combined with the vacuum dynamic simulations, adiabatic maps can indicate of the preferred sampled region for that particular carbohydrate during a dynamics simulation. Employing these two studies in conjunction enable the flexibility of the carbohydrate to be esitmated. These studies reveal that GlcNAc-α-(1→4)-3-amine-Glc has limited glycosidic flexibility in vacuum. Therefore this molecule was chosen for additional investigation involving aqueous condensed phase simulations.

The influence of water molecules to the internal flexibility of the GlcNAc-α-(1→4)-3-amine-Glc was explored. It was found that despite the stronger intra-glycosidic hydrogen bonds in the GlcNAc-α-(1→4)-3-amine-Glc, the correlation times for glycosidic dihedral angle fluctuations are approximately the same as for maltose. Investigation of the first hydration shells associate with the two disaccharides revealed high water probability densities region straddling between the cross-glycosidic linkage functional groups. This probability density corresponds to single water molecules forming

bridging hydrogen bonds between the functional groups on either side of the linkage for periods of 3.66ps in the case of maltose and 8.36ps in the case of the derivative. With the aid of the ΔE^{HB} vs. Δ_{el+lap} correlation curves, *ab initio* studies demonstrated that these sugar-solvent hydrogen bonds are of similar strength to the sugar-sugar hydrogen bonds. This combination of molecular dynamics and *ab initio* computational methods demonstrates that increasing the internal hydrogen bond strength in oligosaccharides does not necessarily lead to slower internal molecular motion of these sugars in solution. The intermolecular hydrogen bonds formed with water compete equally with the intra-molecular hydrogen bonds in the sugar. Hence, one can conclude that an increase in the strength of cross glycosidic hydrogen bonding do not necessary induce an increase in the molecular rigidity.

Addition of hydrophobic substituents commonly reduces the degree of polysaccharide solubility. This is due to the fact that the hydrophobic groups of the solute interrupt the randomness of the water solvent molecules and induce an entropic effect of increasing in the orderliness of water solvent molecules around hydrophobic functional groups of the solute. However, introducing the hydrophobic groups onto a carbohydrate does not necessary lead to a decrease in its aqueous solubility. For instance, the intrinsic aqueous solubility of cyclodextrins can be considerably enhanced when permethylated. Prior to solving this perplexing issue, probing the underlying explanation concerning the anomalous differences in the relative solubility of the α-, β-, and γ-CD is essential.

The relative solubility of α-, β-, and γ-CD is known to deviate from simple mass dependence as β-CD being is much less soluble than the others. The underlying molecular origin of this anomalous trend was inspected by analysing the motions of the CDs from the trajectories generated in molecular dynamics (MD) simulations. In addition, translational diffusion measurements have been carried out for the three CDs using a PGSE NMR techniques and the corresponding diffusion coefficients obtained from the MD trajectories have found to be in good agreement with those observed experimentally. The MD trajectories exhibit that the more soluble CDs tend to trap water in their ring cavity for a shorter time than the others. In addition, the relaxation rate of internal motion for the more soluble CDs, as seen from time correlation functions, suggested they have more flexibility freedom. This anomalous solubility of the CDs can be explained by detailed assessments of their internal motion behaviours. The internal motions analyses showed the macro-cyclic ring and individual monomer tilting is most rigid in the case of β-CD in comparison with the other two native CDs. Hence β-CD is least soluble – in line with the anomalous trend.

Along with two chemically modified CDs (listed in chapter 8), similar investigations involving the internal motion of the native CDs were applied to the di- and trimethylated β-CD derivatives for explaining the contrast in their aqueous solubility. The results illustrate that some of the less bulky hydrophobic functional groups cause an increase of the CD's internal motion and consequently enhance the solubility of that particular CD. The other more bulky hydrophobic functional groups, such as replacing hydroxyl groups with methyl substituents, were found to hinder the glycosidic

flexibility and hence induce a more structured solvent organisation around the cycloamylose ring. Therefore, one can conclude that the bulkiness and the rigidity of the inter-glycosidic substituents played a more significant role to the aqueous solubility of the carbohydrate than hydrogen bonding.

It is essential to have detailed knowledge of the conformational response of a new carbohydrate material to its environment in order to design biodegradable polymers with improved functionality. This work has uncovered some of the fundamental issues by the use of a variety analytical tools. The results presented in this work have important implications when considering the mechanism of hydrophobic and hydrophilic effects in more intricate problems such as folded glycoproteins, enzyme-substrate interactions and host-guest complexation.

9.2 Future Studies

The glycosidic rotational freedom of the α(1→4) saccharides can be controlled by introducing some bulky hydrophobic groups on the C2 and C3' positions. However, such modification may result in undesirable effects such as a reduction in solubility and a decrease in biodegradability. Modifying the atoms involved in the glycosidic linkage (i.e. replacing the glycosidic oxygen with other atoms such as nitrogen, sulphur or carbon) may be an alternative manner to reduce the glycosidic flexibility of the saccharides.

It has been shown in this project that simulations comprise of 5ns trajectory are sufficiently long enough to estimate some important transitions for α- and γ-CDs within the condensed phase. However, use of a longer trajectory is advised when investigating the oligosaccharides which have a slight higher degree of internal rigidity, such as is the case of β-CD. A longer simulation should allow all the important transitions of the molecule to take place and subsequently increase the confidence of the results obtained.

Presently, literatures provide little useful explanatory theory for polysaccharide hydration, with poor predictability. This work demonstrated that with the aid of detailed molecular motion investigations, SDF studies are useful for providing information regarding the water structuring about carbohydrates. However, the predictability of carbohydrate solubility would remain inadequate unless the solvation free energy calculations are explored. The change in solubility resulting from modification on the inter-glycosidic functional groups on the saccharide can be better understood from such studies of the solvation free energy surfaces. This will be particularly useful in rationalising the basis of the perplex solubility trend of the methylated cyclodextrins.

Should a suitable biodegradable saccharide be developed in the near future, the degree of the biodegradability of this *de novo* saccharide need to be determined by use of by introducing a cleavage enzyme. It has long been suggested that the arrangement of hydrogen bonds on the complementary

surfaces of biological molecules play a vital role in the binding specificity of an enzyme. The electron correlation curves developed in this work can assist in locating the conformations associate with molecular binding for systems known to bind but whose exact binding mechanisms are unknown. Instead of dealing with the full biomolecular structure which is often extremely large in size, use of these curves can greatly simplify the complex task by isolating the hydrogen bonding responsible for the binding. In order to account for the dynamic behaviour of the biomolecule and as well as allow the forming and/or breaking of the bonds, other more sophisticated methods, such the combined electronic structural approaches and empirical force field methods, should be employed.