Resolving the Hydration of Hexacyclic Systems Using Ultrasonic Interferometry and Computer Simulations

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

MASTER of SCIENCE

by
Annalisa Boscaino
M.Sc. (Hons) (Universita' degli Studi di Napoli "*Federico II*")

September, 2010

Supervisor: Associated Professor Kevin J. Naidoo

Resolving the Hydration of Hexacyclic Systems Using Ultrasonic Interferometry and Computer Simulations

Annalisa Boscaino   **September 2010**

ABSTRACT

The major part of this thesis focuses on investigating dynamic hydration behaviour of a series of hexacyclic molecules in water, by using computational techniques validated through ultrasound interferometry experiments. The water taking part to the diffusive motion of the solutes has been characterised by performing mainly time-dependant functions (which allows to simulate hydration numbers) and ring puckering analysis, to understand the conformational features and the role of the functional groups on the flexibility of the solutes.

The ultrasonic interferometry has been employed for its ability to accurately calculate the hydration number in the first coordination shell of a solute. Molecular dynamics (MD) simulations have been used to investigate the structure of the solvent surrounding the solute, in a time-dependant manner. In particular, from MD trajectory, pair and spatial distribution functions, residence time of water in the first solvation shell and auto and cross-correlation functions have been calculated, as well as linear diffusion coefficients. Moreover, hybrid QM/MM systems gave insight about conformational behaviour of the molecules.

A section of this dissertation (**Chapter 5***) has been prepared to be submitted to the *Journal of Physical Chemistry B*. It summarises and discusses the results obtained. From the experimental measurements and computer simulations, it has been possible to draw conclusion about the relationship between molecular flexibility and hydration. The paper represents the conclusion of the first part of this thesis.

The dissertation is then completed by giving preliminary hydration and conformational behaviour results about another class of compounds, which represents an interesting future application for this work: a series of inhibitors, and more specifically of glycosidase-inhibitors. These molecules have been found to convert carbohydrates into simple sugars, i.e. monosaccharides, which can be then absorbed through the intestine. Hence, glycosidase-inhibitors can reduce the impact of carbohydrates on blood sugar. Therefore, it is clear their importance as anti-diabetic drugs. Further, these compounds are fascinating for their structural resemblances with glucopyranose, extensively studied in the first five chapters, leading to start a further investigation about analogies and differences in the hydration and flexibility behaviour in water between the two classes of carbohydrates.

"To this wonderful
experience in
South Africa and
all the beauty
that I saw"

AKNOWLEDGMENTS

LIST OF ABBREVIATIONS

MM	Molecular Mechanics
MD	Molecular Dynamics
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
CHARMM	Chemistry at Harvard Macromolecular Mechanics
Å	Angstrom (10^{-10} m)
nm	Nanometre (10^{-9} m)
ps	Picosecond (10^{-12} s)
ns	Nanosecond (10^{-9} s)
RT	Residence Time
TCF	Time Correlation Function
NMR	Nuclear Magnetic Resonance
n_H	Hydration Numbers
PDF (or RDF)	Pair (or Radial) Distribution Function
SDF	Spatial (or Solvent) Distribution Function
NJ	Nojirimycin
DNJ	Deoxynojirimycin
CSFF	Carbohydrate Solution Force Field
HGFB	Ha, Giammona, Field and Brady force field
PHLB	Palma, Himmel, Liang and Brady force field
CGenFF	CHARMM General Force Field
VMD	Visual Molecular Dynamic
IR	InfraRed (Spectroscopy)
FEP	Free Energy Perturbation

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

TABLE OF CONTENTS

INDEX of FIGURES, TABLES AND SCHEMES

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 1

Introduction

Understanding the ability of a molecule to perform unique functions is one of the most important targets of chemistry.

For solution chemistry, this ability strictly depends on the nature of the solute, and more specifically on the nature of the different functional groups of the solute and their geometric proximity. These features mainly determine the solute-solvent interactions and play a fundamental role in the solubility properties [1].

Therefore, by gaining insight on the solubility, one can reveal numerous aspects on the physical and chemical phenomena occurring in solutions.

From electrochemistry (devoted to the study of ions and, in particular, of the way in which the structure of bulk water is modified by the introduction of ions [2]) to the investigation of complex biological compounds, both experimental and, more recently, computational methods have been developed to understand solubility.

The latter is generally related to the maximum quantity of a solute that can be dissolved in a solvent. What factors affect the limit?

In this thesis, an attempt to answer to such a crucial question is provided, by focusing primarily on hexacyclic molecules in solution. Since these molecules have been considered in an aqueous environment, systems characterized by a solute surrounded by water will be exclusively discussed.

A prime role in solubility is played by the competition between hydrophobic and hydrophilic regions within and around the solute [3], which determines the nature and the number of solute-solvent interactions.

This concept can be thermodynamically expressed by the change in the Gibbs free energy of solvation, ΔG, which depends in turn on the variation of two state functions, entropy (ΔS) and enthalpy (ΔH). They contribute differently to the overall ΔG, depending on the nature of the solute [4] and are connected by the following equation:

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

where a more negative ΔG generally means higher solubility.

Although such a thermodynamic approach gives essential information on the tendency of a solute to be solvated in water, it has been found to be incomplete, when applied in the present thesis.

The aim here is to quantitatively describe conformational changes of the molecule which can affect its solubility, together with the ability of each functional group of the molecule, i.e hydroxyl groups, to forming hydrogen-bonds (in particular, intermolecular hydrogen bonds), and the influence of their geometric proximity on the hydrogen-bonds. In other words, a complete description requires an understanding of how the structure changes with time and how the molecule interacts with the solvent in a time-dependent fashion. Since all these aforementioned quantities enter into the determination of the ΔG, the relevance of each single parameter in the solubility of the molecule cannot be disentangled from the ΔG value itself. The ΔG can be considered a "static" parameter which hides dynamic information, as it does not disclose the dynamic conformational changes of the solute and the dynamic interactions between solute and solvent.

Therefore, the research has to be extended to a time-dependant level [5], which is incorporated in the concept of *dynamic* hydration. In numerous cases in literature [2, 6-16], dynamic hydration has been explained in terms of a parameter called the *dynamic hydration number*, n_H, generally referred to as the *primary hydration number*. This is usually described as the number of water molecules that binds the solute in its first coordination shell for times long enough that solute and solvent can diffuse together. In this way n_H encodes the dynamics of the molecule [2]. In this dissertation, the same definition has been applied.

Solubility and hydration refer to the same concept: they both investigate how strongly solute and solvent interact. However, hydration encodes an important feature: it is a time-dependent concept, a dynamic tool to be exploited when evaluating solubility.

Literature reports extensive studies of hydration, through both experiments and computer simulations, either by means of n_H or through measurements of other observables and features of the solutions, with applications to a wide variety of solutes, and from simple to very complex systems. In regard of complex systems, the three major classes of biomolecules - carbohydrates [8, 10-12, 17-19], enzymes [20], and proteins [5, 21-23] - are included. Obtaining hydration information for small molecules such as ions is relatively simple, because they are spherical and the electrostatic interactions dominate in the formation of the solvation sphere [2, 9], whereas in the case of bigger systems, there is a competition between several parameters and the problem ends up being less intuitive.

The most exploited experimental techniques for obtaining information on hydration are NMR (Nuclear Magnetic Resonance) [6, 7, 10, 15], and ultrasound interferometer [2, 9, 12, 13, 24] – these two methods are both able to give n_H values - and X-ray and neutron diffractions. The latter are more related to the investigation of the structure induced by the solute on the surrounding solvent [17, 25-30].

NMR spectroscopy represents the only experimental method that is able to give structural information at the atomic level and insight into the dynamics of the system at the same time, which makes it the method of choice for studying biologically important molecules in solution [5]. Slow exchange between two or more chemically distinct conformations can be directly investigated on the basis of spin lattice relaxation time constant, T_1, chemical shift measurements, line-shape analysis and saturation transfer experiments. nOe (*nuclear Overhauser effect)* enhancement can yield information on molecular reorientation times.

As mentioned above, from NMR it is possible to derive n_H. Its calculation is generally based on the comparison between T_1 of ^{17}O for water and T_1 for solutions at different concentrations. As an example, this procedure has provided a tool for the calculations of n_H in aqueous solutions of mono- and polysaccharides [6, 15, 31]. Although NMR is the most versatile experimental technique available for studying dynamics, with *state of the art* instrumentation it gives information "restricted" to the micro- and millisecond ranges, whereby it can provide only time-averaged dynamic information about events lasting less than one microsecond [7, 23].

Ultrasound interferometry gives insights on how strongly the solute and solvent interact. By using this technique, it is possible to derive numerous parameters (the ultrasonic velocity being one of the most promising) with high accuracy [9], from which one is able to derive different thermodynamic quantities as well as n_H. The theory and the experimental setup of ultrasound interferometry will be extensively described in **Chapter 4**, as this has been chosen, in this thesis, as the most adaptable method to obtain n_H.

A distinct role is performed by X-ray and neutron diffraction: they cannot be used for the calculations of n_H, but they are able to give fundamental insights into the conformation of the water surrounding a solute (i.e. measurements of radial distribution functions [17]).

Neutrons provide an ideal tool for the study of almost all forms of condensed matter. They are non-destructive and can penetrate deeply into matter. Thermal energy neutrons have wavelengths close to 1 Å, thus matching inter-atomic distances. Experiments using neutron radiation, therefore,

simultaneously provide information on the amplitudes and energies of atomic motions. Neutron diffraction has been adopted for examining the effect of the presence of a solute in a solvent, namely the structure induced by the former on the latter. On top of several other applications, this technique allowed focusing on the role of the specific constituents of a molecule in structuring the surrounding water. In the specific case of neutron diffraction applied to carbohydrates, Brady and co-workers discussed in some of their works the rotational conformation of the exocyclic hydroxymethyl group and hydroxyl groups of carbohydrates (in particular, D-glucopyranose and D-xylopranose) [26-28].

Focusing on the techniques where one is able to study hydration through measurement of hydration numbers, it has been found that different experiments can lead to very different n_H values, even for the same compound [2]. The definition of n_H does indeed strictly depends on the chosen experimental technique. The difference stems only from a different normalisation factor and it is independent of the concentrations of the solutions. Therefore, once n_H values from different experimental techniques are normalized for a given solution, one can compare them rigorously. In other words, by just taking the experimental results, one can only compare their trends.

This is mainly due to the fact that distinct techniques explore distinct properties and regions of the system. For example, in comparing NMR and the ultrasound interferometer, both of which are able to give n_H, the former derives n_H from comparing spin lattice relaxation time values of a reference (water) with different solutions at increasing concentrations, whereas the latter uses ultrasound speed coupled with density measurements of solutions at different concentrations. Consequently, by taking into account the different parameters of the systems, the two techniques derive n_H differently.

Further, in both cases, one can define n_H as the number of water molecules lying in the "first" coordination shell of the solute for a given time, but the definition of first coordination shell itself turns out to be slightly different in the two cases. While NMR defines a coordination shell around the solute which corresponds to the first solvation shell (which is the first peak of a Radial Distribution Function), the ultrasound interferometer takes into account a region which lies in between the first and second solvation shell (between the first and second peaks in the Radial Distribution Function). For all these reasons, once again, one can only compare n_H trends coming from two or more different techniques and it is meaningless to compare their absolute values.

These methods are undoubtedly necessary to investigate the chemistry of both simple and complex structures and to derive hydration properties.

However experiments cannot probe fast processes occurring in the picosecond and nanosecond timescale range, even if some progress has recently been made in this regard. One of the current mainstream implementations is to couple NMR and quasielastic neutron scattering (QENS), which allows expanding the time scale of observations down to picoseconds [32].

However they still need to be coupled with computer simulations which are able to cover a wide time range: in principle, they cover timescales from 10^{-10} to 10^{-3} s [33, 34].

Nevertheless, experimental techniques provide an irreplaceable tool for validating computer simulations, which otherwise would remain theory, lacking "real world" confirmation.

In this thesis, n_H measurements obtained through the ultrasound interferometer and computer simulations have been compared; the experiments mainly served as a validation tool.

Different procedures for obtaining n_H from both experimental [9-13, 15, 24, 35, 36] and computational techniques [2, 7, 10, 11, 24] are available in the literature. Here, Impey's computational procedure has been followed here [2], adapting his method of calculating n_H for ions to the case of hexacyclic (organic) molecules.

Similar to the purpose of this thesis, a 2007 publication by Burakowsky and Gliński showed the applicability of the acoustic Pasynski method (which was previously applied only to ions) for determining hydration numbers for some non-electrolytic aqueous solutions [9]. In particular, the authors have focused on a systematic calculation of primary n_H values for linear organic molecules of different lengths. It has been shown that for dilute aqueous solutions, the hydration numbers are strongly and directly dependent on the length of the solute molecules and on their constituents. The data obtained by Burakowski and Gliński suggested that the addition of a –CH$_2$ group to the chain increases the total hydration number of the molecule by approximately one unit: the dependency is thus close to linear. They conclude that the hydration number is an additive parameter even if the hydration mechanisms are different, in the sense that they are affected by several different parameters, strictly depending on the nature of the solute and its substituents.

The strategy for hydration studies applied in this thesis follows some of the aspects by Burakowsky and Gliński. There, they studied the hydration behaviour of a series of linear aliphatic molecules increasing in length through the addition of –CH$_2$ groups. Here, molecules having the same general structure (all hexacyclic) but differing in the number of substituents, i.e. OH around the ring, have been studied. In addition to their approach, special attention has been given to the introduction of

an oxygen atom in the ring, to try to explain its role in the hydration behaviour of two carbohydrates, β-D-glucopyranose and β-D-xylopyranose.
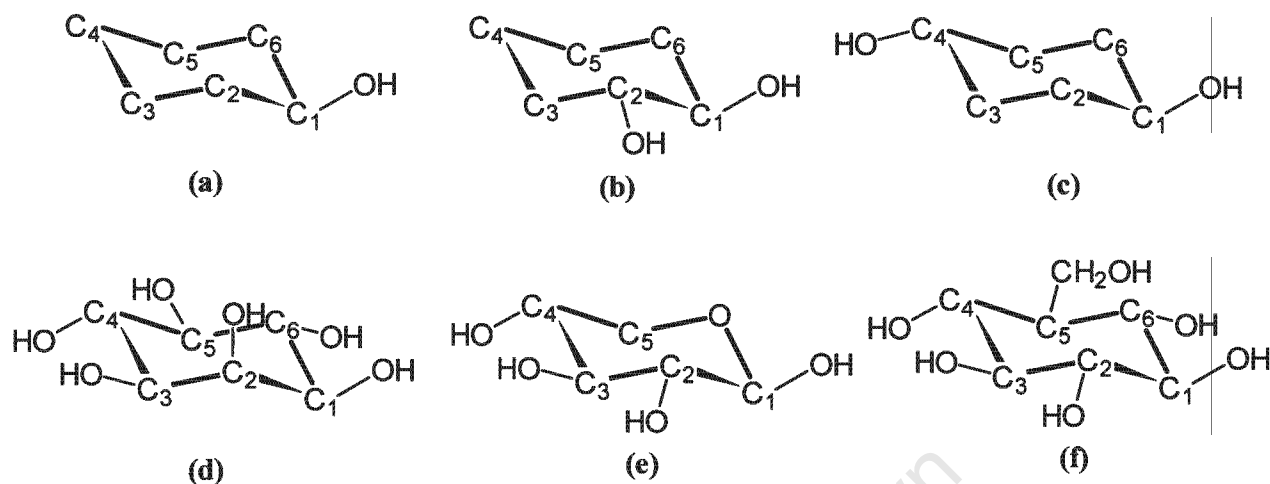
Figure 1.1 Schematic representation of the six hexacyclic molecules studied in this thesis: (a) cyclohexanol, (b) 1,2-cyclohexanediol, (c) 1,4-cyclohexanediol, (d) myo-inositol, (e) β-D-xylopyranose, (f) β-D-glucopyranose.

All these molecules (**Figure 1.1**) contain a different number of attached alcohol groups held in a cyclohexane-like "chair" geometry, as well as nonpolar centers (-CH and $-CH_2$ groups), and in the cases of carbohydrates (**Figure 1.1 (e)** and **(f)**) ether-like ring oxygen atoms are present. Therefore, by comparing such systems with each other, hydration behaviour can be understood as a function of the different number of OH, different functional groups and their relative positions on the ring.

The major aim has been to obtain a set of observations on the above mentioned systems, providing information about the way the solvation sphere develops around a solute as a function of the number and relative positions of interaction sites on the solute [1], which can then be extended to more extensive systems. Further, only "single" carbohydrates have been studied, motivated by the notion that by unveiling the behaviour of a single carbohydrate, it is in principle possible to predict the solute-solvent interactions of aggregates of carbohydrates, i.e. disaccharides and polysaccharides [7].

More specifically in the case of carbohydrates, previous attempts to study their hydration have been conducted by measuring not only hydration numbers [12][1] but many different parameters, such

---

[1] Reference reported here refers exclusively to n_H measured through ultrasound interferometry.

as the ratio of equatorial versus axial OH groups [6, 15, 37], the anomeric effect [38, 39], the hydrophobic/hydrophilic indexes [40], and the hydrophilic volume of the carbohydrate [41].

The n_H experimental results for glucopyranose and xylopyranose obtained in this dissertation agree with the literature values within the experimental error. Hydration numbers for cyclohexanol, 1,2- and 1,4- cyclohexanediol and myo-inositol (**Figure 1.1**, (**a**), (**b**), (**c**) and (**d**), respectively) are not available from literature resources.

For many different reasons, there has been an increasing interest in the hydration and conformational properties of sugars in recent years. Carbohydrates play key roles in biological recognition processes [42, 43], in the development of disease treatments, energy storage, structural support [44], and in many important areas of the food and technical industry. Further, carbohydrate-water interactions are fundamental for many macroscopic quantities of carbohydrates, in which the geometry of water induced by the solute causes significant changes in the properties of the materials [45]. Thus, understanding the aqueous solvation behaviour of single carbohydrates and polymers is of considerable importance. It is the nature of their interactions with water that is responsible for most of their biological functions.

However, from an experimental point of view, carbohydrates remain the least investigated among biomolecules, because they are difficult to synthesize and manipulate, and this makes it difficult to study their biological functions in detail [43, 46].

For all these reasons, computer simulations have been developed to study these compounds. Although in the past one important limitation for a molecular modelling approach was that carbohydrates were especially difficult to model due to their highly polar functional groups, their flexibility, and the differences in their conformations and configurations, fortunately, several contributions have been made to set up appropriate parameters. The CSFF force field, especially developed for carbohydrates simulation [46], has been used throughout this dissertation.

In this short paragraph, structural details of the six hexacyclic systems are given.
Figures 1.1 shows the six molecules studied in this thesis.
Starting from a structural description of the carbohydrates, β-D-glucopyranose is represented in **Figure 1.1-(f)**. Further, in **Figure 1.2-(b)**, the conventional numbering of the heavy atoms for a general carbohydrate has been reported.

Figure 1.2 The numbering procedure for a general cyclohexane-like molecule (a) and for a general carbohydrate (b). Hydrogens have been omitted in both cases for clarity.

This sugar is a six-membered ring in the chair conformation $^4C_1$, composed of five carbon atoms (C1, C2, C3, C4 and C5) and one oxygen atom (O5) all connected by single bonds. A methyl alcohol group, CH_2OH, is attached to C5. Every other carbon of the ring carries both a hydroxyl group (OH) and a hydrogen atom. The OH attached to C1 can be either in an axial or an equatorial position with respect to the ring, giving rise to two possible anomers, α and β respectively. The anomeric ratio in aqueous solutions is β:α 64%:36% [38, 39]. Here, only the β anomer has been considered. The experimental X-ray crystal structure is available for β-D-glucose [47]. A subsequent refinement for the same structure is also available [48].

Like glucose, β-D-xylopranose (**Figure 1.1-(e)**) exists in the $^4C_1$ chair conformation in aqueous solution, as determined by NMR [49], and it has the same anomeric ratio as glucose [50]. Xylopyranose differs from glucopyranose because there are no OH and methyl alcohol groups on C5.

Structures for cyclohexanol, 1,2-cyclohexanediol and 1,4-cyclohexanediol (**Figure 1.1-(a)**, **(b)** and **(c)** respectively) are available in the Cambridge Structural Database [51-53] (their respective codes are: *QQQEJV08*, *PIWXIK* and *POVSEY*). They are all puckered in the $^4C_1$ chair conformation. *Ab initio* studies for cyclohexanol have reported that the preferred conformation of the hydroxyl group on C1 is equatorial [54]. This result is validated by the corresponding crystal structure.
While the 1,2-cyclohexanediol has the two OH groups bonded to adjacent carbon atoms in the ring (C1 and C2), in 1,4-cyclohexanediol, the two hydroxyl groups are positioned on C1 and C4. Their hydroxyl groups have been observed in both the *cis* and *trans* positions, with respect to the ring. More specifically, 1,4-cyclohexanediol shows a cis-trans ratio of 2:1 [53].

The structure of myo-inositol, represented in **Figure 1.1-(d)**, corresponds to the crystal structure [55]. Like the previous carbohydrates, this molecule puckers in the $^4C_1$ chair conformation. Each carbon of the ring carries both one OH group and one hydrogen. All the OH groups are found in equatorial positions, except for the OH on C2, which is in axial position.

Objectives

Ultrasound interferometer experiments and computer simulations of the dynamic hydration of six small hexacyclic systems in water were the main objectives of this dissertation. The former was aimed to give hydration number values of the six hexacyclic molecules in water, while the latter gave, in addition to dynamic hydration numbers calculations, more insight into the role played by each functional group on the hydration of each molecule. Moreover, from computer simulations, important information on the conformational changes (i.e. flexibility of the ring) of these molecules with time has been obtained.

In **Chapter 2**, theoretical notions of modelling and simulation methods applied in this dissertation, namely Molecular Mechanics (MM), Molecular Dynamics (MD), Quantum Mechanics (QM), and hybrid Quantum Mechanics/Molecular Mechanics (QM/MM), are given. More particularly, **Section 2.1** contains a discussion about Classical Methods (MM and MD). Quantum Mechanical methods are included in **Section 2.2**, while **Section 2.3** reports a brief description of hybrid systems, such as QM/MM. This overview is then followed by procedures for validating simulation results (**Section 2.4**) and by the details for performing a parameterisation procedure.

Chapter 3 gives an overview of the computational analytical techniques applied to characterise the six systems. From the two "traditional" *Pair* and *Spatial Distribution Functions,* usually applied for characterising solvent surrounding solute, a description of various time-dependant functions, such as residence time, auto and cross-correlation functions and linear diffusion, is given. In this chapter, the conformational analysis is also included: ring puckering, following the Cramer and Pople method, has been performed, and the theory behind this powerful tool is described. Furthermore, the importance of Free Energy Perturbation for validating MD results is presented, together with a related applicative example.

Chapter 4 is the experimental chapter, where the ultrasound interferometer technique is described. Through this technique, hydration numbers, n_H, have been obtained for all the six hexacyclic molecules studied in the thesis. The procedure for the calculation of n_H is reported in **Section 4.2**, after an overview of the theory behind the technique and the working set-up (**Section 4.1**). A statistical analysis of the data is given in **Section 4.3**. The errors on n_H experimental results have been computed by applying a Gaussian propagation of errors.

Chapter 5* reports the results obtained by applying the computational techniques and the experimental procedure explained in the first four chapters of this dissertation. The chapter is given in

the format of paper; it has been prepared to be submitted to the *Journal of Physical Chemistry B*. Experiments and simulations made possible to draw conclusion about the relationship between molecular flexibility and hydration.

In the concluding **Chapter 6**, the methods and the techniques discussed in the first part of the thesis are applied to another interesting class of molecules, strictly related to carbohydrates from a structural point of view, which are important as anti-diabetic drugs: glycosidase-inhibitors. A general overview of these compounds and their important role is given in **Section 6.1**, followed by the details of the parameterisation procedure, which is based on QM calculations (**Section 6.2**). Preliminary results on the static and dynamic water characterising iminosugars are reported (through pair distribution function, residence time and n_H). Ring puckering results, with related relaxation times of θ and ϕ angles, are also reported (**Section 6.3**).

This chapter gives an example of a possible extension of the procedure for obtaining hydration and conformational information to another important class of compounds. It also represents a suggestion for a future work: investigating the complete series of glycosidase-inhibitors and the influence of their functional groups on their inhibitory activity.

References

1. Esquivelzeta-Rabell, M., J. Peon, and G. Cuevas, *Rotational Diffusion of Dihydroxy Coumarins: Effect of OH Groups and Their Relative Position on Solute-Solvent Interactions.* Journal of Physical Chemistry B, 2009. **113**: p. 8599-8606.
2. Impey, R.W., P.A. Madden, and I.R. McDonald, *Hydration and Mobility of Ions in Solution.* Journal of Physical Chemistry, 1983. **87**: p. 5071-5083.
3. Ruckenstein, E., *On the Phenomenological Thermodynamics of Hydrophobic Bonding.* Journal of Dispersion Science and Technology, 1998. **19**(2).
4. Schravedijk, P. and N.F.A. Van der Veg, *From Hydrophobic to Hydrophilic Solvation: An Application to Hydration of Benzene.* Journal of Chemical Theory and Computation, 2005. **1**(4).
5. Kay, L.E., *Protein Dynamics from NMR.* Nature Structural Biology, 1998: p. 513-517.
6. Uedaira, H., et al., *Hydration of Oligosaccharides.* Bulletin of the Chemical Society of Japan, 1990. **63**(12): p. 3376-3379.
7. Engelsen, S.B., et al., *The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and NMR spectroscopy.* Biophysical Chemistry, 2001. **93**: p. 103-127.
8. Aeberhardt, K., et al., *Ultrasonic wave spectroscopy study of sugars oligomers and polysaccharides in aqueous solutions: The hydration lenght concept.* International Journal of Biological Macromolecules, 2005. **36**: p. 275-282.
9. Burakowsky, A. and J. Glinski, *Hydration numbers of non-electrolytes – Application of the acoustic method of Pasynski.* Chemical Physics, 2007. **332**: p. 336-340.
10. Corzana, F., et al., *A hydration study of 1,4 and 1,6 linked alpha-glucans by comparative 10 ns molecular dynamics simulations and 500MHz NMR.* Journal of Computational Chemistry, 2004. **25**(4): p. 573-586.
11. Gaida, L.B., C.G. Dussap, and J.B. Gros, *Variable hydration of small carbohydrates for predicting equilibrium properties in diluted and concentrated solutions.* Food Chemistry, 2006. **96**(3).
12. Galema, S.A. and H. Hoiland, *Stereochimical Aspects of Hydration of Carbohydrates in Aqueous Solutions. 3. Density and Ultrasound Measurements.* Journal of Physical Chemistry, 1991. **95**: p. 5321-5326.
13. Junquera, E., D. Olmos, and E. Aicart, *Carbohydrate–water interactions of p-nitrophenylglycosides in aqueous solution. Ultrasonic and densitometric studies.* Journal of Physical Chemistry Chemical Physics, 2002. **4**: p. 352-357.
14. Kiriukhin, M.Y. and K.D. Collins, *Dynamic hydration numbers for biologically important ions.* Biophysical Chemistry, 2002. **99**(2): p. 155-168.
15. Uedaira, H., M. Ikura, and H. Uedaira, *Natural-Abundance Oxygen-17 Magnetic Relaxation in Aqueous Solutions of Carbohydrates* Bulletin of the Chemical Society of Japan, 1989. **62**: p. 1-4.
16. Tait, M.J., et al., *Hydration of Monosaccharides: A Study by Dielectric and Nuclear Magnetic Relaxation.* Journal of Solution Chemistry, 1972. **1**(2): p. 131-151.
17. Andersson, C. and S.B. Engelsen, *The Mean Hydration of Carbohydrates as Studied by Normalized Two-Dimensional Radial Pair Distributions.* Journal of Molecular Graphics and Modelling, 1999. **17**(2): p. 101-105.

18.   Galema, S.A., *The effect of stereochemistry on carbohydrate hydration in aqueous solutions*, in *Faculty of Mathematics and Natural Sciences*. 1992, Ph.D. Thesis. University of Groningen: Groningen, The Netherlands.

19.   Galema, S.A., M.J. Blandamer, and J.B.F.N. Engberts, *Stereochemical Aspects of the Hydration of Carbohydrates. Kinetic Medium Effects of Monosaccharides on a Water-Catalyzed Hydrolysis Reaction*. Journal of American Chemical Society, 1990. **112**: p. 9666-9668.

20.   Iyer, P.V. and L. Ananthanarayan, *Enzyme Stability and Stabilization of Aqueous and Non-aqueous Environment*. Process Biochemistry, 2008. **43**(10): p. 1019-1032.

21.   Chalikian, T.V., A.P. Sarvazyan, and K.J. Breslauer, *Hydration and partial compressibility of biological compounds*. Biophysical Chemistry, 1994. **51**(2-3): p. 89-109.

22.   Israelachvili, J. and H. Wennerstrom, *Role of hydration and water structure in biological and colloidal interactions*. Nature, 1996. **379**(6562): p. 219-225.

23.   Pal, S.K., et al., *Biological water: femtosecond dynamics of macromolecular hydration*. Journal of Physical Chemistry B, 2002. **106**(48): p. 12376-12395.

24.   Zavitsas, A.A., *Properties of Water Solutions of Electrolytes and Non-Electrolytes*. Journal of Physical Chemistry B, 2001. **105**: p. 7805-7817.

25.   Enderby, J.E., *Ions Solvation via Neutron Scattering*. Chemical Society Reviews, 1995. **24**: p. 159-168.

26.   Mason, P.E., et al., *Neutron Diffraction and Simulation Studies of the Exocyclic Hydroxymethyl Conformation of Glucose*. Journal of Chemical Physics, 2006. **125**(22): p. 224505/1-224505/9.

27.   Mason, P.E., et al., *Neutron Diffraction and Computer Simulation Studies of D-Xylose*. Journal of American Chemical Society, 2005. **127**: p. 10991-10998.

28.   Mason, P.E., et al., *Determination of a Hydroxyl Conformation in Aqueous Xylose Using Neutron Scattering and Molecular Dynamics*. Journal of Physical Chemistry B Letters, 2006. **110**(7): p. 2981-2983.

29.   Mason, P.E., et al., *Neutron diffraction and simulation studies of the exocyclic hydroxymethyl conformation of glucose*. Journal of Chemical Physics, 2006. **125**(22): p. 224505/1-224505/9.

30.   Brady, J.W., et al., *Molecular dynamics and neutron diffraction studies of the structuring of water by carbohydrates and other solutes*. Modelling Molecular Structure and Reactivity in Biological Systems. 2006. 76-82.

31.   Tait, M.J., et al., *Hydration of Monosaccharides: A Study by Dielectric and Nuclear Magnetic Relaxation*. Journal of Solution Chemistry, 1974. **1**(2): p. 131-151.

32.   Smith, L.J., et al., *Molecular dynamics of glucose in solution: A quasielastic neutron scattering study*. Journal of Chemical Physics, 2004. **120**(8).

33.   Brady, J.W., *Molecular Dynamics Simulations of α-D-Glucose in Aqueous Solution*. Journal of American Chemical Society, 1989. **111**: p. 5155-5165.

34.   Hajduk, P.J., D.A. Horita, and L.E. Lerner, *Picosecond Dynamics of Simple Monosaccharides As Probed by NMR and Molecular Dynamics Simulations*. Journal of American Chemical Society, 1993. **115**: p. 9196-9201.

35.   Stokes, R.H. and R.A. Robinson, *Interactions in Aqueous Nonelectrolyte Solutions. I. Solute-Solvent Equilibriums*. Journal of Physical Chemistry, 1966. **70**(7): p. 2126-2131.

36.   Hinton, J.F. and E.S. Amis, *Nuclear Magnetic Resonance Studies of Ions in Pure and Mixed Solvents*. 1966.

37. Uedaira, H. and H. Uedaira, *Sugar-Water Interaction from Diffusion Measurements*. Journal of Solution Chemistry, 1985. **14**(1): p. 27-34.

38. Schmidt, R.K., M. Karplus, and J.W. Brady, *The Anomeric Equilibrium in D-Xylose: Free Energy and the Role of Solvent Structuring*. Journal of American Chemical Society, 1996. **118**(3): p. 541-6.

39. Ha, S., et al., *Solvent Effect on the Anomeric Equilibrium in D-Glucose: A Free Energy Simulation Analysis*. Journal of American Chemical Society, 1991. **113**: p. 1553-1557.

40. Miyajima, K., K. Machida, and M. Nakagaki, *Hydrophobic Indexes for Various Monosaccharides*. Bulletin of Chemical Society Japan, 1985. **58**: p. 2595-2599.

41. Walkinshaw, M.D., *Variation in the Hydrophilicity of Hexapyranose Sugars Explains Features of the Anomeric Effect*. Journal of Chemical Society Perkin Transactions 2, 1987. **2**: p. 1903-1906.

42. Dwek, R.A., *Glycobiology: Toward Understanding the Function of Sugars*. Chemical Reviews, 1996. **96**: p. 683-720.

43. Duus, J., H. Gotfrendsen, and K. Bock, *Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations*. Chemical Reviews, 2000. **100**: p. 4589-4614.

44. Molteni, C. and M. Parrinello, *Condensed Matter Effects on the Structure of Crystalline Glucose*. Chemical Physics Letters, 1987. **275**: p. 409-413.

45. Di Bari, M., et al., *Mean Square Fluctuations of Hydrogen Atoms and Water-Biopolymer Interactions in Hydrated Saccharides*. Biophysical Journal, 2001. **81**: p. 1190-1194.

46. Kuttel, M., J.W. Brady, and K.J. Naidoo, *Carbohydrate Solution Simulations: Producing a Force Field with Experimentally Consistent Primary Alcohol Rotational Frequencies and Populations*. Journal of Computational Chemistry, 2002. **23**(13): p. 1236-1243.

47. Ferrier, W.G., *The Crystal and Molecular Structure of β-D-Glucose*. Acta Crystallographica, 1963. **16**.

48. Jeffrey, G.A. and S.S.C. Chu, *The Refinement of the Crystal Structures of β-D-Glucose and Cellobiose*. Acta Crystallographica, 1968. **B24**.

49. Jeffrey, G.A., et al., *Neutron Diffraction Refinement of Partially Deuterated β-L-Arabinopyranose and α-L-Xylopyranose at 123K*. Acta Crystallographica, 1980. **B36**: p. 373-377.

50. Hoog, C. and G. Widmalm, *Free Energy Simulations of D-Xylose in Water and Methyl D-Xylopyranoside in Methanol*. Journal of Physical Chemistry B, 2001. **105**: p. 6375-6379.

51. Ibberson, R.M., et al., *Polymorphism in Cyclohexanol*. Acta Crystallographica, 2008. **B64**: p. 573-582.

52. Lloyd, M.A., et al., *Solid-state compounds of stereoisomers: cis and trans isomers of 1,2-cyclohexanediol and 2,3-tetra-lindiol*. Acta Crystallographica, 2007. **B63**: p. 433-447.

53. Loehlin, J.H., M. Lee, and C.M. Woo, *Hydrogen-bond Patterns and the Structures of 1,4-cyclohexanediol: 2:1 cis:trans-1,4-cyclohexanediol*. Acta Crystallographica, 2008. **B64**: p. 583-588.

54. Jansen, C., et al., *The energetically preferred orientation of the hydroxyl group in cyclohexanol. Ab initio and force field calculations*. Journal of Molecular Structure (Teochem), 1997: p. 398-399; 395-404.

55. Rabinowitz, I.N. and J. Kraut, *The Crystal Structure of Myo-Inositol*. Acta Crystallographica, 1964. **17**: p. 159.

Chapter 2
Modelling Molecular Systems

Molecular modelling is a field in chemistry concerned with the use of physical principles to simulate and predict the behaviour of chemical systems, in order to derive properties that are too difficult to obtain from experiments.

An overview of modelling and simulation methods used in this thesis, namely classical methods (Molecular Mechanics (MM) and Molecular Dynamics (MD)), and quantum methods (Quantum Mechanics (QM) and hybrid Quantum Mechanics/Molecular Mechanics (QM/MM)), is given in this chapter. They were used to characterize the six hexacyclic systems shown in **Figure 2.1**.
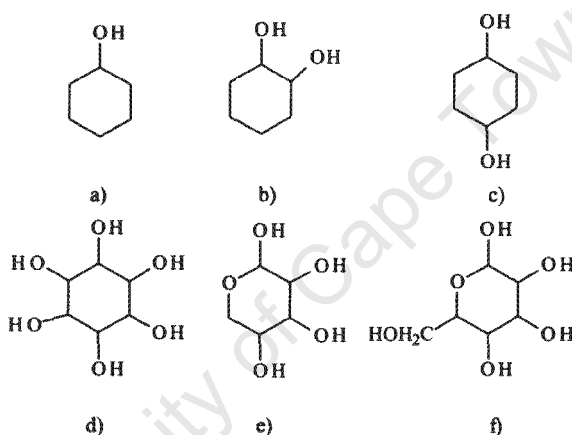
Figure 2.1. The six molecules studied in this thesis: a) cyclohexanol, b) 1,2-cyclohexanediol, c) 1,4-cyclohexanediol, d) myo-inositol, e) β-D-xylopyranose, f) β-D-glucopyranose. They have been studied through different computational methods, thus having information at different levels of detail.

Several texts on the matter are available [56-60], where details related to the following description can be found.

2.1 Classical Methods and Molecular Force Field

The description of systems on atomic and molecular scale is given by quantum mechanics. However, for systems consisting of a great number of atoms like biomolecules, quantum mechanical

calculations are still unrealistic, mainly due to the complexity of solving the Schrödinger equation for such big systems and the consequent large computer demand. The very first target in the strategy for simulating a molecular system is indeed to reduce computer time demand as much as possible and, at the same time, still perform an accurate simulation.

In this regard, classical methods offer a very good compromise for the two requirements.

Classical methods make use of some assumptions. They assume that the potential energy of an isolated molecule in the electronic ground state is uniquely determined if the coordinates of the constituents nuclei are known. This implies that it is possible to ignore the coordinates of the electrons, or more precisely that the latter will not influence the calculation of potential energy. This is possible by invoking the *Born-Oppenheimer* approximation: it states that the Schrödinger equation for a molecule can be separated into a part describing the motions of the electrons and another part describing the motions of the nuclei and that the two motions can be studied independently. This can be done because electrons are much lighter and consequently move much faster than the nuclei, so energy can be written in terms of nuclei position. Under this condition, the total wavefunction for a molecule can be written in the following form:

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

The total energy equals the sum of the electronic energy and the nuclear energy (the electrostatic repulsion between the positively charged nuclei), so that:

$$E_{tot} = E(electrons) + E(nuclei) \tag{2.2}$$

The potential energy of the molecule is then expressed by:

$$V = V(x_1, y_1, z_1, ..., x_N, y_N, z_N) \tag{2.3}$$

where the index N counts for the number of atoms in the molecule. Because only an isolated molecule is here considered, it follows that the forces acting on the individual atoms are the result of interactions between the atoms of the molecule only. This is expressed by the following equation:

$$F_{x_i} = -\frac{\partial V}{\partial x_i}, F_{y_i} = -\frac{\partial V}{\partial y_i}, F_{z_i} = -\frac{\partial V}{\partial z_i} \tag{2.4}$$

At the equilibrium, when the potential energy of the molecule is at its minimum, all the forces in equation **2.4** vanish:

$$F_{x_i} = F_{y_i} = F_{z_i} = ... = F_{z_N} = 0 \tag{2.5}$$

It is common to describe molecules in term of internal coordinates. As the potential energy of an isolated molecule is only the result of interaction between the atoms of the molecule, its value is independent by translation and rotation motions. Therefore, internal coordinates (bonds, lengths, torsion, etc.) describe the relation between potential energy and geometry.

If $q_1, q_2, q_3, ... q_K$ are the internal coordinates of the molecule, then the potential energy in equation 2.3 can also be rewritten as:

$$V = V(q_1, q_2, q_3, ... q_K) \tag{2.6}$$

where each internal coordinate q_P is also function of the Cartesian coordinates of the atoms:

$$q_P = q_P(x_1, y_1, z_1, ..., x_N, y_N, z_N) \tag{2.7}$$

The previous equation 2.4 can be then reformulated as (for x):

$$F_{x_i} = -\frac{\partial V}{\partial x_i} = -\sum_{P=1}^{K} \frac{\partial V}{\partial q_P} \frac{\partial q_P}{\partial x_i} = 0 \tag{2.8}$$

and the same applies for y and z.

Consequently, when one wants to describe the properties of a molecule in the equilibrium state, it is necessary to know the potential energy of the molecule as a function of the internal coordinates, and also the coordinate transformation between internal and Cartesian coordinates.

The potential energy of the molecule in terms of its internal coordinates is called *molecular force field*.

In other words, the force field is a combination of physical parameters able to give the energy of the system based on the coordinates of all the atoms and on information of the structure.

Molecular Mechanics (MM), Molecular Dynamics (MD) and Monte Carlo simulation methods are "traditionally" referred to as *force field methods*. In both MM and MD, the two most widely used molecular modelling methods, the energy of the system is calculated using a classical *ball and spring* model to represent the bonds and the interactions between atoms.

Regarding MD calculation, appropriately designed force fields can be very accurate. MD is an efficient and suitable choice for studying even very large, solvated complex systems. An advantage of the method is that a detailed time evolution of the molecular interactions and conformational changes is obtained.

In MD, successive configurations of the system are generated by integrating Newton's law of motion, over a short period of time, also known as the *timestep*. The result is a trajectory which shows how the

positions and the velocities of the particles vary over the time. More specifically, the trajectory is obtained by solving the differential equation (Newtons's second law):

$$\frac{d^2x}{dt^2} = \frac{F_{x_i}}{m_i} \tag{2.9}$$

From equation **2.9**, the motion of the particle of mass m_i along the coordinate x_i caused by a force F_{x_i} acting on the particle in that direction can be derived.

All the MD simulations in this thesis have been performed with the program CHARMM [61].
Other commonly used molecular dynamics packages are: AMBER [62], GROMOS [63] and DL_POLY [64].

The form of the CHARMM potential energy function used to calculate the potential energy V (with $V = V_{intra} + V_{inter}$) is reported in the following equations (**2.10-2.11**) [65, 66]. The intramolecular portion of the potential energy function, V_{intra}, is described by:

$$\sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi (1 + \cos(n\phi - \delta)) + \\ \sum_{\substack{improper \\ dihedrals}} K_\phi (\phi - \phi_0)^2 + \sum_{Urey-Bradley} K_{UB} (r_{1,3} - b_{1,3;0})^2 \tag{2.10}$$

While the intermolecular or external (or nonbonded) portion, V_{inter}, has the form:

$$\sum \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] \tag{2.11}$$

Equation **2.10** includes terms for the bonds, valence angles, torsion or dihedral angles, improper dihedral angles and a Urey-Bradley 1,3-term, where b_0, θ_0, ϕ_0, and $r_{1,3;0}$ are the bond, angle, improper, and Urey-Bradley equilibrium terms, respectively; n and δ are the dihedral multiplicity and phase and the all the Ks stands for the respective force constants. The intermolecular terms include a Lennard-Jones potential for the Van Der Waals interactions, $\left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right]$ and a Coulomb potential for electrostatic interactions, $\frac{q_i q_j}{4\pi D r_{ij}}$. q_i and q_j are the partial atomic charge of atom i and j, respectively, D is the dielectric constant, ε_{ij} is the well depth, $R_{min,ij}$ is the radius in the Lennard-Jones (LJ) 6–12 term, and r_{ij} is the distance between i and j.

bond strecthing

angle bending

bond rotation
(torsion)

Non-bonded interactions
(electrostatic)

δ^+ δ^+

δ^+

Non-bonded interactions
(van der Waals)
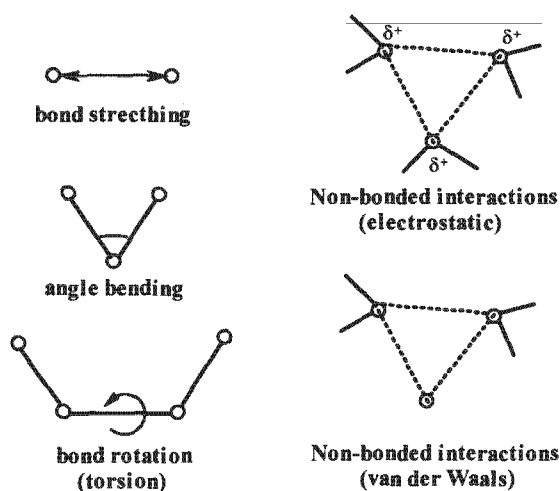
Figure 2.2 Representation of the main contributions to a force field: bond stretching, angle bending, torsional terms and non-bonded interactions (as depicted in reference [56]).

Summarizing, to completely define a force field, it is necessary to express all the parameters appearing in the previous equations. In other words, it is necessary to *parameterise* the force field.

Dynamics. Once all the parameters are available, a molecular dynamics simulation which involves the following steps can be performed:

• Minimization: an initial structure of the molecule that will be simulated is needed. This can come, in the best cases, from experimental data (crystal structure) or from a minimized ideal geometry (usually, this can be obtained from a molecule-builder program). Energy minimization is performed prior to dynamics to remove steric overlap and move the molecule to a lower potential energy;

• Heating: since the above minimisation only decreases potential energy while in the dynamics the kinetic energy is included as well, it is not possible to run a simulation without further steps - especially from a crystal structure - at room temperature; the kinetic energy might be too high, leading the simulation to crash. In a heating step, initial random velocities are assigned according to a Gaussian distribution appropriate for an initial low temperature and then dynamics can be run. The temperature is gradually increased by assigning greater random velocities to each atom at chosen time intervals. In this way, the total energy of the system is gradually increased;

• Equilibration: this is achieved by allowing the system to evolve spontaneously for a certain period of time (depending on the system) and integrating the equations of motion until the average temperature and structure remain stable.

• Production and analysis: in the final molecular dynamics simulation, CHARMM takes the equilibrated structure as its starting point. In a typical simulation, the outcome trajectory traces the motion of the molecule through a chosen period of time.

2.1.1 Ensembles

MD can be performed under conditions of constant number of particles (N), volume (V) and energy (E) (the so called *micro-canonical* or constant *NVE* ensemble). A *canonical* ensemble, taking constant N, V and T (temperature), can also be used, as well as an *isothermal-isobaric* ensemble, with fixed N, T, P (pressure) and a *grand-canonical* ensemble, with fixed μ (chemical potential), V, T.

In the *canonical*, *micro-canonical* and *isothermal-isobaric* ensembles the number of particles is constant, while in a *grand-canonical* simulation the composition can change.

In this thesis, all the simulations have been performed in *NPT* conditions.

2.1.2 Periodic boundary conditions and the Minimum Image Criterion.

MD simulations are most commonly used to model liquids rather than isolated molecules in the gas phase. In that case, one has to properly choose the environment conditions, which implies to choose the most representative number of molecules, the right temperature, the right volume etc. Usually, simulations are run in a box of molecules or atoms (unit cell); an example is depicted in **Figure 2.3** (which represents the box as used in this thesis). The box should have the right dimensions (sides a, b, and c) such that the total mass of the molecules included the box, divided by the volume, will give the right bulk density, or better the density that the same system would have in a "real" experiment performed under the same conditions of temperature and pressure.

One can argue that, by choosing such a way of simulating a system, the molecules closer to the boundaries of the box can experience different environment (such as different forces and interactions) in respect to the bulk molecules. To overcome this possibility, it is assumed that the system consists of infinite number of copies of the unit cell (which is also what happens in a "real" crystal, where the lattice is made of copies of the same unit cell).
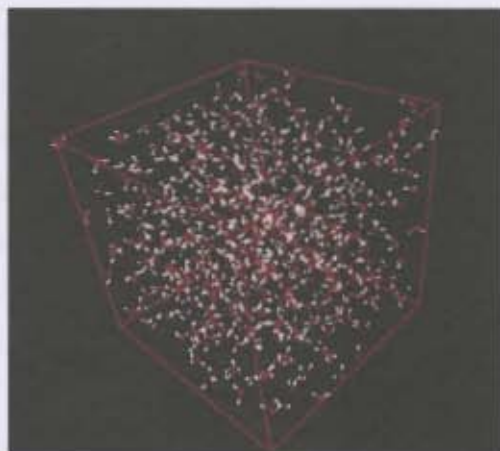
Figure 2.3 A representation of a 24.8626x24.8626x24.8626 Å box containing 512 water (TIP3P) molecules.

If during the MD simulation, a molecule or atom moves out of the original cell, another molecule (equivalent) will move into the cell to replace it (**Figure 2.4**) This "new" molecule comes from one of the adjacent cells. In this way, not only the total number of molecule in the original unit cell will remain constant, but it is also ensured that all the molecules, in particular the one at the boundaries, will experience the same forces of the molecules in the bulk. This procedure is known as *Periodic Boundary Conditions* (PBC).

In addition to the interactions among molecules in the same box, it is necessary to consider interactions between a molecule and other molecules in the surrounding cells. One can argue that, by having considered an infinite number of boxes surrounding the central box, a molecule in the central box will interact with an infinite number of surrounding molecules, by increasing enormously the time demand of the simulation. However, with the *minimum image convention*, only the interactions between a molecule *i* and its closest images are taken into account. The closest images are all the particles surrounding molecule *i* at the distance not greater that the side of the box chosen as unit cell (**Figure 2.5**).
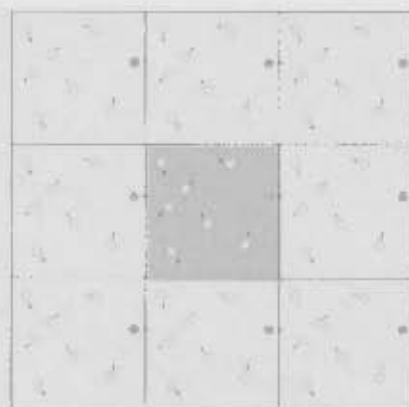
Figure 2.4. The *PBC* (in two dimensions) ensures that all the molecules on the border of the central box (highlighted in light blue) experience forces as if they were in the bulk fluid.
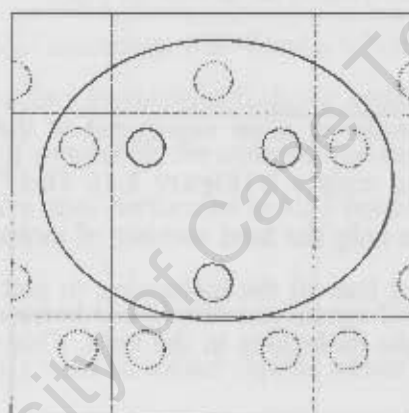
Figure 2.5. Representation of the *Minimum Image Convention*. This method requires that the interactions for a molecule *i* are only calculated between that molecule and closest image of every other molecule in the system.

2.1.3 The SHAKE Algorithm in Dynamics

The *timestep* to be used in MD simulations is chosen by considering the fastest motion present in the system. Bond vibrations for light atoms, such as hydrogens, are so fast that an extremely short *timestep* would be required to solve the equations of motion. However, these very high frequency motions are usually of little interest than the lower frequency modes, which often correspond to major conformational changes. For this reason, one is allowed to constraint the fastest motions.

By constraining them, the other degrees of freedom are not affected and stronger constraints allow to larger *timestep* (which decreases simulation time).

The most commonly used constraining method in MD simulations is the SHAKE algorithm [67]. This is a special technique for treating the dynamics of molecular system where some selected degrees of freedom (such as bond lengths) are constrained (usually hydrogen atoms, as said before). It is also used to keep fixed both bond lengths and angles for water molecules in the case of simulations in solution [68].

2.1.4 Carbohydrate Force Fields.

As mentioned in **Chapter 1**, carbohydrates are molecules especially difficult to model, mainly because of their flexibility, resulting in a myriad of possible conformations and configurations [43]. They also have highly flexible and polar OH groups, and tendency to form intramolecular bonds. Fortunately, several works have been contributing to set up appropriate parameters.

Carbohydrate parameter in CHARMM. The first carbohydrate parameter set in CHARMM was the HGFB force field, developed by Brady and co-workers [69]. This force field was parameterised using the CHARMM parameter optimisation feature to fit the experimental vibrational and crystallographic information for α-D-glucose and it was optimised for the TIP3P water model. Partial charges were based on standard CHARMM values. Then, Palma *et al.* [70] developed a force field (the PHLB), replacing the general dihedral angle term in the previous HGFB with a specific function (three terms for each dihedral: 1-, 2- and 3-fold cosine terms). The dihedral angles were parameterised to reproduce the experimental vibrational frequency data and to fit *ab initio* energy surfaces (by using small molecules – usually called fragments – as reference) [71].

Kuttel *et al.* [46] corrected the PHLB force field for the hydroxymethyl behaviour and from their work the new CSFF (Carbohydrate Solution Force Field) was delivered. It was validated by calculating the potential of mean force (PMF) profiles both in solution and vacuum for β-D-glucose and β-D-galactose. The PMF profiles confirm that the MD simulations using this force field will produce equilibrium population distributions and rotational frequencies in water that are consistent with experimental evidence. In addition, normal modes calculated for α-D-glucose in vacuum compare favourably with experimental data.

2.1.5 Water models.

Because of the aqueous environment of biological systems, water models have been developed to try to reproduce the properties of bulk liquid water. A very popular model used in biomolecular simulations is the TIP3P [72]. This treats H_2O as a rigid three-point system, with point charges on each of the atoms and Lennard-Jones parameters only on the oxygen. Although this water model is the most common and it has been traditionally used in the parameterisation procedure, especially in CHARMM, however it has been found that it does not accurately reproduce the second solvation shell seen experimentally in liquid water. The TIP4P and TIP5P [73] models use more interaction sites (4 or 5, respectively) and show a better second solvation shell [74]. Physical properties of these mentioned water models are reported in **TABLE 2.1**, while the following **Figure 2.5** reports their schematic representation. In this thesis, TIP3P water model for all the MD simulations as implemented in CHARMM has been used [72] and, in the case of QM/MM method, the $TIP4P_{EW}$, which allows to use the TIP4P water model with Ewald techniques [75].

TABLE 2.1: Properties of different water models most commonly used in a simulation.

	TIP3P	TIP4P	TIP4P$_{EW}$	TIP5P
r (OH) (Å)	0.9572	0.9572	0.9572	0.9572
∠HOH	104.52	104.52	104.52	104.52
q(O)	-0.834	0.0	0.0	0.0
q(H)	0.417	0.52	+0.52422	0.241
q(M)	-	-1.04	-1.04844	0.0
q(L)	-	-	-	-0.241
r(OM) (Å)	-	0.15	0.125	0.70
θ_{LOL}	-	-	-	109.47
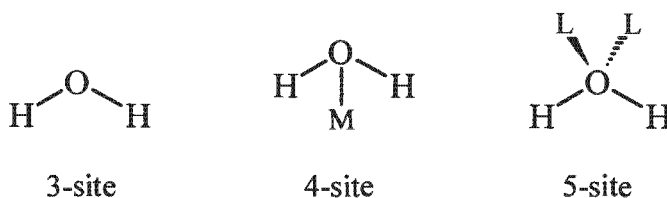
Figure 2.5. Schematic representation of the most widely used water models. Their physical properties are reported in TABLE 2.1.

2.2 The Quantum Mechanical "*ab initio*" Method (QM)

In Latin, *ab initio* means "*from the beginning*": these methods are based indeed on first principles and fundamental physical constants (like Planck constant, mass and charge of elementary particles, speed of light, etc.) without referring to experimental data. Because they are very reliable although very slow, they are applied to the study of small and medium-sizes molecules.

A QM method is based on the time dependant Schrödinger equation.

For a particle with mass m travelling in one-dimension, the Schrödinger equation is:

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Psi(x)+V(x)\Psi(x) = E\Psi(x) \tag{2.12}$$

$\psi(x)$ is the wavefunction describing the state of the particle, $V(x)$ is the potential energy, and E is the total energy of the particle. In three dimensions, equation 2.12 becomes:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(\vec{x})+V(\vec{x})\Psi(\vec{x}) = E\Psi(\vec{x}) \tag{2.13}$$

or, in a shorter notation:

$$\hat{H}\Psi = E\Psi \tag{2.14}$$

where $\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V$ is the Hamiltonian. An exact solution to equation 2.14 is possible only for single-electron systems, such as the hydrogen atom, while multi-electron systems do not admit any exact solution.

However, a highly accurate solution can be obtained in the latter case, by making certain approximations which are described in the following paragraphs.

24

2.2.1 Molecular Orbitals and the Variation Principle

Valence bond theory (VB) was the first developed quantum mechanical theory of bonding. It assumes that bonds are formed by localised electron pairs. The wavefunction, Ψ, describing the bond is given by the following equation:

$$\Psi(r_1, r_2) = \psi_1(r_1)\psi_2(r_2) \tag{2.15}$$

where electron 1 resides on nucleus 1 and electron 2 on nucleus 2 or more precisely, considering that electrons are indistinguishable, as:

$$\Psi(r_1, r_2) = \psi_1(r_1)\psi_2(r_2) + \psi_1(r_2)\psi_2(r_1) \tag{2.16}$$

In VB theory all the basilar concepts of spin pairing, s- and p-bonds, and hybridisation are already encoded.

Later, the Molecular Orbital (MO) theory considered that electrons occupy an orbital that spreads all over the complete molecule. The approximation used in this case is that of a linear combination of atomic orbitals (LCAO), which gives, for the hydrogen molecule:

$$\varphi(r) = c_1\psi_1(r) + c_2\psi_2(r) \tag{2.17}$$

where ϕ refers to a molecular orbital, while ψ stands for an atomic orbital, and c_1 and c_2 are constants.

In the simple hydrogen molecule, one has that $c_1 = c_2$, but for molecules formed by atoms of different nature this is not true.

By multiplying Equation 2.14 by the complex conjugated, $\varphi*$, of the wave function and integrating over all the possible positions, the energy of the electron can then be calculated as:

$$E = \frac{\int \varphi*(r)\hat{H}\varphi(r)dr}{\int \varphi*(r)\varphi(r)dr} \tag{2.18}$$

By substituting equation **2.17** in **2.18**:

$$E = \frac{c_1^2 \int \psi_1 \hat{H}\psi_1 dr + 2c_1c_2 \int \psi_1 \hat{H}\psi_2 dr + c_2^2 \int \psi_2 \hat{H}\psi_2 dr}{c_1^2 \int \psi_1^2 dr + 2c_1c_2 \int \psi_1\psi_2 dr + c_2^2 \int \psi_2^2 dr} \tag{2.19}$$

All the atomic constants c_i and the atomic orbitals in equation **2.19** are considered to be real.

A note has to be done on the Hamiltonian $\hat{H}$: it is a *Hermitian* operator, which implies:

$$\int \psi_1 \hat{H}\psi_2 dr = \int \psi_2 \hat{H}\psi_1 dr \tag{2.20}$$

For simplifying equation **2.19**:

$$H_{11} = \int \psi_1 \hat{H} \psi_1 dr \tag{2.21}$$

and the same is for H_{22}. H_{11} and H_{22} which are called Coulomb integral. H_{11} represents the energy of an electron occupying its own atomic orbital. Again, in the simple case of the hydrogen molecule, $H_{11}=H_{22}$.

$$H_{12} = \int \psi_1 \hat{H} \psi_2 dr = \int \psi_2 \hat{H} \psi_1 dr = H_{21} \tag{2.22}$$

H_{12} is a *resonance* integral and represents the energy of an electron moving under the influence of both nuclei. Usually, both Coulomb and resonance integrals give a negative value since bond systems are considered.

$$S_{12} = S_{21} = \int \psi_1 \psi_2 dr \tag{2.23}$$

is the *overlap* integral, whose value is positive or zero, being $\psi_{1,2}$ definite positive (and normalised to 1, i.e. $\int \psi_i^2 dr = 1$).

Consequently, equation **2.19** becomes:

$$E = \frac{c_1^2 H_{11} + 2c_1 c_2 H_{12} + c_2^2 H_{22}}{c_1^2 + 2c_1 c_2 S_{12} + c_2^2} \tag{2.24}$$

The variation theorem states that the energy calculated by equation **2.24** will always be greater than, or equal to, its lowest energy eigenvalue. Therefore, even if it is not possible to solve the Schrödinger equation (but only numerically approximate the solution), it is instead possible to choose the constants c_1 and c_2 such that the calculated energy is as close as possible to the lowest (unknown) energy eigenvalue.

This means finding the minimum of the energy with respect to the two constants c_1 and c_2, which can be done by solving the equation **2.25**:

$$\frac{\partial E}{\partial c_1} = 0 \text{ and } \frac{\partial E}{\partial c_2} = 0 \tag{2.25}$$

By applying equation **2.25** to the previous **2.24**, one derives:

$$\begin{aligned} c_1(H_{11} - E) + c_2(H_{12} - E) = 0 \\ c_1(H_{12} - ES_{12}) + c_2(H_{22} - E) = 0 \end{aligned} \tag{2.26}$$

known as *secular* equations.

The LCAO notation discussed previously has to be refined, considering the spin of the electrons. Each electron in any orbital ϕ is characterized by a spin quantum number $\frac{1}{2}$ and, in the presence of

26

an external magnetic field, there are two possible states corresponding to an alignment along or against the field: α and β will denote the two states. The total wave function of an electron is given by:

$$\phi(r_i, \delta_i) = \phi(r_i)\delta_i \qquad (2.27)$$

where r_i is the spatial coordinate of electron i and $\delta_i = \alpha$ or β is its spin coordinate.

Then, one can apply the simplifying notation $\vec{x}_i$ for spatial and spin coordinates:

$$\phi(\vec{x}_i) = \phi(r_i, \delta_i) \qquad (2.28)$$

The *Pauli Exclusion Principle* states that two electrons with the same wavefunction cannot exist which means that in a multi-electron system and assuming that each molecular orbital is fully occupied (this is known as *closed shell* system) a spatial orbital is expected to have two electrons with opposite spin.

The total wavefunction of a system is expected to be asymmetric with respect to the exchange of the two electrons:

$$\Psi(x_1, x_2, ..., x_i, x_j, ..., x_N) = -\Psi(x_1, x_2, ..., x_j, x_i, ..., x_N) \qquad (2.29)$$

This can be ensured by constructing the electron wavefunction from a *Slater determinant*. For a system of N-electrons, this is given by:

$$\Phi_0 = \frac{1}{\sqrt{N!}} \cdot \begin{vmatrix} \phi_1(\vec{x}_1) & \phi_2(\vec{x}_2) & ... & \phi_N(\vec{x}_N) \\ \phi_1(\vec{x}_2) & \phi_2(\vec{x}_2) & ... & \phi_N(\vec{x}_2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \phi_1(\vec{x}_N) & \phi_2(\vec{x}_N) & ... & \phi_N(\vec{x}_N) \end{vmatrix} \qquad (2.30)$$

where the same spin-orbital appears in the columns, and the same electron in the rows; $\dfrac{1}{\sqrt{N!}}$ is a normalisation factor.

As an approximation, one can assume that the electronic wavefunction can be written as a Slater determinant: $\Psi = \Phi_0$. This is not always true, and more explicitly it is not true for *open shell* systems. In the latter case, one would use:

$$\Psi = \sum_{k=0}^{K} a_k \Phi_k \qquad (2.31)$$

The spin-orbitals are orthonormal:

$$\langle \phi_i(x) | \phi_j(x) \rangle = \delta_{ij} \tag{2.32}$$

and so are the spinfunctions:

$$\langle \alpha | \alpha \rangle = \langle \beta | \beta \rangle = 1 \text{ and } \langle \alpha | \beta \rangle = 0 \tag{2.33}$$

2.2.2 Hartree-Fock Self Consistent Field.

The electronic energy of a system of N particles can be calculated from:

$$E = \langle \Phi_0 | \hat{H} | \Phi_0 \rangle \tag{2.34}$$

where the wavefunction Φ_0 is canonically normalised.

It can be shown from Equation **2.13** that:

$$E = 2 \sum_i^{N/2} h_i + \sum_i^{N/2} \sum_j^{N/2} (2J_{ij} - K_{ij}) \tag{2.35}$$

where:

$$h_i = \langle \phi_i(r_1) | \hat{h}_1 | \phi_i(r_1) \rangle \tag{2.36}$$

where $\hat{h}_i = -\frac{1}{2} \nabla_{r_i}^2 - \sum_k^N \frac{Z_k}{r_{ik}}$ and h_i is a one-electron contribution and it is known as *core* integral. The factor in front of the summation in equation **2.35** refers to the contribution of two electrons in the same orbital.

The term J_{ij} in equation **2.35** is known as *Coulomb* integral:

$$J_{ij} = \langle \phi_i(r_1)\phi_j(r_2) | \hat{g}_{12} | \phi_i(r_1)\phi_j(r_2) \rangle \tag{2.37}$$

where $\hat{g}_{ij} = \frac{1}{r_{ij}}$. A Coulomb integral represents the classical repulsion between two electron charge distributions.

The term K_{ij} is the *exchange* integral:

$$K_{ij} = \langle \phi_i(r_1)\phi_j(r_2) | \hat{g}_{12} | \phi_i(r_2)\phi_j(r_1) \rangle \tag{2.38}$$

This integral has no classical analogue. This is a quantum mechanical manifestation of the fact that the electrons are indistinguishable and that the wavefunction must be antisimmetric. Only electron pairs in different orbitals but with equal spin contribute.

2.2.3 The Hartree-Fock Equation.

The method discussed in **Paragraph 2.2.2** is necessary when finding a set of orbitals which makes the energy a minimum with respect to variations in the wavefunction Φ_0.

To simplify, an operator can be introduced. $\hat{F}_1$ is called *Fock* operator and it is defined as following:

$$\hat{F}_1 = \hat{h}_1 + \sum_{j}^{N/2}(2\hat{J}_j - \hat{K}_j) \tag{2.39}$$

Now the Coulomb operator and the exchange operator are defined as:

$$J_j\phi_i(r_1) = \left\langle \phi_j(r_2)\left|\hat{g}_{12}\right|\phi_j(r_2)\right\rangle \cdot \phi_i(r_1) \tag{2.40a}$$

and

$$K_j\varphi_i(r_1) = \left\langle \phi_j(r_2)\left|\hat{g}_{12}\right|\phi_i(r_2)\right\rangle \cdot \phi_j(r_1) \tag{2.40b}$$

The Fock operator is a one-electron operator, which describes the kinetic energy of an electron, the attraction to all the nuclei (both through $\hat{h}$), and the repulsion to all other electrons (through the two operators $\hat{J}_j$ and $\hat{K}_j$). The summation over the Coulomb and exchange operators represents a one-electron approximation to the behaviour of one electron in the field of the others.

Applying the variation method means that the variations in equation **2.41** are considered:

$$E = \left\langle \Phi_0\left|\hat{H}\right|\Phi_0\right\rangle \tag{2.41}$$

under the constraints that all the individual orbitals, $\phi_i(r)$ are orthonormal.

Skipping the mathematical details, the process of finding orbitals that yield a minimum on the energy results in a set of equations:

$$\hat{F}_i\phi_i(\vec{r}_i) = \varepsilon_i\phi_i(\vec{r}_i) \tag{2.42}$$

known as the Hartree-Foch equation, where the Hartree-Fock energy is:

$$E = 2\sum_{i}^{N/2}\varepsilon_i - \sum_{i}^{N/2}\sum_{i}^{N/2}(2J_{ij} - K_{ij}) \tag{2.43}$$

In order to try to solve the previous equation **2.43**, different approaches have been developed. In each of the possible approaches, the main challenge is the construction of an appropriate basis set using a *linear combination of atomic orbital*.

There are many types of basis sets being used in the different methods of computation. Here, a discussion about the basis sets in general, followed by the description of a specific basis set, i.e. the notation developed by Pople and co-workers, are reported.

2.2.4 Basis Sets. Slater and Gaussian Type Orbitals.

Assuming that the atomic orbital can be described by the product of a radial and an angular functions, it is given by:

$$X(\vec{r}) = R_{nl}(r)Y_{lm}(\theta,\phi) \tag{2.44}$$

where Y_{lm} is a spherical harmonic function, an angular part which describes the shape of the orbitals; r, ϕ and θ are the spherical coordinates; R_{nl} is the radial part of the basis function; n, l, and m are the principal, angular momentum and magnetic quantum numbers, respectively.

To simplify, the radial part R_{nl} is usually expressed as a product of polynomial and exponentially decaying functions. This gives a form:

$$X_S(\vec{r}) = X_{\bar{R}\zeta nlm}(\vec{r}) = Nr^{n-1}e^{-\zeta r}Y_{lm}(\theta,\phi) \tag{2.45}$$

where $\vec{R}$ is the radial function expression for a specific atom, N is the normalization constant and ζ is the exponent. The subscript S indicates that the orbital is a *Slater orbital*.

The functions in **2.45** are the so-called *Slater type orbitals* (STOs): they lead to complex equations, which make calculations complicated and very slow. For this reason, STOs are mainly used for atomic and diatomic systems only, where high accuracy is required, and in semi-empirical methods. Most *ab initio* methods are carried out by performing the *Gaussian type orbitals* (GTOs).

This is possible because the shape of STOs functions can be approximated using a summation of a number of GTOs with different exponents and coefficients.

The GTOs can be written in terms of polar (**2.46**) and Cartesian coordinates (**2.47**):

$$X_g(\vec{r}) = X_{R\alpha nlm}(\vec{r}) = Nr^{n-1}e^{-\alpha r^2}Y_{lm}(\theta,\phi) \tag{2.46}$$

$$X_g(\vec{r}) = X_{R\alpha nlm}(\vec{r}) = Ne^{-\alpha r^2}x^l y^m z^n \tag{2.47}$$

where α is the exponent and the subscript g indicates that the orbital is a Gaussian type orbital.

From equation **2.46** one derives that the GTOs differ from the previous STOs only in the value of their exponents, namely GTOs depend on r^2 rather than r.

2.2.5 A note on primitive Gaussians type of orbitals.

The contracted Gaussian type of orbitals (CGTO) can be obtained from a linear combination of primitive (uncontracted) Gaussians, mentioned in the previous paragraph:

$$X(CGTO) = \sum_{i}^{K} a_i X_i (PGTO) \qquad (2.48)$$

In equation **2.48**, X is a contracted orbital, PGTOs are primitive Gaussian type orbitals, and a_i are the contraction coefficients.

Most energy optimised basis sets have the limitation that the wavefunction depends mainly on the inner shell electrons. The foundation of contracted basis sets aims to overcome the fact that many basis functions go into the description of the energetically important but chemically unimportant core electrons.

There are two different way of contracting a set of PGTOs into a CGTOs, namely *segmented* and *general* contractions. The segmented are the most widely used and the Pople method is one of them. The latter has been applied in this thesis and thus will be briefly described in the coming paragraph.

Pople basis sets. In this method, basis sets are denoted as the general notation *K-nlG* or *K-nlmG* where K indicates how many primitive Gaussian type of orbitals (the above mentioned PGTOs) are used for representing the core orbitals and the *n,l* or *n,l,m* indicate both how many functions of the valence orbitals are split into and how many PGTOs are used for their representation.

One simple example is:

• 6-31G: in this notation, the core orbitals are a contraction of six primitive Gaussian orbitals (PGTOs), the inner part of the valence orbital is a contraction of three PGTOs and the outer is represented by only one PGTO.

Pople notation has been improved as well for the case of treating anions and molecules with lone pairs, by adding a diffuse function, which is denoted as *K-N l ++ or K-N l m G ++* .

2.2.6 Electron Correlation Methods.

It is important to highlight that the Hartree-Fock method can give molecular property calculation results that can be accurate enough to be compared with experimental results for small and medium-sized molecules. On the other hand, it is also true that, when one wants to study mechanisms sensitive to energy changes (bond dissociation energy, ionization potential, electronic affinities, etc.), this method doesn't consider electron correlation.

To simply describe electron correlation, one can think that, in a real system, the motion of electrons are correlated and they move far apart from each other because this allows them to have an energy value that is lower than the one obtained through the Hartree-Fock method. Therefore, to consider electron correlation is very important in improving calculations.

Also in the case of electron correlation different approaches are possible.

Because this dissertation has been focused on the application of a particular electron correlation approach, i.e. the *Møller-Plesset Perturbation Theory*, the basic concepts related to this approach will be briefly reported herein.

As the name suggests, the electron correlation parameter is added via a perturbation from the Hartree-Fock equation. Therefore, in this method, the true Hamiltonian is described as the sum of a zeroth-order Hamiltonian $\hat{H}_0$ plus a perturbation $\hat{V}$:

$$\hat{H} = \hat{H}_0 + \hat{V} \tag{2.49}$$

The unperturbed Hamiltonian $\hat{H}_0$ is obtained by the summation of the one-electron Fock operators for the N electrons:

$$\hat{H}_0 = \sum_{i=1}^{N} \hat{F}_i = \sum_{i=1}^{N} (\hat{h}_i + \sum_{j=1}^{N} (\hat{J}_j - \frac{\hat{K}_j}{2})) \tag{2.50}$$

The value of the true Hamiltonian is described by the sum of the nuclear attraction and repulsion terms:

$$\hat{H} = \sum_{i=1}^{N} (\hat{H}_i^{core}) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{r_{ij}} \tag{2.51}$$

Now the perturbation term $\hat{V}$ needs to be known. This is given by the difference between **2.50** and **2.51**, which gives (being $\hat{H}^{core} = \hat{h}$):

$$\hat{V} = \sum_{i=1}^{N} (\sum_{j=i+1}^{N} \frac{1}{r_{ij}} - \sum_{j=1}^{N} (\hat{J}_j - \frac{\hat{K}_j}{2})) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{r_{ij}} \tag{2.52}$$

Equation **2.52** implies that the energy calculated with the 1st order Møller-Plesset perturbation theory gives the Hartree-Fock energy.

2.3 Hybrid System Quantum Mechanics/Molecular Mechanics (QM/MM).

This is a combination of methods which treats different parts of a system at varying levels of precision and, as the name suggests, make use of both classical and quantum methods, described above.

Therefore, in the specific case of a hybrid system QM/MM, the accuracy of a quantum mechanical description is coupled with the low computational cost of molecular mechanics. As results, one can treat with a great accuracy (at QM level) a localised region of the system and treat the rest at MM level (usually portion of the system that are not directly involved in the chemical and physical properties of interest and/or the surrounding water), as schematically represented in **Figure 2.6**.

Figure 2.6. Representation of a hybrid QM/MM system.

The Hamiltonian for a hybrid QM/MM system (H_S) is described by the following equation:

$$H_S = H_{QM} + H_{MM} + H_{QM/MM} \tag{2.53}$$

where H_{QM} is the quantum potential energy of the quantum part of the system, H_{MM} describes the MM potential for the MM part of the system and the interaction between the QM and MM components of the system are described by the $H_{QM/MM}$ [76].

An increasing development of algorithms for hybrid systems has recently occurred. This is mainly due to the fact that MM and MD force fields neglect the changes in the electronic structure of a system. Such changes in electronic structure in processes that involve bond-breaking and bond-forming, charge transfer, and/or electronic excitation, require quantum mechanics (QM) for a proper treatment. However, due to the high demanding computational cost, the application of QM is still limited to relatively small systems consisting of up to tens or several hundreds of atoms, or even

smaller systems when the highest levels of theory are employed. For all these reasons, combined QM/MM methods are now the most complete answer.

2.4 Parameterisation of Simulations Terms and Their Validations Against Experimental and Theoretical Estimates.

2.4.1 Parameterisation Procedure.

When performing a simulation by using classical methods, it could be necessary to check the reliability of the employed parameters. Because it is necessary to run dynamic production starting from the right structure of the molecule of interest, it follows that having the right connection of atoms, conformation and geometry is of crucial importance. In addition, one needs to specify many more parameters - torsion angles and dihedrals, electrostatic interactions, etc. - as explained in **Paragraph 2.1**.

In some lucky cases, one has all the necessary experimental data available (X-ray structure of the molecule of interest, plus previously validated parameters) and, ideally, even experimental data to validate the final results of the simulation.

In some other cases, it is necessary to begin from scratch, which could clearly represent a source of error (i.e. not well minimized initial geometry, wrong bond lengths or angles values). In this second case, going through a parameterisation procedure is a necessary step, in order to obtain all the missing terms necessary for the simulation. The most widely used parameterisation approaches are discussed herein. When the parameterisation step is completed, a further analysis is needed: the validation of the parameterised terms, which is discussed in the following **Paragraph 2.4.2**.

Once a functional form for the force field has been chosen and the data to be used in the parameterisation identified, the so called *by analogy* approach [77] is usually applied to obtain parameters.

In this approach, one chooses a similar fragment (or an entire molecule) close to the target (the molecule to be simulated). Of course, the geometry, angles and dihedrals could be different between the two cases. However, one of the principal rules in a parameterisation choosing the *by analogy* procedure relies on considering the single fragments forming a molecule as nearly independent from the whole geometry of the molecule. Therefore the parameters of a fragment derived for one molecule

are suitable for other molecules having the same fragment (see **Chapter 6**, **Section 6.2** for an applicative example of this procedure).

The parameters are gradually refined to give better and better fits to the data. In this case, it could be difficult to simultaneously modify a large number of parameters in such a strategy. Therefore, it is usual to perform the parameterisation in stages. It could be necessary to get information by experimental data (i.e. in the case of torsional barriers for the torsional function) – when available – or by quantum mechanical calculations. The strategy in these two cases is the following:

- a molecular fragment that adequately represents the bond of interest and its immediate environment is chosen;
- a series of structure are generated by rotating around the bond and their energies determined using quantum mechanics.

The torsional potential is then fitted to reproduce the energy curve, in conjunction with the Van der Waals potential and partial charges.

The error is given by the sum of squares of the differences between the observed and calculated values for the set of properties. The objective is to change the force field parameters to minimize the error.

2.4.2 Validation.

Once the parameterisation procedure is chosen and started, in order to verify the reliability of the parameters at each step of the parameterisation, different methods of validation can be applied and the criteria to choose the suitable one strictly depend on the nature of the needed information. In this regard, it is possible to distinguish within two main classes of terms (as previously explained in **Paragraph 2.1.1**):

• Intramolecular (internal, bonded), which include terms for the bonds, valence angles, torsion or dihedral angles, improper dihedral angles and a torsion potential (this will described in the following **Paragraph 2.4.3**).

• Intermolecular (external, non-bonded) terms, which account for interactions between all pairs of atom (i and j) that are in different molecules or that are in the same molecule but separated by at least three bonds. These are usually modelled using a Coulomb potential term for electrostatic interactions and a Lennard-Jones potential for Van der Waals interactions (description about these terms is included in the following **Paragraph 2.4.5**).

While the intramolecular terms are relatively simple to obtain, because they are all observables and they can be compared with experimental values, the intermolecular ones require more efforts.
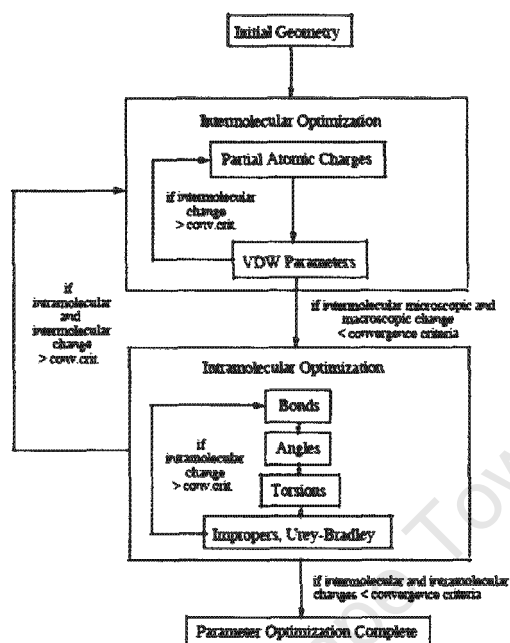
Figure 2.7. General parameterisation procedure scheme, for both bonded and non-bonded terms, as depicted in reference [66]. At each step, it is necessary to check if the obtained parameters are close enough to the target values (which can be either experimental or calculated quantum mechanics values).

However, the accuracy of the simulations results depends on both. Consequently, they need to be carried out carefully.

2.4.3 Bonded parameters.

Validation of bonded parameters can be reached through experiments and different techniques are suitable for this aim.

Because of their nature, the most widely used validation tool is the Infra-Red (IR) spectroscopy, from which a normal mode analysis can be obtained.

This procedure has been applied in this dissertation, in particular in the study of N-sugars (**Chapter 6**); for these compounds indeed, the entire parameterisation set was not available. Therefore

vibrational analysis represented a necessary source of comparison and optimisation of bonded parameters (the computational details of the technique are described in major details in **Paragraph 2.4.4**).

X-Ray is also widely exploited for validation purposes, as it gives bond lengths values and spatial geometries with very high resolution (one can think of *state of the art* synchrotron sources, which produce data better than 1.0Å in resolution). Further, on-line crystallographic data-bases collect an enormous number of X-Ray structures. This undoubtedly represents an optimal starting point for modelling structures. Moreover, all the simulation programs packages are able to read X-ray structures directly from on-line data bases, which makes the procedure easier.

2.4.4 Infra-Red.

IR spectroscopy is one of the commonly used spectrophotometric methods of analysis. It makes use of infrared radiation and its interaction with matter.

Since organic and inorganic compounds absorb various frequencies in the electromagnetic radiation spectrum which lie in the IR region, this technique is considered a very powerful tool for studying the chemical nature and structural properties of many compounds.

The IR region cover the wavenumber range of about 14000 to 10 cm^{-1} (or in micrometers: 0.70 – 1000 μm) and this whole region is divided into three sub-regions: the near-infrared, from 14000 to 4000 cm^{-1} (or 0.70-2.0 μm), the middle-infrared region, from 4000 to 200 cm^{-1} (or 2-6 μm) and the far-infrared region, 200-10 cm^{-1} (or from 6 -1000 μm).

An IR radiation causes a vibration transition. When atoms and molecules absorb such radiation, they get excited from a lower vibrational state (ground state) to a higher state. These transitions are ruled by precise quantum mechanical restrictions.

Selection rules for IR absorption. The selection rule for IR absorption can be determined by evaluating the transition moment, R. This moment refers to the electric dipole moment transitions of the interaction of the oscillating electric vector of the radiation field with the electric dipole moment of an atom or molecule. R is calculated from wave function:

$$R = \int \psi_i * \mu \psi_j d\tau \qquad (2.54)$$

2.54 refers to a transition between states *i* and *j*.
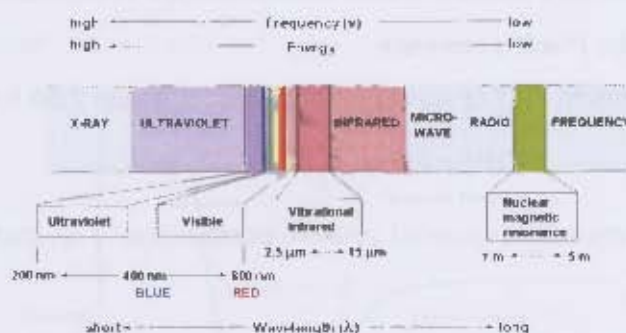
Figure 2.8 The electromagnetic spectrum. InfraRed (IR) region is a part of the electromagnetic spectrum, just beyond the visible region. It is divided into three sub-regions: the near-infrared, from 14000 to 4000 cm^{-1} (or 0.70-2.0 μm), the middle-infrared region, from 4000 to 200 cm^{-1} (or 2-6 μm) and the far-infrared region, 200-10 cm^{-1} (or from 6-1000 μm).

μ is the electric dipole moment operator and for vibrational motion can be expressed as:

$$\mu = \mu_0 + (r - r_e)(\delta\mu/\delta r)_0 + \frac{1}{2}(r - r_e)^2(\delta^2\mu/\delta r^2)_0 + \ldots \tag{2.55}$$

where μ_0 is the permanent dipole moment, r is the distance between two nuclei and r_e is the equilibrium bond distance.

If in equation **2.54** the wave functions have symmetry it gives R—0 and no vibrational transition is observed. In the opposite case, the transitions involved are allowed transitions.

This is mathematically expressed in equation **2.56**, where by neglecting all but the first two terms in equation **2.55** and then substituting the resulting expression for μ in equation **2.54**, one can obtain:

$$R = \int \psi_i * \lfloor \mu_0 + (r - r_e)(\delta\mu/\delta r)\rfloor \psi_j d\tau \tag{2.56}$$

Equation **2.56** shows that there must be a change in the dipole moment of the molecule during the vibration in order to observe absorption of the IR radiation

When a displacement of the bonds of a molecule occurs and they are far from their equilibrium bond distance r_e, these motions are described as harmonic oscillations. Due to the quantum mechanical restrictions, such molecules can assume only certain value of vibrational energy and from the Schrödinger equation for a harmonic oscillator, the allowed energy states are:

$$E_v = h\nu(V + \frac{1}{2}) \tag{2.57}$$

where V is the vibrational quantum number, whose values can be $V=0,1,2\ldots$, while ν is the vibrational frequency in Hertz and h is the Plank's constant.

Expressing the energy in equation **2.57** in wave number units, equation **2.58** follows:

$$T(V) = \widetilde{\nu}(V + \tfrac{1}{2}) \qquad\qquad (2.58)$$

where $\widetilde{\nu}$ is the vibrational frequency expressed in wave number (cm^{-1}) or, that is equivalent, $\widetilde{\nu} = 1/\lambda$, with λ the wavelength in cm.

The basic quantum mechanical selection rule for an harmonic oscillator permits vibrational energy transitions between adjacent states only, with $\Delta V = \pm 1$.

Any molecule of N-atoms has 3N numbers of degree of freedom of motion. These are a sum of the translational, rotational and vibrational motions. According to this, if one wants to predict how many vibrational motions characterise a molecule, the rule to apply is as following: linear polyatomic molecules will undergo 3N-5 vibrational modes, while non-linear molecules 3N-6. Indeed, it is necessary to subtract from the total 3N degrees of motion the non-vibrational motions.

Simulation of IR Spectra. When one has computed an IR spectrum, the following step is to assign each peak to the specific fragment of the molecule.

This can be difficult, because an IR spectrum is usually formed by a large number of peaks, even for very small and simple molecules. However, various tables and texts are helpful in these cases: the most important vibrational frequencies of the main functional groups have been collected and, even before computing the spectrum, it is possible to predict where main peaks will appear.

On the other hand, there are computer simulations.

It is possible to compute very rigorous IR spectra and this can be carry out by using two methods of approximations: harmonic and anharmonic approximations.

- The harmonic approximation treats the displacement of the bonds of a molecule away from their equilibrium distance r_e as an harmonic oscillator. This is described by Hook's law and the potential energy curve varies parabolically as a function of the displacement. This approximation gives an equally spaced energy separation.

- The anharmonic approximation states that the variation of potential energy of a system with nuclear distance usually is not a symmetric parabola but a *Morse curve*. In this case, the separation between two adjacent energy states decreases progressively as the quantum number increases.

However, the anharmonic oscillator approximation requires more CPU time and thus the harmonic approximation is widely used.
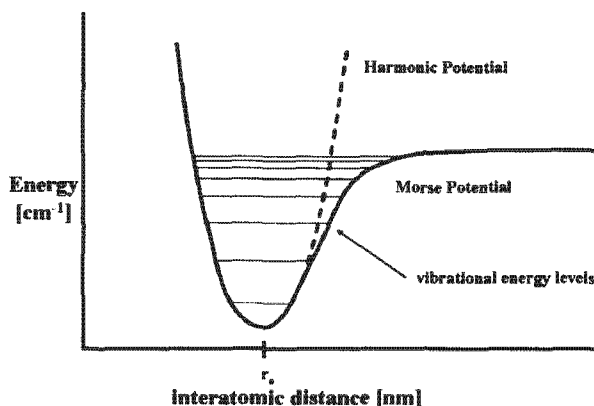
Figure 2.9 Representation of harmonic and anharmonic curves and the vibrational energy levels.

It follows that, when computing vibrational frequencies, approximations have to be used and this leads to a deviation between experimental and computed IR spectra.

Nevertheless, publications help us to use specific "scaling factor" in order to compare computed normal mode analysis, even with different basis sets, with experimental data [78].

In a parameterisation procedure, the latter represent the target data: the computed vibrational frequencies will be adjusted, step by step, to the experimental values.

2.4.5 Non-bonded Parameters.

The difficulty of obtaining and validating this class of parameters (mainly atomic charges) lies in the fact that they are not directly observable. Furthermore, they are not formally quantum mechanical observables, and therefore cannot be extracted from molecular wavefunctions in a simple manner. The most widely used methods for estimating atomic charges are derived from a least-squares fit to the electrostatic potential (ESP) calculated on a large number of points around the molecule of interest. Examples of potential-based methods are the charge from electrostatic potential (CHELP), charge from ESP grid-based developed by Breneman and Wiberg (CHELPG) and the Merz and Kollman (MK) methods. They differ mainly in the choice of points where the ESP is calculated, and respectively, on concentric spherical shells around each atom, on a regularly spaced cubic grid around the molecules, on nested Connolly surfaces. Since the potential is dependent on the total molecular

charge distribution (and consequently on the nuclear positions), changes in structure may produce significant differences in the calculated charges values. So, the level of theory and geometry optimization used to get electrostatic potential, as well as the choice of potential point, will affect the resulting charges [79].

Usually, the procedure foresees numerous steps of comparison between QM calculations performed at different level of theories and furthermore, a comparison with the chosen classical simulation program (in this case, CHARMM) is required. When comparing QM with MD values – or vice versa - scaling factors have to be used.

As previously mentioned, in this thesis a parameterisation procedure for N-sugars has been performed (see **Chapter 6**), by employing Merz-Kollman charges, mainly because the majority of the parameters have been selected – *by analogy* – from the CGenFF, CHARMM General Force Field developed by Mackerell and co-workers [66], which indeed makes use of MK charges. Therefore, the following discussion will be focused exclusively on the optimisation procedure for this type of charges.

The entire procedure has been performed in *Gaussian 03* [80].

Initial estimates for the partial atomic charges may be made *by analogy* – when a similar compound is available - or from the MP2/6-31G(d) Merz-Kollman charges (this charge-fitting scheme provides a good initial guess for the CGenFF partial atomic charges, especially in the case of heterocycles). This can be done, by building the model compound and optimising it with an MP2/6-31G(d) level of theory.

Once the initial charges have been determined, the first comparison step can be performed with classical methods (in this case, with CHARMM).

The next step usually foresees the building up of a compound+water complex for obtaining the interaction energy values. It is common to build different complexes with water in a variety of orientations, generally one complex for each hydrogen bond donor or acceptor. If the functional groups in the molecule can form more hydrogen bonds (e.g., an alcohol that can donate as well as accept a hydrogen bond), additional complexes have to be constructed, one for each possible hydrogen bond. These complexes are optimised with MP2/6-31G(d) level of theory and the water molecule is built in the TIP3P geometry (**Figure 2.10**).

The complexes are set up with an ''ideal'' linear geometry, with the interaction distances optimised at the HF/6-31G(d) level, keeping all other degrees of freedom fixed. The HF/6-31G(d) level of theory is used to maintain compatibility with CHARMM (especially in the case of the additive force field

CGenFF, as in this case). Finally, the optimised interaction distance is measured and the interaction energy is determined.
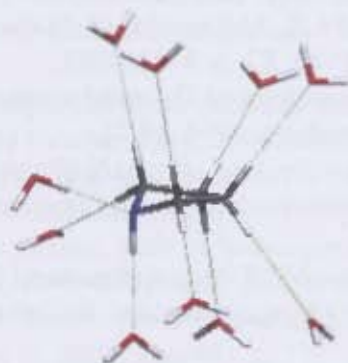
Figure 2.10 All the possible complexes compound–water have to be constructed, in order to obtain interaction energies. This example represents pyrrolidine complexes with TIP3P water as it is depicted in reference [66].

Having obtained the charges for the compounds–water interaction and the relative interaction energies, the same models can be used in CHARMM and a comparison between the obtained charges can be performed.

While higher levels of QM theory may lead to more accurate hydrogen bond energies and geometries, their use as target data will lead to a wrong comparison with the classical method (in particular, for the non-bond interactions). For polar, neutral molecules, empirical results should overestimate the magnitude of the QM dipole moment by 20 to 50% and should reproduce its orientation [66].

Some rules can facilitate charge fitting:

   • Charges are adjusted to maintain integer charges on groups of atoms, such as rings.

   • Aliphatic hydrogen atoms are always assigned a charge of +0.09, except when they are located on an aliphatic carbon atom directly adjacent to a positively charged nitrogen atom, where they are assigned a standard charge of +0.28. Similarly, aromatic C-H moieties not adjacent to a heteroatom are assigned charges of +0.115 and −0.115 on the C and the H atom, respectively [66].

   In this dissertation, as aforementioned, the parameterisation and validation procedures have been applied in order to study a class of compound, i.e. iminosugars, which are described in Chapter 6.

References

1. Esquivelzeta-Rabell, M., J. Peon, and G. Cuevas, *Rotational Diffusion of Dihydroxy Coumarins: Effect of OH Groups and Their Relative Position on Solute-Solvent Interactions.* Journal of Physical Chemistry B, 2009. **113**: p. 8599-8606.

2. Impey, R.W., P.A. Madden, and I.R. McDonald, *Hydration and Mobility of Ions in Solution.* Journal of Physical Chemistry, 1983. **87**: p. 5071-5083.

3. Ruckenstein, E., *On the Phenomenological Thermodynamics of Hydrophobic Bonding.* Journal of Dispersion Science and Technology, 1998. **19**(2).

4. Schravedijk, P. and N.F.A. Van der Veg, *From Hydrophobic to Hydrophilic Solvation: An Application to Hydration of Benzene.* Journal of Chemical Theory and Computation, 2005. **1**(4).

5. Kay, L.E., *Protein Dynamics from NMR.* Nature Structural Biology, 1998: p. 513-517.

6. Uedaira, H., et al., *Hydration of Oligosaccharides.* Bulletin of the Chemical Society of Japan, 1990. **63**(12): p. 3376-3379.

7. Engelsen, S.B., et al., *The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and NMR spectroscopy.* Biophysical Chemistry, 2001. **93**: p. 103-127.

8. Aeberhardt, K., et al., *Ultrasonic wave spectroscopy study of sugars oligomers and polysaccharides in aqueous solutions: The hydration lenght concept.* International Journal of Biological Macromolecules, 2005. **36**: p. 275-282.

9. Burakowsky, A. and J. Glinski, *Hydration numbers of non-electrolytes – Application of the acoustic method of Pasynski.* Chemical Physics, 2007. **332**: p. 336-340.

10. Corzana, F., et al., *A hydration study of 1,4 and 1,6 linked alpha-glucans by comparative 10 ns molecular dynamics simulations and 500MHz NMR.* Journal of Computational Chemistry, 2004. **25**(4): p. 573-586.

11. Gaida, L.B., C.G. Dussap, and J.B. Gros, *Variable hydration of small carbohydrates for predicting equilibrium properties in diluted and concentrated solutions.* Food Chemistry, 2006. **96**(3).

12. Galema, S.A. and H. Hoiland, *Stereochimical Aspects of Hydration of Carbohydrates in Aqueous Solutions. 3. Density and Ultrasound Measurements.* Journal of Physical Chemistry, 1991. **95**: p. 5321-5326.

13. Junquera, E., D. Olmos, and E. Aicart, *Carbohydrate-water interactions of p-nitrophenylglycosides in aqueous solution. Ultrasonic and densitometric studies.* Journal of Physical Chemistry Chemical Physics, 2002. **4**: p. 352-357.

14. Kiriukhin, M.Y. and K.D. Collins, *Dynamic hydration numbers for biologically important ions.* Biophysical Chemistry, 2002. **99**(2): p. 155-168.

15. Uedaira, H., M. Ikura, and H. Uedaira, *Natural-Abundance Oxygen-17 Magnetic Relaxation in Aqueous Solutions of Carbohydrates* Bulletin of the Chemical Society of Japan, 1989. **62**: p. 1-4.

16. Tait, M.J., et al., *Hydration of Monosaccharides: A Study by Dielectric and Nuclear Magnetic Relaxation.* Journal of Solution Chemistry, 1972. **1**(2): p. 131-151.

17. Andersson, C. and S.B. Engelsen, *The Mean Hydration of Carbohydrates as Studied by Normalized Two-Dimensional Radial Pair Distributions.* Journal of Molecular Graphics and Modelling, 1999. **17**(2): p. 101-105.

18. Galema, S.A., *The effect of stereochemistry on carbohydrate hydration in aqueous solutions*, in *Faculty of Mathematics and Natural Sciences.* 1992, Ph.D. Thesis. University of Groningen: Groningen, The Netherlands.

19. Galema, S.A., M.J. Blandamer, and J.B.F.N. Engberts, *Stereochemical Aspects of the Hydration of Carbohydrates. Kinetic Medium Effects of Monosaccharides on a Water-Catalyzed Hydrolysis Reaction.* Journal of American Chemical Society, 1990. **112**: p. 9666-9668.

20. Iyer, P.V. and L. Ananthanarayan, *Enzyme Stability and Stabilization of Aqueous and Non-aqueous Environment.* Process Biochemistry, 2008. **43**(10): p. 1019-1032.

21. Chalikian, T.V., A.P. Sarvazyan, and K.J. Breslauer, *Hydration and partial compressibility of biological compounds.* Biophysical Chemistry, 1994. **51**(2-3): p. 89-109.

22. Israelachvili, J. and H. Wennerstrom, *Role of hydration and water structure in biological and colloidal interactions.* Nature, 1996. **379**(6562): p. 219-225.

23. Pal, S.K., et al., *Biological water: femtosecond dynamics of macromolecular hydration.* Journal of Physical Chemistry B, 2002. **106**(48): p. 12376-12395.

24. Zavitsas, A.A., *Properties of Water Solutions of Electrolytes and Non-Electrolytes.* Journal of Physical Chemistry B, 2001. **105**: p. 7805-7817.

25. Enderby, J.E., *Ions Solvation via Neutron Scattering.* Chemical Society Reviews, 1995. **24**: p. 159-168.

26. Mason, P.E., et al., *Neutron Diffraction and Simulation Studies of the Exocyclic Hydroxymethyl Conformation of Glucose.* Journal of Chemical Physics, 2006. **125**(22): p. 224505/1-224505/9.

27. Mason, P.E., et al., *Neutron Diffraction and Computer Simulation Studies of D-Xylose.* Journal of American Chemical Society, 2005. **127**: p. 10991-10998.

28. Mason, P.E., et al., *Determination of a Hydroxyl Conformation in Aqueous Xylose Using Neutron Scattering and Molecular Dynamics.* Journal of Physical Chemistry B Letters, 2006. **110**(7): p. 2981-2983.

29. Mason, P.E., et al., *Neutron diffraction and simulation studies of the exocyclic hydroxymethyl conformation of glucose.* Journal of Chemical Physics, 2006. **125**(22): p. 224505/1-224505/9.

30. Brady, J.W., et al., *Molecular dynamics and neutron diffraction studies of the structuring of water by carbohydrates and other solutes.* Modelling Molecular Structure and Reactivity in Biological Systems. 2006. 76-82.

31. Tait, M.J., et al., *Hydration of Monosaccharides: A Study by Dielectric and Nuclear Magnetic Relaxation.* Journal of Solution Chemistry, 1974. **1**(2): p. 131-151.

32. Smith, L.J., et al., *Molecular dynamics of glucose in solution: A quasielastic neutron scattering study.* Journal of Chemical Physics, 2004. **120**(8).

33. Brady, J.W., *Molecular Dynamics Simulations of α-D-Glucose in Aqueous Solution.* Journal of American Chemical Society, 1989. **111**: p. 5155-5165.

34. Hajduk, P.J., D.A. Horita, and L.E. Lerner, *Picosecond Dynamics of Simple Monosaccharides As Probed by NMR and Molecular Dynamics Simulations.* Journal of American Chemical Society, 1993. **115**: p. 9196-9201.

35. Stokes, R.H. and R.A. Robinson, *Interactions in Aqueous Nonelectrolyte Solutions. I. Solute-Solvent Equilibriums.* Journal of Physical Chemistry, 1966. **70**(7): p. 2126-2131.

36. Hinton, J.F. and E.S. Amis, *Nuclear Magnetic Resonance Studies of Ions in Pure and Mixed Solvents.* 1966.

37. Uedaira, H. and H. Uedaira, *Sugar-Water Interaction from Diffusion Measurements*. Journal of Solution Chemistry, 1985. **14**(1): p. 27-34.

38. Schmidt, R.K., M. Karplus, and J.W. Brady, *The Anomeric Equilibrium in D-Xylose: Free Energy and the Role of Solvent Structuring*. Journal of American Chemical Society, 1996. **118**(3): p. 541-6.

39. Ha, S., et al., *Solvent Effect on the Anomeric Equilibrium in D-Glucose: A Free Energy Simulation Analysis*. Journal of American Chemical Society, 1991. **113**: p. 1553-1557.

40. Miyajima, K., K. Machida, and M. Nakagaki, *Hydrophobic Indexes for Various Monosaccharides*. Bulletin of Chemical Society Japan, 1985. **58**: p. 2595-2599.

41. Walkinshaw, M.D., *Variation in the Hydrophilicity of Hexapyranose Sugars Explains Features of the Anomeric Effect*. Journal of Chemical Society Perkin Transactions 2, 1987. **2**: p. 1903-1906.

42. Dwek, R.A., *Glycobiology: Toward Understanding the Function of Sugars*. Chemical Reviews, 1996. **96**: p. 683-720.

43. Duus, J., H. Gotfrendsen, and K. Bock, *Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations*. Chemical Reviews, 2000. **100**: p. 4589-4614.

44. Molteni, C. and M. Parrinello, *Condensed Matter Effects on the Structure of Crystalline Glucose*. Chemical Physics Letters, 1987. **275**: p. 409-413.

45. Di Bari, M., et al., *Mean Square Fluctuations of Hydrogen Atoms and Water-Biopolymer Interactions in Hydrated Saccharides*. Biphysical Journal, 2001. **81**: p. 1190-1194.

46. Kuttel, M., J.W. Brady, and K.J. Naidoo, *Carbohydrate Solution Simulations: Producing a Force Field with Experimentally Consistent Primary Alcohol Rotational Frequencies and Populations*. Journal of Computational Chemistry, 2002. **23**(13): p. 1236-1243.

47. Ferrier, W.G., *The Crystal and Molecular Structure of β-D-Glucose*. Acta Crystallographica, 1963. **16**.

48. Jeffrey, G.A. and S.S.C. Chu, *The Refinement of the Crystal Structures of β-D-Glucose and Cellobiose*. Acta Crystallographica, 1968. **B24**.

49. Jeffrey, G.A., et al., *Neutron Diffraction Refinement of Partially Deuterated β-L-Arabinopyranose and α-L-Xylopyranose at 123K*. Acta Crystallographica, 1980. **B36**: p. 373-377.

50. Hoog, C. and G. Widmalm, *Free Energy Simulations of D-Xylose in Water and Methyl D-Xylopyranoside in Methanol*. Journal of Physical Chemistry B, 2001. **105**: p. 6375-6379.

51. Ibberson, R.M., et al., *Polymorphism in Cyclohexanol*. Acta Crystallographica, 2008. **B64**: p. 573-582.

52. Lloyd, M.A., et al., *Solid-state compounds of stereoisomers: cis and trans isomers of 1,2-cyclohexanediol and 2,3-tetra-lindiol*. Acta Crystallographica, 2007. **B63**: p. 433-447.

53. Loehlin, J.H., M. Lee, and C.M. Woo, *Hydrogen-bond Patterns and the Structures of 1,4-cyclohexanediol: 2:1 cis:trans-1,4-cyclohexanediol*. Acta Crystallographica, 2008. **B64**: p. 583-588.

54. Jansen, C., et al., *The energetically preferred orientation of the hydroxyl group in cyclohexanol. Ab initio and force field calculations*. Journal of Molecular Structure (Teochem), 1997: p. 398-399; 395-404.

55. Rabinowitz, I.N. and J. Kraut, *The Crystal Structure of Myo-Inositol*. Acta Crystallographica, 1964. **17**: p. 159.

56. Leach, A.R., *Molecular Modelling: Principles and Applications*. 1ˢᵗ and 2ⁿᵈ ed. 2001: Longman.

57. van Gunsteren, W.F. and J.C. Berendsen, *Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry.* Angewandte Chemie International Edition English, 1990. **29**: p. 992-1023.

58. Allen, M.P. and D.J. Tildesley, eds. *Computer Simulations of Liquids.* 1987, J. W. Arrowsmith Ltd.: Oxford.

59. Karplus, M. and G.A. Petsko, *Molecular Dynamics Simulations in Biology.* Nature, 1990. **347**: p. 631-639.

60. Jensen, F., *Introduction to Computational Chemistry.* 1999, England: J. Wiley and sons.

61. Brooks, B.R., et al., *CHARMM: The Biomolecular Simulation Program.* Journal of Computational Chemistry, 2009. **30**(10).

62. Case, D.A., et al., *The Amber Biomolecular Simulation Programs.* Journal of Computational Chemistry, 2005. **26**: p. 1668-1688.

63. Scott, W.R.P., et al., *The GROMOS Biomolecular Simulation Program Package.* Journal of Physical Chemistry, 1999. **103**: p. 3596-3607.

64. Smith, W., T.R. Forester, and I.T. Todorov, *DL_POLY: Application to Molecular Simulation,* ed. S.D. Laboratory. 2009, Daresbury, Warrington WA4 4AD Cheshire, UK.

65. Brooks, B.R., et al., *Charmm - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.

66. Vanommeslaeghe, K., et al., *CHARMM General Force Field:A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields* Journal of Computational Chemistry, 2010. **31**(671-690).

67. Ryckaert, J., G. Ciccotti, and J.C. Berendsen, *Numerical Integration of the Cartesian Equations of Motion of a System with Constrains: Molecular Dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.

68. Berendsen, H.J.C., J.R. Grigera, and T.P. Straatsma, *The Missing Term in Effective Pair Potentials.* Journal of Physical Chemistry, 1987. **91**(24): p. 6269-6271.

69. Kuttel, M., *Developing Analytical Tools for Saccharides in Condensed Phase,* in *Faculty of Science. Department of Computational Chemistry.* 1999, M.Sc. Thesis. University of Cape Town: Cape Town. South Africa.

70. Brady, J.W., et al., *A revised potential surface for molecular studies of carbohydrates.* Carbohydrate Research, 1988. **180**: p. 207-221.

71. Kony, D., et al., *An improved OPLS-AA force field for carbohydrates.* Journal of Computational Chemistry, 2002. **23**(15): p. 1416-1429.

72. Brooks, B.R., et al., *CHARMM: The Biomolecular Simulation Program.* Journal of Computational Chemistry, 2009. **30**(10).

73. Mahoney, M.W. and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.* Journal of Chemical Physics, 2000. **112**(20).

74. Shaw, K.E., C.J. Woods, and A.J. Mulholland, *Compatibility of Quantum Chemical Methods and Empirical (MM) Water Models in Quantum Mechanics/Molecular Mechanics Liquid Water Simulations.* Physical Chemistry Letters, 2009. **1**: p. 219-223.

75. Horn, H.W., et al., *Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew.* Journal of Chemical Physics, 2004. **120**(20).

76. Lin, H. and D.G. Truhlar, *QM/MM: what have we learned, where are we, and where do we go from here?* Theoretical Chemistry Accounts, 2007. **117**: p. 185-199.

77. Hopfinger, A.J. and R.A. Pearlstein, *Molecular Mechanics Force-Field Parameterization Procedures.* Journal of Computational Chemistry, 1984. **5**: p. 486-499.

78. Scott, A.P. and L. Radom, *Harmonic Vibrational Frequencies: An Evaluation of Hartee-Fock, Moller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors.* Journal of Physical Chemistry, 1996. **100**: p. 16502-16513.

79. Maciel, G.S. and E. Garcia, *Charges Derived from Electrostatic Potentials: Exploring Dependence on Theory and Geometry Optimization Levels for Dipole Moments.* Chemical Physics Letters, 2005. **409**: p. 29-33.

80. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A., *Gaussian 03, Revision C.02.* 2004, Gaussian, Inc.: Wallingford, CT.

Chapter 3

Techniques for Analysis of Simulations

In this chapter, the different analytical techniques applied for characterizing the six hexacyclic systems (**Figure 2.1**) are described.

Starting from the two "traditional" conformational correlation functions (*Pair* and *Spatial Distribution Functions*), which give a static snapshot of the water structure around a solute, time-dependant functions will then be introduced (such as water Residence Time (RT), auto and cross-correlation functions (TCF) and linear diffusion), followed by ring puckering analysis and free energy perturbation method.

3.1 Conformational Correlation Functions

3.1.1 Pair Distribution Function

To analyse the local structuring of liquids, the approach traditionally used is the calculation of the pair distribution function (PDF), usually indicated as *g(r)*. This gives the probability of finding pairs of atoms at a distance *r* apart, relative to the probability expected for a completely random distribution at the same density [1].

In molecular liquids, such as water, several different *g(r)* can be calculated: the site-site distributions (for example, O-O and O-H *g(r)*) or the solvent distribution around solute sites using the *g(r)* in which where one of the two sites can either be a solute atom or the entire residue.

The PDF, for solvent molecules around any solute is defined by [2]:

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

where *r* is the interatomic distance from the selected atom or atoms in the solute molecule and the solvent, ρ is the density of the bulk solvent and *N(r)* is the number of water molecules within a sphere of radius *r* around the solute. The factor of 4π normalises *g(r)* to unity at positions in the bulk solvent far from the selected solute atom [3].

The PDF can be determined experimentally using X-ray, neutron or electron diffraction structure factors [4], or computed from the trajectories of an MD simulation.

Gases, liquids and solids show very different pair distribution functions. In the gas phase, only one definite peak appears and then a rapid decay to unity indicating no long-range structure (or interactions).
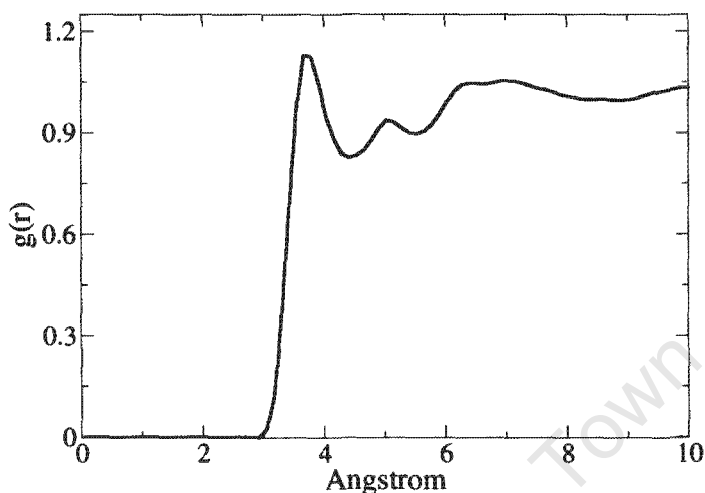
Figure 3.1 $g_{CO}(r)$ of β-D-glucopyranose (calculated for C1). The first maximum (in Å) gives the most probable distance (in Å) for the first coordination shell of the solute. Integrating under the first peak, the number of water molecules around the solute, in that region, can be obtained. The second maximum gives the second coordination shell.

In a solid, the PDF has an infinite number of sharp peaks with well defined separations, which indicates long-range structure (or interactions) typical of a crystal structure. The PDF for liquids is an intermediate case, because it shows few peaks, less than the solid but more than the gas, at short distances and then rapidly decays to unity at positions far away from the solute [3, 4].

From the PDF, it is possible to obtain the distance of the nearest neighbour to an atom of interest, plus the total number of atoms in the first shell by integrating under the first peak.

Consequently, PDF is used to give important insight into the structure of liquids.

However, this function alone is able to give a *static* picture of the hydration sphere of a solute. Nevertheless, it happens that during a simulation of a specific length (5 ns in the present case), water moves in and out the first coordination shell and not all the "PDF water molecules" take part to the diffusive motion of the solute. In other words, not all the H_2O molecules calculated from a PDF are *dynamic* water.

3.2.2 Spatial Distribution Function

The PDF, reported in **Section 3.2.1**, is able to give information in one-dimension, i.e. the number of neighbours and the distance from each site.

In order to visualise the location of water molecules around a solute a more detailed three-dimensional picture is required.
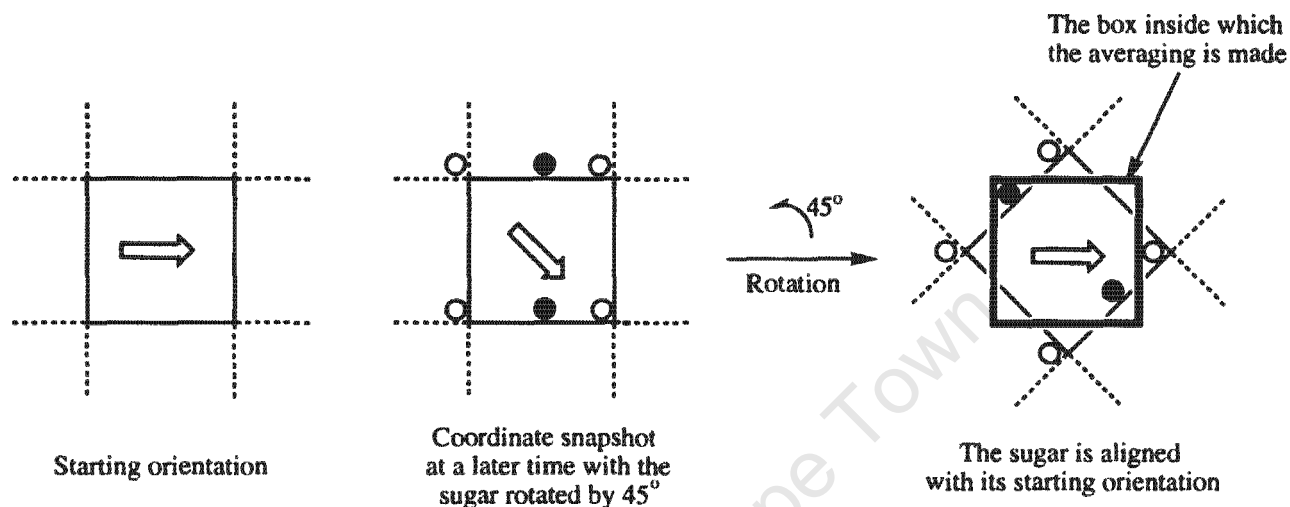
Figure 3.2 Scheme of the reorientation procedure used for the spatial distribution function, as showed in reference [3].

The spatial distribution function (SDF), or water probability density calculation, is a method suitable for this purpose. A brief description of this technique is given below.

When generating a map of the water structuring a solute, firstly all the rotational and translational diffusion throughout the simulation have to be removed. In order to achieve this, the instantaneous positions and orientations of the solute in each coordinate set, which make up the trajectory, are translated and rotated to obtain a best least-squares overlap with a reference frame, as shown in **Figure 3.2**.

The coordinate transformations performed during this reorientation procedure are applied to all the atoms in the system. If two or more trajectories are to be compared, then the original reference frame is used in order that the resulting density maps will have the same relative orientation. Care must be taken when choosing the atoms for the least-squares fit: by using atoms that move around too much, an unsatisfactory overlay of solute molecules and a consequent possible blurring of the water densities will be obtained [3]. Therefore, choosing only heavy atoms is the best option.

50

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

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

In equation **3.2**, *elec* stands for the number of electrons for the atom and *a* is calculated so that the function drops to 10% of its maximum value at the atom's van der Waals radius (1.4 Å for oxygen).

In this method, the simulation boxes are divided into a certain number of bins approximately 0.5 Å wide: to simplify, one can think of a Rubrik's cube, in which each side of the box is divided into a number of smaller cubes (**Figure 3.3**). The densities in each box are summed for all the selected frames from the dynamics simulation. The final water density matrix is normalised using equation **3.3**, so that the density of the bulk water corresponds to a value of 1, and 50% above bulk density corresponds to a value of 1.5 etc.

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

In eq. **3.3**, the numerators n_{xbin}, n_{ybin} and n_{zbin} are the number of divisions in the x, y and z directions in the simulation box, respectively. At the denominator, n_{atom} is the number of atoms whose density is being calculated (in our case the water oxygens) while $n_{electrons}$ is the number of electrons for that atom (i.e. 8 for oxygen, etc). This method of calculating the water densities is essentially the same as a three-dimensional pair distribution function but without the radial averaging [3]. The resultant electron densities are contoured in three-dimensions to produce diagrams of the water structuring around the solute molecule.

These two conformational correlation functions are such powerful resources of information for characterising and visualising solvent structure.

However, for further and more accurate inspections of the dynamics of the water binding the solute, it has been necessary to couple these two static functions (PDF and SDF) with time-dependant properties. The aim has been the calculation of the number of water molecules that characterises the chemistry of the solute, by binding the former in its first coordination shell for times long enough. In

51

other words, the aim has been to give a differentiation between "generic" water molecules entering and leaving the first coordination shell too fast from "specific" water molecules entering in that region for a long enough time range, thus taking place to the diffusive motion of the solute.

a   b   c
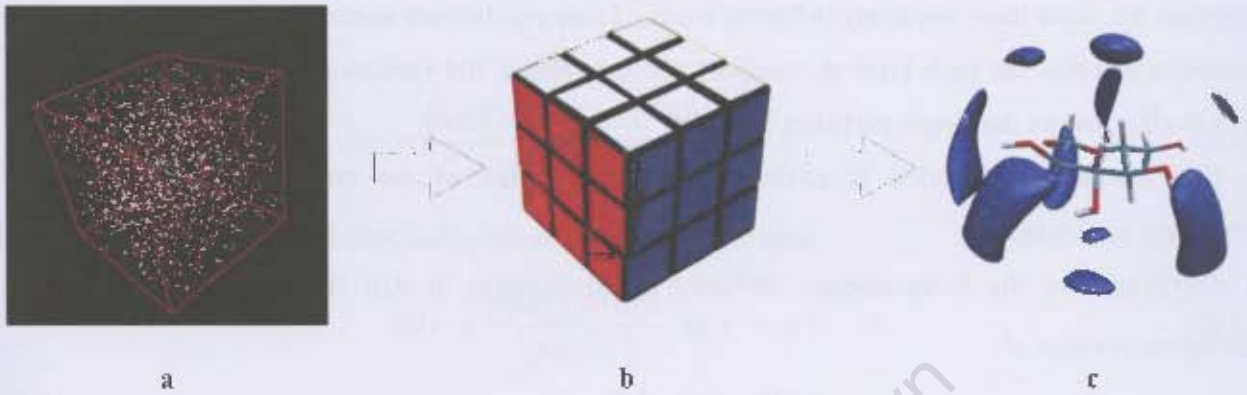
Figure 3.3. Schematic representation of the binning procedure for the spatial distribution function. The simulation box (a) is divided into n_{xbin}, n_{ybin} and n_{zbin} (b) along x, y and z axes, respectively. After the binning procedure, we generate the SDF. (c) depicts the contour of the probability density for the oxygen atoms of the water molecules around glucopyranose (as calculated from MD simulation).

3.3 Dynamic Properties

3.3.1 Time Correlation Function

In the fifties, the non-equilibrium statistical mechanics saw its beginning: this area showed that the phenomenological coefficients describing many transport processes and time-dependent phenomena in general could be written as integrals over a type of function called *time correlation function*. This function plays a similar role in non-equilibrium statistical mechanics as the partition function plays in equilibrium statistical mechanics [5].

The great advantage of the time correlation functions is that the resulting formulas for the transport coefficients do not depend on the details of any particular model and are not limited to any particular density region. For example, the *self-diffusion coefficient* can be expressed in terms of a velocity time correlation function by:

$$D_N = \frac{1}{3}\int_0^\infty \langle v(t) \cdot v(0) \rangle dt \tag{3.4}$$

this equation works for any density, for angle-dependent intermolecular forces, for polyatomic molecules and generally for any classical system in which the diffusion (or some transport process) is governed by the diffusion equation (or by some transport process equation) [5].

Since the state of equilibrium is unique, a single partition function gives all the thermodynamic properties, but since there are many different kinds of non-equilibrium states, we need a different time correlation function for each type of transport process. Hence the various time correlation functions play a similar role as the single partition function.

The time correlation function is a time-dependent measure of the correlation of spontaneous fluctuations in a system.

Let $\delta A(t)$ represent the instantaneous deviation or fluctuation in $A(t)$ from its time-independent equilibrium average, A:

$$\delta A(t) = A(t) - \langle A \rangle \tag{3.5}$$

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

$$\partial A(t) = \partial (t; r^N, p^N) = \partial A[r^N(t), p^N(t)] \tag{3.6}$$

Unless A is constant of the motion (e.g. the energy), $A(t)$ will look chaotic even in an equilibrium system. While the equilibrium average of $\delta A(t)$ may seem "uninteresting" (i.e. $\langle \delta A \rangle = 0$), the generalised order parameter, S^2, is an important average of the Legendre polynomial. Furthermore, one can extract non-chaotic information by considering the equilibrium correlations between the fluctuations at different times.

The correlation between $\delta A(t)$ and an instantaneous or spontaneous fluctuation at time zero is given by:

$$C(t) = \langle \delta A(t) \delta A(0) \rangle = \langle A(t) A(0) \rangle - \langle A \rangle^2 \tag{3.7}$$

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

$$C(t) = \int dr^N dp^N f(r^N, p^N) \delta A(0; r^N, p^N) \delta A(t; r^N, p^N) \tag{3.8}$$

where $f(r^N, p^N)$ is the equilibrium phase space distribution function. In an equilibrium system, the correlations between the dynamical variable, A, at different times should depend upon the separation between these times only and not the absolute value of time [15]. Thus the non-normalised time correlation function is:

$$C(t) = \langle \delta A(t') \delta A(t'') \rangle, \text{ for } t = t'' - t' \tag{3.9}$$

and the normalised time correlation function:

$$c(t) = \frac{\langle \delta A(t') \delta A(t'') \rangle}{\langle (\delta A(0)^2) \rangle} \tag{3.10}$$

At short times,

$$c(0) = \frac{\langle \delta A(0) \delta A(0) \rangle}{\langle (\delta A(0)^2) \rangle} = \frac{\langle (\delta A(0))^2 \rangle}{\langle (\delta A(0)^2) \rangle} = 1 \tag{3.11}$$

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

$$c(t) \rightarrow \frac{\langle \delta A(t) \rangle \langle \delta A(0) \rangle}{\langle (\delta A(0))^2 \rangle}, \text{ as } t \rightarrow \infty \tag{3.12}$$

and since $\langle \delta A \rangle = 0$, $c(t) \rightarrow 0$ for $t \rightarrow \infty$, that is confirmed by the Schwartz inequality. The latter guarantees that the absolute value of $c(t)$ lies between 0 and 1, with values close to 1 indicating a high degree of correlation, between the dynamical variables. Going toward zero we have lesser and lesser correlation, thus a value of zero indicates no correlation.

Lower correlation with increasing time is expressed by the "*regression of spontaneous fluctuations*" in the Onsager's hypothesis: the relaxation of macroscopic non-equilibrium disturbances is governed by the same laws as the regression of spontaneous microscopic fluctuations in an equilibrium system.[6] Simplifying, one can think of a system, which has been disturbed, that will relax or tend toward the equilibrium state much in the same way as spontaneous fluctuations occur in an equilibrium system.

For a correlation with the same dynamical variable, A, with a time integral of $t \in [0,\infty]$ an *autocorrelation function* can be obtained, **Figure 3.4.** These functions are of particular interest for several reasons: they give a clear picture of the dynamics in the solution, their time integrals τ may often be directly related to macroscopic transport properties, and in addition their Fourier transform, $I_{AA}(\omega)$, may often be related to the experimental spectra (on this regard, see equation **3.15**).[3]

Correlation between two different dynamical variables, A and B, is also possible and it gives a so-called *cross-correlation function*:

$$C(t) = \langle \delta A(t) \delta B(0) \rangle \tag{3.13}$$

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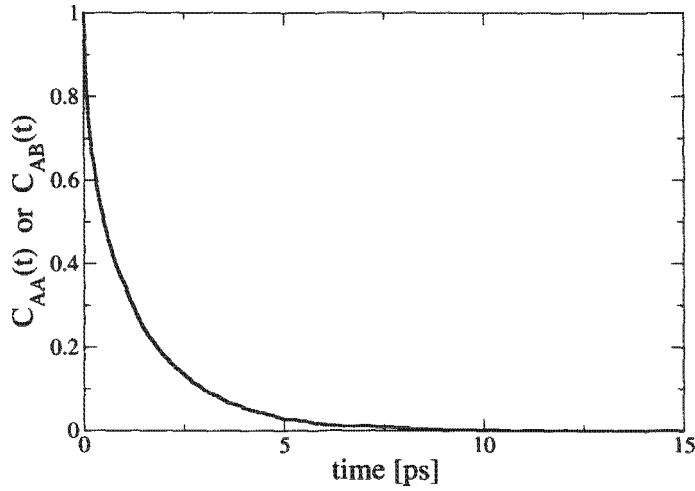
Figure 3.4: Normal exponentially decaying nature of time correlation function (AB) and autocorrelation function (AA).

$$c(t) = \frac{\langle \delta A(t) \delta B(0) \rangle}{\langle \delta A(0) \delta B(0) \rangle} \tag{3.14}$$

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

$$I(\omega) = \int_{0}^{\infty} \cos \omega t \langle \delta A(t) \delta B(0) \rangle dt \tag{3.15}$$

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

3.3.2 Residence Time

The time that a water molecule spends in the first solvation shell of the solute, before diffusing back into the bulk solution is known as the *residence time*. This is calculated from a number correlation

function, similar to a time correlation function, but differing in the fact that the dynamical variable has a value of either 0 or 1. The correlation time, τ, calculated from the time correlation function is the average residence time that a water molecule will spend in the first solvation shell.

To calculate the residence time, following a previous work on ions [7], a function $Pj(t,t_n; t^*)$ has been used. The latter takes the value 1 if the solvent molecule j lies within the coordination shell of a given solute atom at both time steps t_n and t_{n+1} and in the interim does not leave the coordination shell for any continuous period longer than t^*. Otherwise, the function takes the value 0.[2] From Pj one can derive a probability for a solvent molecule in the coordination shell to remain there during the time t, which is given by:

$$p(t) = \frac{\left\langle P_j(t,t_i;t^*) \right\rangle_{i,j}}{\left\langle P_j(0,t_i;t^*) \right\rangle_{i,j}} \tag{3.17}$$

The first minimum in the pair distribution function gives the most probable distance (in Å) for the first coordination shell of the solute. In the computation of the residence time, the latter has been approximated as a sphere – around both the carbons of the ring and the oxygen of the hydroxyls - with radius r. The latter is called spherical *cutoff*.

Once the PDF and residence time have been performed, one is able to calculate hydration numbers, n_H, by applying the following formula [7]:

$$n_H = n_{solute} \exp^{(-\tau_{bulk}/\tau_{solute})} \tag{3.18}$$

where n_{solute}, is the number of coordination water lying in the first solvation shell (from the PDF), τ_{bulk} and τ_{solute} are the residence time for water and solute, respectively.

3.3.3 Calculation of Hydration Numbers for a Series of Ions.

It can be useful to report an example, in which the hydration numbers for a series of atoms have been calculated. This particular case is extracted from Impey's publication [7] and it has represented a model for the hydration number calculations performed in this thesis. Some adjustments have been done, in order to apply Impey's method (on ions) to the present case (hexacyclic organic molecules). In Impey's work, a series of $[M^+]_{aq}$ with M = Li, Na, K, and $[X^-]$ with X = F, Cl have been considered.

---

[2] Impey fixed the value of t^* for ions at 2ps. In the present work, comparison between different t^* has been made. Eventually, the results with $t^*=2$ps have been seen to better reproduce experimental results.

The radial distribution function for all of them has been calculated as a function of temperature. The results for Li$^+$ only are reported in **Figure 3.5**.

Figure 2. Radial distribution function $g_{10}(R)$ for Li$^+$ as a function of temperature.
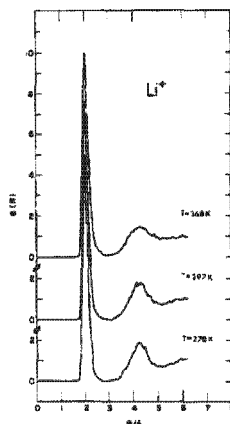
Figure 3.5 Radial distribution function for Li$^+$ ion, as depicted in Impey's work [7].

By integrating under the first peak of each radial distribution function for each ion, the coordination numbers have been obtained. Equation **3.19** has been then applied, in order to obtain the hydration numbers:

$$n_{ion}(t) = \frac{1}{N_t} \sum_{n=1}^{N_t} \sum_{j} P_j(t_n, t; t^*) \qquad (3.19)$$

Equation **3.19** is described as following: "*n_{ion} is the coordination number (from RDF), whereas $n_{ion}(t)$ measures the number of molecules which lie initially within the first coordination shell and are still there after a time t has elapsed. The parameter t* is introduced to take account of molecules which leave the first coordination shell only temporarily and return to it without ever having properly entered the bulk*".

Before calculating the n_H for the hexacyclic systems studied in this thesis, the same procedure applied by Impey in his publication has been followed and applied to Cl$^-$ (present in his series), to compare the results and check the reproducibility of the method.

The details of the simulations were different in the two cases, as summarized in **TABLE 3.1**.

TABLE 3.1: The details of the two simulations, Impey's case and this thesis, are summarised.

Details of Simulation	Impey's Case	This Thesis
geometry	cubic	cubic
number of water molecules	64 and 125	512
water model	MCY	TIP3P

The results of the two simulations for Cl⁻ anion are given in the following **TABLE 3.2**.

TABLE 3.2: Results of the two simulations on Cl⁻ are reported.

Cl-	Impey's Case	This Thesis
coordination water molecules	7.2	7.6
RT [ps] for *bulk* water	4.5	4.8
RT [ps] for *ion*	4.5	4.8
n_H	2.6	2.8

3.3.4 Linear diffusion

Time correlation functions are not only useful for analysing conformational events; they can also be used to calculate transport properties such as diffusion and viscosity and physical properties such as the shear strain. Furthermore, in this case, they have been used in the calculation of relaxation times for the two ring puckering parameters, θ and ϕ, as shown in **Section 3.6**.

And linear diffusion is another of the properties that has been calculated for the six hexacyclic systems.

This quantity is related to the time dependant displacement of the molecular centre of mass by a time correlation function that for long time periods can be approximated by an Einstein relation:

$$6tD = \left\langle \left| r(t) - r(0) \right|^2 \right\rangle \tag{3.18}$$

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

3.4 Puckering Analysis

Conformational behaviour of simple rings plays a critical role in determining the chemistry of these structures and the description of nonplanar medium-sized nonaromatic rings is a problem with a long history. Even if a qualitative description of puckering for rings (specifically, in six-membered rings) dates back to the studies of Sachse (see [8] and references herein) on cyclohexane, however its quantitative solution began in 1947 with Kilpatrick, Pitzer, and Spitzer, with their discussion of the molecular structure of cyclopentane (N=5) [9].

They started from the normal modes studies of *out-of-plane* motions of a planar regular pentagon, reporting that the displacement of the *j*-th carbon atom perpendicular to the plane of the molecule can be described by:

$$z_j = \left(\frac{2}{5}\right)^{\frac{1}{2}} q \cos\left(2\psi + 4\pi(j-1)/5\right) \tag{3.19}$$

where q represents the puckering amplitude and ψ is a phase angle describing various kinds of puckering. However, when this coordinate system was applied to cyclic molecules with unequal bond length and angles, difficulties were encountered.

About 1972, when Geise, Altona, Romers and Sundaralingam [10] introduced the torsion angle parameter θ instead of displacement perpendicular to a chosen reference plane, but their attempt to give a quantitative value to the displacement of atoms in a ring was approximate. These torsion angles are directly derivable from the atomic coordinates and are all zero in the planar form. For these torsion angles, they proposed a relationship of the form:

$$\theta_j = \theta_m \cos(P + 4\pi(j-1)/5) \tag{3.20}$$

where θ_m is an amplitude and P is a phase angle. Fixing the values for the five θ_j, the phase P is obtained (as consequence of **3.20**) from:

$$\tan P = \frac{\theta_3 + \theta_5 - \theta_2 - \theta_4}{2\theta_1(\sin \frac{1}{5}\pi + \sin \frac{2}{5}\pi)} \tag{3.21}$$

and θ_m follows from (**3.20**) for one particular atom (actually atom 1). This procedure has been used to obtain puckering coordinates for a large number of five-membered rings.

Although the method of Geise *et al.,* can be applied directly to any five-membered ring given only the torsion angles, it is subject to certain disadvantages: the puckering amplitude calculated by the above procedure will depend indeed on which atom is chosen as number one.

Then, Dunitz gave a much more precise quantitative definition for displacement from planarity of a ring molecule, which related to the original expression given by Kilpatrick.

Subsequently, Cremer and Pople proposed a general definition of ring-puckering coordinates which was close to the cyclopentane treatment of Kilpatrick *et al.,* but which can be applied without approximation to any cyclic molecule given only the coordinates of the nuclear positions of the atoms in the ring [11].

Their definition of ring puckering coordinates has been applied here to understand which is the most favourable conformation adopted by the six-membered cyclic solutes in water. Therefore, even if their method includes quantitative description for cyclic rings with various N atoms, this dissertation is focused on six-membered rings.

The Cremer and Pople general definition of ring puckering makes use of the projections z_j of the position vector of each ring atom into the normal of the mean ring plane. In the particular case of a six-membered ring, this method defines coordinates in terms of the distances z_j as a functions of 3 parameters (q_2, ϕ_2 and q_3), by using the following set of equations, as fully derived and reported by Autieri and co-workers [8, 12]:

$$q_2 \cos\phi_2 = \sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \cos\left[\frac{2\pi}{3}(j-1)\right]$$

$$q_2 \sin\phi_2 = -\sqrt{\frac{1}{3}} \sum_{j=1}^{6} z_j \sin\left[\frac{2\pi}{3}(j-1)\right] \qquad (3.22)$$

$$q_3 = \sqrt{\frac{1}{6}} \sum_{j=1}^{6} (-1)^{j-1} z_j$$

where q_2 and ϕ_2 are a single amplitude-phase pair while q_3 is a single puckering coordinate. By changing a coordinate to a spherical coordinate frame (Q, θ, ϕ) [8, 11, 12], all kind of possible puckering can be visualized on the surface of a globe:

$$q_2 \cos\phi_2 = Q\sin\theta\cos\phi$$

$$q_2 \sin\phi_2 = Q\sin\theta\sin\phi \qquad (3.23)$$

$$q_3 = Q\cos\theta$$
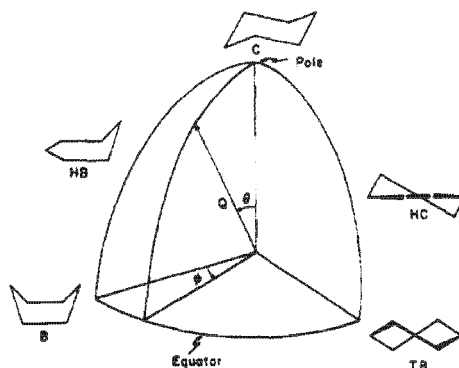
Figure 3.5 One octant of the sphere on which the conformations of six-membered rings can be mapped (for a constant Q). Special conformations are indicated: C = chair for $\theta = 0$, $\phi = 0$; B = boat for $\theta = 90°$, $\phi = 0$; TB = twist boat for $\theta = 90°$, $\phi = 90°$; HB = half-boat; HC =half-chair (as in reference [11]).

In **3.23**, ϕ represents the azimuthal angle and lies in the range [0,360], θ is the zenithal angle [0, π] and it is referred as the puckering amplitude, and Q is the radial coordinate ($Q^2 \equiv \sum_{j=1}^{N} z_j^2 = \sum_m q_m^2$ and Q>0) [12].

In order to visualize this system of coordinates, **Figure 3.5** – as depicted in reference [11] – could be helpful.

Therefore, different combinations of ϕ and θ are able describe all the different conformations that a cyclohexane-like molecule can adopt, namely the two chair, $^4C_1$ and the "inverted" $^1C_4$, the boats and skew shapes and the conformation in between, while Q is the total puckering amplitude. In **Figures 3.6** and **3.7**, a schematic representation of Q and of different combinations for ϕ and θ, in a 2-dimensional representation are reported.
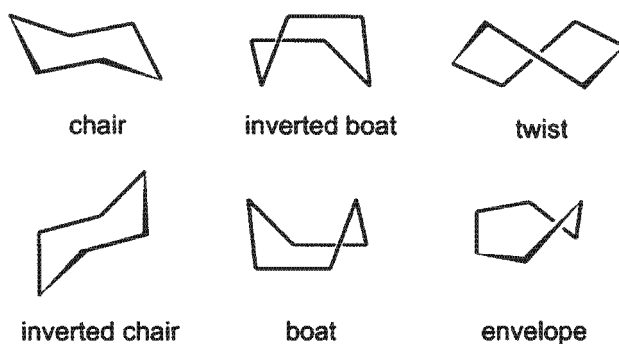
Figure 3.6. Schematic representation of the possible conformations for a general hexacyclic ring that can be described by adopting the Cremer and Pople notation.
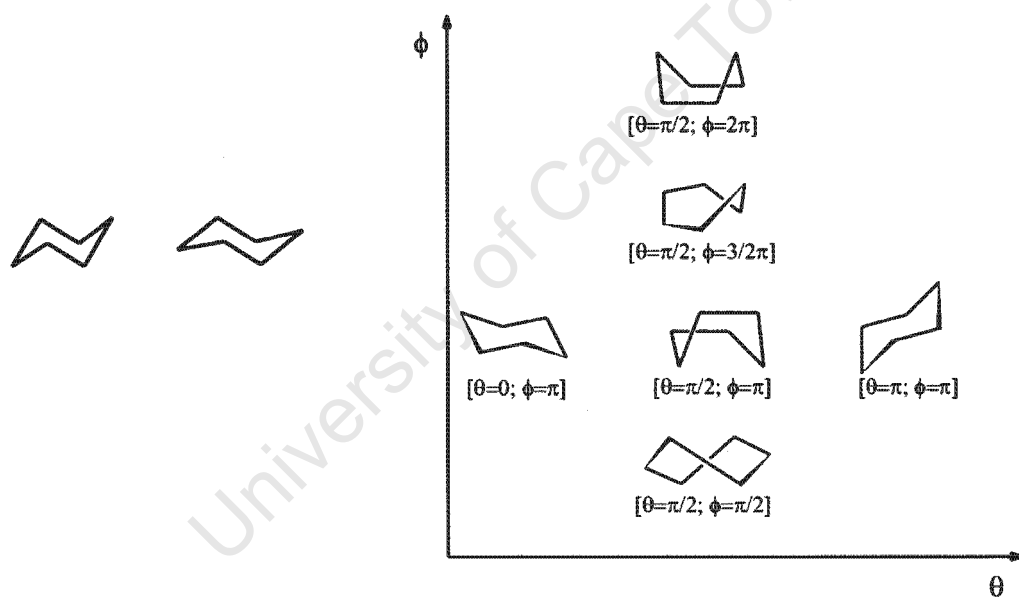
Figure 3.7. Schematic representation of the puckering parameters. The two conformations are both chair but they differ in Q values (left panel). Different combinations of the two puckering parameters, θ and ϕ, and their corresponding conformations reported in a 2-D representation (right panel).

3.5 Free Energy Perturbation.

The concept of Free energy is the most important in physical chemistry. The free energies of molecular systems describe their tendencies to associate and react or, as in this case, the tendency of a molecule to be solvated in water. Furthermore, by calculating the relative free energy of a particular system, a tool to organize and interpret the results of MD simulations is given and it is then possible to compare them with experimental data (when available) [13].

The statistical mechanical definition of free energy is in terms of the partition function, a sum of the Boltzmann weights of all the energy levels of the systems. However, only for the simplest model system can this free energy be represented by an analytical function. One can write a classical analog of the quantum mechanical partition function where the energy is viewed as a continuous function, rather than discrete. This is likely to be a good approximation in most system involving non covalent interactions at room temperature. Unfortunately, the free energy represented in this way requires an integration over all $3N$ degrees of freedom, where N = number of atoms in the system. Thus, this is impractical in most cases, namely for large systems.

However, it is possible to use free energy differences between related systems A and B ($\Delta G = G_B - G_A$) represented by the Hamiltonians H_A and H_B. This free energy difference can be represented as:

$$G_B - G_A = \Delta G = -RT \ln \left\langle e^{-\Delta H / RT} \right\rangle_A \qquad (3.24)$$

where $\Delta H = H_B - H_A$ and $\langle \ \rangle_A$ refers to an ensemble average over a system represented by the Hamiltonian H_A. Equation **3.24** is the fundamental equation of free energy perturbation calculations. There are different possible approaches to perform free energy calculations.

In the *double wide sampling* approach, the Hamiltonian H_A is described as:

$$H(\lambda) = \lambda H_B + (1 - \lambda) H_A \qquad (3.25)$$

where A can vary from 0 *(H = HA)* to 1 *(H = HE)*. One can then generalize **3.24** as follows:

$$\Delta G = G_B - G_A = \sum_{\lambda=0}^{1} -RT \ln \left\langle e^{-\Delta H' / RT} \right\rangle_\lambda \qquad (3.26)$$

where $\Delta H' = H_{\lambda + d\lambda} - H_\lambda$. In other words, one breaks up the free energy calculation into windows, each one involving a small enough interval in λ to allow the free energy to be calculated accurately. The *double wide sampling* is a useful test for convergence of the free energy. This involves calculating the free energy difference for both the $\lambda \to \lambda'$ and $\lambda' \to \lambda$ intervals. If one applies equation **3.26** and evaluates the ensemble at state λ and evaluates the free energy to mutate this into λ' this

should be the negative of the free energy determined by using the ensemble characteristic of λ' and calculating the free energy to mutate this to state λ.

An alternative approach in the free energy perturbation calculations is represented by the *thermodynamic integration*, where the free energy difference between two systems (one characterized by $H = H_A$ or $\lambda = 0$ in equation **3.25** and the other by $H = H_B$ or $\lambda = 1$ in equation **3.25**) is given by:

$$\Delta G = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{3.27}$$

The application of equation **3.27** requires evaluating the ensemble average of the derivative of the Hamiltonian with respect to λ, $\left\langle \partial H / \partial \lambda \right\rangle_\lambda$ at various values of λ. One can then use numerical integration methods to calculate ΔG by equation **3.27**.

The third commonly used method for free energy calculations is called *slow growth*, in which the Hamiltonian is changed an infinitesimal amount over each step of the simulation:

$$\Delta G = \sum_{no.steps\lambda=0}^{\lambda=1} (H_{n+1} - H_n) \tag{3.28}$$

where H_n is the Hamiltonian for a given λ and H_{n+1}, is the Hamiltonian for the next larger λ.

This equation can be derived from equations **3.24** or **3.27**, using the assumption in **3.24** that ΔG is small and in **3.27** that $\partial H / \partial \lambda = \Delta H / \Delta \lambda$ [14].

In this thesis, the *double wide sampling* has been applied as free energy perturbation calculation method, when evaluating the Gibbs energy of solvation for the six systems. The thermodynamic cycle that has been chosen to apply is represented in **SCHEME 1**.

SCHEME 3.1 Since the Gibbs Free Energy G is a state function, one is independent from the path followed for its calculation.
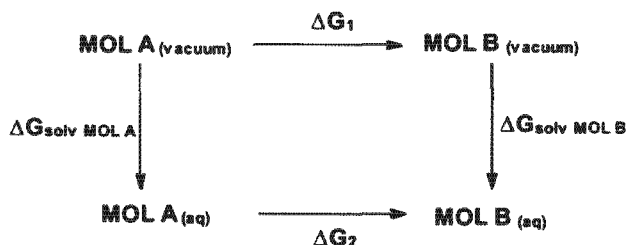
Therefore, by calculating the difference: $\Delta\Delta G = \Delta G_{solv}MOL\ B - \Delta G_{solv}MOL\ A = \Delta G_2 - \Delta G_1$, where ΔG_1 and ΔG_2 are the free energies of the process mutating a particular molecule A into a molecule B in solution and in vacuum respectively, the relative free energy of solvation can be obtained. Molecule A and B are usually indicated as hybrids. When performing FEP, going from a hybrid A to a hybrid B, or going from 0 to 1 – where 0 is referred as reactant and 1 as the product - the difference in energy of the hybrids should be as small as possible, generally of the order of 2 kcal/mol [14].

All the systems for which FEP have been performed start from the same reactant structure, namely cyclohexane. Therefore, it has been referred as 0, while the products cyclohexanol, 1,2-cyclohexanediol, 1,4-cyclohexanediol, myo-inositol, xylopyranose and glucopyranose are referred as 1.

Simulations where performed at $\lambda = 0.1, 0.2, 0.3, 0.4, \ldots 0.9$ and in order to check the consistency of the results, the simulations where run in both directions $\lambda \to \lambda'$ and $\lambda' \to \lambda$ except at the two end points (0 and 1), as it is for a *double-ended sampling* [15]. Furthermore, two additional steps were included: $\lambda = 0.05$ and 0.95.

3.5.1 FEP calculations in CHARMM. Example: Transformation of Methanol to Ethane

When performing FEP calculations in CHARMM, the previous equation **3.25** more specifically becomes:

$$H(\lambda) = H_E + (1 - \lambda)^N H_R + \lambda^N H_P \tag{3.29}$$

where:

- H_E is the environment part of the Hamiltonian;

- H_R is the reactant part of the Hamiltonian;

- H_P is the product part of the Hamiltonian;

- λ is the coupling parameter (as described above);

- N is an integer exponent.

The molecular system is divided into four types of atom:

▶ Reactant and product atoms, which are the atoms that are changed in the simulation, when going from the starting molecule (hybrid 0) to the final molecule (hybrid 1);

▶ Co-located atoms, for which only the charges are changed during the simulation;

▸Environment atoms, which are all the other atoms that are not changed throughout the simulation (solvent atoms, parts of the molecule in common to both reactant and products) [16].
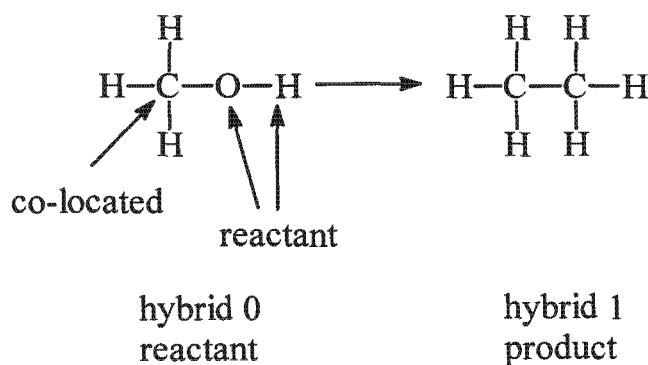
Figure 3.8 Schematic representation of the types of atoms, when performing a perturbation from methanol (reactant, left-hand side) to ethane (product, right-hand side). Oxygen and hydrogen in methanol are the reactant atoms (they are changed into the final CH_3 group of ethane), while carbon of methyl group of methanol is considered a co-located atom, as its charge is changed during the simulation. Environment atoms are all the other atoms (hydrogens in the methyl group in hybrid 0, as well as the surrounding water).

When considering the case of going from methanol (reactant, hybrid 0) to ethane (product, hybrid 1), the reactant atoms are the O and H in methanol molecule, as indicated in **Figure 3.8**; C of the methyl group in methanol is considered as co-located, as it starts (at $\lambda=0$) with the methanol methyl group charge and ends up (at $\lambda=1$) with the ethyl-methyl group charge.

All potential energy terms involving reactant atoms, the electrostatic interactions involving co-located charge atoms with their reactant charges, and the kinetic energy of the reactant atoms, are all included into H_R term. Potential energy terms involving product atoms and co-located product charge, together with their kinetic energy, are instead included into H_P.

In CHARMM, reactant, product and co-located atoms are defined in the topology file.

References

1.   Allen, M.P., *Introduction t to Molecular Dynamics Simulation*. NIC series, ed. N. Attig, et al. Vol. 23. 2004, Julich: John von Neumann Institute for Computing.

2.   Rossky, P.J. and M. Karplus, *Solvation. A Molecular Dynamics Study of Dipeptide in Water*. Journal of American Chemical Society, 1979. **101**(8): p. 1913-1937.

3.   Kuttel, M., *Developing Analytical Tools for Saccharides in Condensed Phase*, in *Faculty of Science. Department of Computational Chemistry*. 1999, M.Sc. Thesis. University of Cape Town: Cape Town. South Africa.

4.   Best, R., *Combined NMR and Simulation Study of Carbohydrate Linkage Dynamics*, in *Faculty of Science. Department of Computational Chemistry*. 2000, M.Sc. Thesis. University of Cape Town: Cape Town. South Africa.

5.   McQuarrie, D.A., *Statistical Mechanics*. 2000, Sausalito, California: University Science Books.

6.   Chandler, D., *Introduction to Modern Statistical Mechanics*. 1987, Oxford: Oxford University Press.

7.   Impey, R.W., P.A. Madden, and I.R. McDonald, *Hydration and Mobility of Ions in Solution*. Journal of Physical Chemistry, 1983. **87**: p. 5071-5083.

8.   Autieri, E., et al., *Puckering Free Energy of Pyranose Using a Combined Metadynamics-Umbrella Sampling Approach*. arXiv:1001.3667v1 [physics.chem-ph], 2010.

9.   Kilpatrick, J.E., K.S. Pitzer, and R. Spitzer, *The Thermodynamics and Molecular Structure of Cyclopentane*. Journal of American Chemical Society, 1947. **69**: p. 2483-2488.

10.   Altona, C. and M. Sundaralingam, *Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. A New Description Using the Concept of Pseudorotation*. Journal of American Chemical Society, 1972. **94**(23): p. 8205-8212.

11.   Cremer, D. and J.A. Pople, *A General Definition of Ring Puckering Coordinates*. Journal of the American Chemical Society, 1975. **97**(6): p. 1354-1358.

12.   Sega, M., E. Autieri, and F. Pederiva, *On the Calculation of Puckering Free Energy Surfaces*. Journal of Chemical Physics, 2009. **130**.

13.   Vanden-Eijnden, E., *Some Recent Techniques for Free Energy Calculations*. Journal of Computational Chemistry, 2009. **30**: p. 1737-1747.

14.   Kollman, P., *Free-Energy Calculations: Applications to Chemical and Biochemical Phenomena*. Chemical Reviews, 1993. **93**(7): p. 2395-2417.

15.   Jorgensen, W.L., *Monte Carlo simulations of differences in free energies of hydration*. Journal of Chemical Physics, 1985. **6**: p. 3050-3054.

16.   Fleischman, S.H. and C.L. Brooks, *Thermodynamics of aqueous solvation: Solution properties of alcohols and alkanes*. Journal of Chemical Physics, 1987. **87**(5): p. 3029.

17.   Jorgensen, W.L. and J. Tirado-Rives, *The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin*. Journal of American Chemical Society, 1988. **110**(6).

CHAPTER 4

Ultrasound Interferometer

The theory behind the ultrasound interferometer is presented in this chapter. Through this experimental technique, hydration numbers, n_H, for all the six hexacyclic molecules studied in the thesis (**Figure 2.1**) have been obtained. Furthermore, by comparing experimental and computed n_H (from both MD and QM/MM) trends, it has been possible to validate computer simulations.

4.1 Ultrasound Interferometer

4.1.1 Interference Phenomenon

Interference is observed when two coherent waves superimpose in space.

Ultrasound interferometry is based on the phenomenon of interference: sound is a wave which propagates in a medium and, for its wave nature, can undergo interference.

If two coherent waves interact with a particle present in the propagation medium, the resultant displacement and velocity of the particle are the vector sums of the two separate effects due to each wave.

An ultrasound interferometer – as used in this case, or an acoustic interferometer in general - can be referred to as a type of differential device: its function is the translation of the phase difference of the initial coherent wave into intensity modulation [1].

Ultrasonic interferometry has been chosen in this context, because it presents several advantages all suitable for the aqueous systems studied here:

• Its applicability to small samples (like small hexacyclic molecules);

• Its ability to simultaneously measure both velocity and attenuation (from which n_H can be derived);

• It is based on a direct, simple device for the determination of the ultrasonic velocity in solution, which gives results with a high degree of accuracy.

• Only a small amount of solution is required.

• There is no danger of structural changes in the sample, as low ultrasonic energy is used for measurement.

A number of thermodynamic properties of a liquid system can be derived from the measurement of ultrasonic velocity coupled with density data, such as adiabatic compressibility [2, 3], Poisson ratio,

Debye temperature [3], isochoric molar heat capacities, molar volumes, free volumes, intermolecular lengths [4, 5]. In the present case, isentropic compressibility, β, and hydration numbers, n_H, have been estimated.

Ultrasonic interferometry has been used to study molecular interactions between the components of a solution [3, 6], as well for investigating the structure of liquids [7].

4.1.2 Brief History of Ultrasonic Interferometry

The theory behind the application of ultrasonic interferometry technique dates back to the beginning of 20th Century. This is a continuous wave method of ultrasonic measurements in solution in the MHz region, which has been used mainly for determination of wavelength, speed [8-11], and absorption coefficient [11] in solution. In 1925, G. W. Pierce presented a work in which, for the first time, he developed methods of high frequency interferometry [12]. A further extensive analysis on the subject was made by D. W. Dye [13], in the same year. In a following publication by Hubbard and co-workers [14], it has been shown that the output current of an ultrasonic interferometry is a complicated function of the absorption coefficient, the reflection coefficient of the reflector and the distance moved by the reflector. More recently, this technique has been also studied in detail in a comprehensive paper by W.G. Cady [15].

4.1.3 Experimental Device

The speed of the ultrasonic waves, which travel through the solution, is measured using a multi frequency ultrasonic interferometer of a single crystal, with variable path type (**Figure 4.1** depicts the device used in this thesis, multi-frequency generator and measuring cell, plus a detailed image of the latter).

The device consists of two parts:

1) A high frequency generator. This is a high frequency crystal controlled oscillator in the form of a modified circuit, operating in the megacycle range. It is designed to excite the transducer, which is a quartz crystal fixed at the bottom of the measuring cell, to produce ultrasonic waves at its resonant frequency in solution. The generator is provided with a DC micrometer to detect the current.

For initial adjustment two knobs are provided on the generator, one to adjust the position of the needle on the Ammeter and the other to increase the sensitivity of the instrument with greater deflection, if desired [16];

2) The measuring cell (**Figure 4.1**): this is a cylindrical metal container placed vertically on a heavy metal base, which serves as the coupler between the quartz crystal and the high frequency generator. The measuring cell is a double-walled cell containing the experimental solution and around it there is also provision for circulation of water from the thermostatically regulated water bath, to maintain a constant temperature. A quartz crystal acting as a transducer is fixed at the bottom of the cell and a movable metallic reflecting plate, controlled from the outside by micrometer screw, is kept parallel to the quartz crystal. The measuring cell consists of three parts: (i) metal base, (ii) container and (iii) metallic plate reflector. The measuring cell is connected to the high frequency generator through a co-axial cable [30].

4.1.4 Principles of the Ultrasonic Interferometer Set-up

An ultrasonic wave is produced when the high frequency generator excites the quartz crystal, which acts as a generating transducer. The resulting ultrasonic wave travels through the solution toward a movable reflecting metallic plate, which is kept parallel to the quartz crystal. The same quartz crystal also acts as a detecting transducer. The reflecting plate can be moved using a micro-meter screw and it can therefore approach or recede from the vibrating crystal, a change in amplitude and phase is brought about in this way. The reflector can be made to pass through successive positions of maxima and minima of the ultrasonic wave in the solution. When the distance between the movable metallic plate and the quartz crystal is exactly a whole multiple of the ultrasonic wavelength, standing waves are formed in the solution. In other words the incident and reflected ultrasonic waves are superimposed in solution. The aforementioned change in amplitude and phase of the vibrating crystal gives rise to piezoelectricity, and thus the current can be measured. The acoustic propagation produces an electrical current in the quartz crystal and the anode current of the generator is a maximum when standing waves are formed in solution. If the distance between the reflector plate and the crystal is increased or decreased, and the variation is a multiple of half of the wavelength ($\lambda/2$), the anode current is also at a maximum.
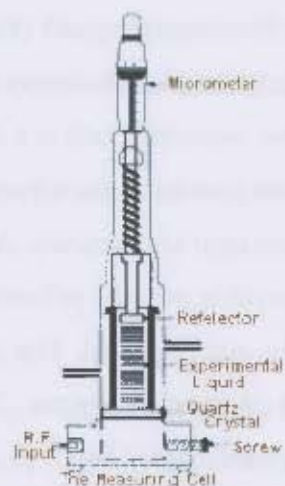
Figure 4.1. The ultrasound interferometer device (generator and measuring cell) used for experiments in this thesis, on the left hand side. On the right, a detailed image of the measuring cell, as depicted in reference [16].

4.2 Measurements of Hydration Numbers

4.2.1 Ultrasonic speed

Measurements of the ultrasonic speed, which travels through the solution, are based on finding the wavelength, λ, of the ultrasonic wave in solution. When a multiple of half the wavelength is achieved, by moving the adjustable metallic plate, a maximum is obtained in the observed current.

The ultrasonic speed, u, can easily be calculated using the relation:

$$u = f \times \lambda \tag{4.1}$$

where f is the frequency (of the high-frequency generator) and λ is the wavelength. The wavelength is calculated as follows:

$$\lambda = 2 \times d \tag{4.2}$$

where d is the distance between the quartz crystal and the reflector plate, obtained by turning the micrometer screw, for a maximum or a minimum deflection. Hence the ultrasonic speed can be calculated using:

$$u = f \times 2 \times d \tag{4.3}$$

4.2.2 Isentropic Compressibility

Isentropic compressibility, β, depends upon the compressible volume of molecules around a solute; this volume is very sensitive to the intermolecular interactions in solutions. In this regard, one generally observes that the stronger the interaction between the molecules in solution is, the lower the compressibility becomes.

The determination of the solution density, ρ, is required in the measurement of β as shown in equation **4.4**. The relationship between β and ultrasonic speed, u, is given by the Newton-Laplace equation:

$$\beta = \frac{1}{\rho u^2} \tag{4.4}$$

4.2.3 Hydration Numbers

Once β is computed for both solute and solvent, the following classical relationship can be applied to obtain hydration numbers, n_H:

$$n_H = \lim_{n_d \to 0} \frac{n_W}{n_S}(1 - \frac{\beta_S}{\beta_W}) \tag{4.5}$$

where n_w and n_S are the number of moles of water and solute in the solution, β_S and β_W are the isentropic compressibility (expressed in $m*s^2*kg^{-1}$) of the solution and of pure water, respectively.

Equation **4.5** implicitly assumes that: (i) n_H is the number of water molecules in the first layer surrounding the solute (it is usually referred to as the primary hydration number [6, 17]), and (ii) these water molecules are trapped so tightly that they can be considered as incompressible [6].

In other words, n_H generally refers to the number of incompressible water molecules attached to one molecule – the solute – and furthermore it is usually considered a "dynamical" parameter as it represents the number of water molecules which remain bound to the solute for times long enough to participate in its diffusive motion [18].

4.3 Statistical Analysis through Least Squares Regression Method

After obtaining the experimental data for the distance d and density, ρ, (obtained by using a densitometer device), ultrasonic speed, u, compressibility, β, and hydration numbers, n_H, can be derived from Eqs **4.3**, **4.4** and **4.5** respectively.

Understanding *how precisely* the derived values are obtained, or, in other words, knowing what the *confidence intervals* for the parameters are, is crucial.

Then, the model chosen to fit the results has to be checked, in order to understand if it is appropriate, specifically if it fits the data accurately.

In this dissertation all the data have been fitted by using a linear function. The discussion which follows will be exclusively referred to a linear fitting case.

The *goodness-of-fit* can be obtained by calculating the *chi-squared* (χ^2):

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{ax_i + b - y_{sp,i}}{\Delta_i} \right)^2 \tag{4.6}$$

where a and b are the coefficients of the linear-fit (the slope and the intercept, respectively), x_i are the *x-axes* values (in this case, concentrations of the solutions expressed in molarity, M [mol/V]) and $y_{sp,i}$ are the experimental values (e.g. u, β or n_H).

Δ_i in the denominator is the *uncertainty* on each individual experimental data $y_{sp,i}$, which can be obtained by computing the sets of equations reported in **Section 4.3.1** (Eqns. **4.14-4.16**).

Once all the terms in equation **4.6** have been defined, χ^2 has to be minimized; indeed the best fit in the least squares regression method is given by the parameters minimizing the chi-squared. This is carried out by using the following set of Eqns. **4.7-4.11**.

The sum in equation **4.6** can be rewritten as:

$$\sum_{i=1}^{N} \left(\frac{a^2 x_i^2 + b^2 + y_{sp,i}^2 + 2abx_i - 2ax_i y_{sp,i} - 2by_{sp,i}}{\Delta_i^2} \right) \equiv Aa^2 + Bb^2 + C + Dab + Ea + Fb \tag{4.7}$$

where the coefficients of the terms on the right hand side are defined as:

$$A = \sum_i \left(\frac{x_i^2}{\Delta_i^2} \right), \ B = \sum_i \left(\frac{1}{\Delta_i^2} \right), \ C = \sum_i \left(\frac{y_{sp,i}^2}{\Delta_i^2} \right), \ D = 2\sum_i \left(\frac{x_i}{\Delta_i^2} \right), \ E = 2\sum_i \left(\frac{x_i y_{sp,i}}{\Delta_i^2} \right) \text{ and}$$

$$F = -2\sum_i \left(\frac{y_{sp,i}}{\Delta_i^2} \right)$$

(4.7.1)

The minima (and maxima) of a generic function are characterized by the first derivative of the function going to zero. By differentiating the χ^2 with respect to each of the coefficients, a and b, and set the derivatives equal to zero, two equations in two unknowns are obtained:

$$\begin{cases} \dfrac{\partial \chi^2}{\partial a} = 0 \\[2em] \dfrac{\partial \chi^2}{\partial b} = 0 \end{cases} \tag{4.8}$$

which gives the best fit values for a and b.

By differentiating the previous formula (Eq **4.7**), one obtains:

$$2Aa + Db + E = 0 \quad \text{and} \quad 2Bb + Da + F = 0 \tag{4.9}$$

from which the values for a and b follow:

$$a = \frac{\dfrac{-D^2E}{2A} + DF}{4AB - D^2} - \frac{E}{2A} \rightarrow a = \frac{DF - 2BE}{4AB - D^2} \tag{4.10}$$

$$b = \frac{\dfrac{DE}{2A} - F}{2B - \dfrac{D^2}{2A}} \rightarrow b = \frac{DE - 2AF}{4AB - D^2} \tag{4.11}$$

The uncertainty in the two values, a and b can be calculated, by firstly defining a function σ:

$$\sigma = B \times A - \frac{D^2}{4} \tag{4.12}$$

where A, B and D are the parameters previously defined (Eq. **4.7.1**).

From 4.12, the two errors follow:

$$\pm a = \sqrt{\frac{B}{\sigma}} \text{ and } \pm b = \sqrt{\frac{A}{\sigma}} \tag{4.13}$$

4.3.1 Uncertainty in Coefficients

Assuming there are no systematic errors in the measurements (systematic errors induce covariances, not random errors), then for any function (as in equation **4.12**):

$$f(x_i) \text{ for } i=1,\ldots,N \tag{4.14}$$

the uncertainty in $f(x_i)$ is defined as (considering the uncertainty on each parameter x_i as independent and normally distributed):

$$\Delta_i^2 = \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \right)^2 \Delta x_i^2 \tag{4.15}$$

where the Δ_i^2 is the denominator in the previous equation **4.6**. By calculating the square-root, equation **4.13** becomes:

$$\Delta f = \sqrt{\sum_{i=1}^{N}\left(\frac{\partial f}{\partial x_i}\right)^2 \Delta x_i^2} = \sqrt{\left(\frac{\partial f}{\partial x_1}\right)^2 \Delta x_1 + \left(\frac{\partial f}{\partial x_2}\right)\Delta x_2 + ... + \left(\frac{\partial f}{\partial x_N}\right)\Delta x_N} \qquad (4.16)$$

References

1. Batra, N.K., et al., eds. *Acoustic Wave Interferometers*. The Wiley Encyclopedia of Electrical and Electronic Engineering, ed. J.W. Sons. Vol. 1. 1999.
2. Gekko, K., *Compressibility gives new insight into protein dynamics and enzyme function.* Biochimica et Biophysica Acta, 2001. **1595**: p. 382-386.
3. Stahl, F.A. and R.P. Wolf. *Thermodynamic Properties of Solid Methane from Ultrasound Measurements.* in *Ultrasonic Symposium Proceedings*. 1977.
4. Bebek, K., *Speed of Ultrasound and Thermodynamic Properties of 1-butanol in Binary Liquid Mixtures at 293.15K.* Molecular and Quantum Acoustics, 2005. **26**.
5. Galema, S.A. and H. Hoiland, *Stereochimical Aspects of Hydration of Carbohydrates in Aqueous Solutions. 3. Density and Ultrasound Measurements.* J. Phys. Chem., 1991. **95**: p. 5321-5326.
6. Junquera, E., D. Olmos, and E. Aicart, *Carbohydrate–water interactions of p-nitrophenylglycosides in aqueous solution. Ultrasonic and densitometric studies.* Journal of Physical Chemistry Chemical Physics, 2002. **4**: p. 352-357.
7. Burakowsky, A. and J. Glinski, *Hydration numbers of non-electrolytes – Application of the acoustic method of Pasynski.* Chemical Physics, 2007. **332**: p. 336-340.
8. Freyer, E.B., J.C. Hubbard, and D.H. Andrews, *Sonic Studies of the Physical Properties of Liquids. I. The Sonic Interferometer. The velocity of sound in Some Organic Liquids and Their Compressibilites.* Journal of American Chemical Society, 1929. **51**: p. 759.
9. Hubbard, J.C., *The Acoustic Resonator Interferometer. II. Ultrasonic Velocity and Absorption in Gases.* Physical Review, 1932. **41**(4): p. 523.
10. Hubbard, J.C. and A.L. Loomis, *CXXII. The Velocity of Sounds in Liquids at High Frequencies by the Sonic Interferometer.* Philosophical Magazine, 1928. **5**(33): p. 1177-1190.
11. Klein, E. and W.D. Hershberger, *Supersonic Interferometers.* Physical Review, 1931. **37**(6): p. 760.
12. Pierce, G.W. in *Proc. Nat. Am. Acad.* 1925.
13. Dye, D.W. *The Piezo-Electric Quartz Resonator and its Equivalent Electrical Circuit.* in *Proceedings of Physical Society*. 1925. London.
14. Hubbard, J.C., *Problems in Acoustic Interferometry with Gases.* Physical Review, 1930. **36**(11): p. 1668.
15. Cady, W.G. in *Proc. I. R. E.* 1992.
16. *Manual for Multifrequency Ultrasonic Interferometer*, Mittal Enterprises: New Patel Nagar, New Delhi, India.
17. Zavitsas, A.A., *Properties of Water Solutions of Electrolytes and Non-Electrolytes.* J. Phys. Chem. B, 2001. **105**: p. 7805-7817.
18. Impey, R.W., P.A. Madden, and I.R. McDonald, *Hydration and Mobility of Ions in Solution.* Journal of Physical Chemistry, 1983. **87**: p. 5071-5083.

Chapter 5*

The Extent of Conformational Rigidity Determines Hydration in Non Aromatic HexaCyclic Systems

ABSTRACT

We conducted an ultrasonic study of hydration number for hexacyclic systems. We find, from these experiments, that cyclohexane based molecules such as cyclohexanol and myo-inositol show a very small increase in hydration number despite the large difference in the number of hydroxyl groups present in each of the molecules. There is however, a dramatic increase in hydration number when shifting from molecules with a cyclohexane frame to molecules with a cyclopyranose frame particularly glucose. An analysis of classical and quantum molecular dynamics simulation trajectories reveal that the hydration number is strongly linked to the conformational flexibility within the molecule. Cyclopyranose is a more rigid ring system compared with cyclohexane and so its ring fluctuates in a smaller range and frequency. The effect of the ring rigidity is that the hydroxyls tethered to the cyclopyranose ring undergo less positional diffusion compared with those attached to the cyclohexane ring. This allows for long time intermolecular hydrogen bonds between the hydroxyls bonded to cyclopyranose rings and the surrounding waters, which leads to an increase in the hydration numbers of carbohydrates compared with those of hydroxylated cyclohexanes.

Introduction.

The relative solubilities of simple atomic and molecular systems make it tempting to ascribe the underlying reasons for observed differences to primarily electrostatic interactions between solutes and solvents. Solubility equilibrium is achieved when the solvation and precipitation processes proceed at a constant rate. At the molecular length scale this is reduced to an equilibrium that is achieved between the interactions of solvent molecules with that of the interactions between solvent and solute molecules (solvation). Analysis of the energetics of solvation effects in classical condensed phase simulations are therefore often described in terms of strong electrostatic (Coulomb) and weaker dispersion (van der Waal terms) interactions.[1] Although not simply electrostatic in nature, the hydrogen bond has been singled out as a key reason for carbohydrate hydration[2-4] and its subsequent behaviour in water.[5,6] An early cautionary note made from a lengthy analysis of the character of organic compounds warned that only the presence of an hydroxyl group is insufficient to improve the favourability of solute water interactions.[7]

The most direct measure of the solute water interaction is through the measure of molecular hydration numbers. Numerous experimental techniques such as NMR[8] neutron scattering,[9] X-ray, and ultrasound interferometery[10,11] have been used to measure hydration numbers. These different experimental techniques lead to very different hydration number magnitudes while preserving the trend between a molecular series.[11] This is primarily because each method infers hydration by measuring physical properties unique to the experiment and on timescales restricted to the detection limits of the method. The definitions of hydration number range from the average number of waters that are rotationally slowed down by the solute[12] to the number of water molecules that are translationally bound to the solute.[13] Computer simulations are well suited to extract molecular events

on the 10^{-10} to 10^{-6} s timescales. It is therefore very informative to combine experimental hydration trends with those calculated from computer simulations.[14]

Studies of linear [15,16] and aliphatic alcohols[17] concluded that the extent of hydration is dependent on the number of hydroxyl groups present in a similar set of molecules. For example the dynamic hydration numbers (n_{dhn}) measured by the ^{17}O spin relaxation times, ^{1}T, for ethanol and ethylene glycol are 10.5 and 5.7 respectively.[15] Ishihara *et. al.* argued that the hydrophobic hydration for linear alcohols is disrupted by hydrophilic groups and so an increase in the number of hydroxyls for diols compared with their corresponding *n*-alcohols leads to a relative decrease in n_{dhn}. It is not only hydrophilic groups that alter the diffusional behaviour of the surrounding water molecules, hydrophobic groups have been shown by ^{2}H NMR relaxation[18] and femtosecond mid-infrared spectroscopy[19] to slow down the orientational mobility of water molecules in the surrounding hydration shell. Therefore, when attempting to associate the solute's molecular features responsible for the degree to which it is hydrated, both the hydrophilic and hydrophobic nature of the solute must be taken into account.

In an experiment measuring the partial molar volumes of polyols and comparing them with those measured from their corresponding sugars it was seen that the linear alcohols were more hydrated than the cyclic carbohydrates (sugars).[20] Carbohydrates are biologically important examples of hydroxylated cyclic systems. While the conformational flexibility of the linear alcohols is greater in range compared with the cyclic sugars the rate of conformational motion of the linear alcohols is far slower than that of the cyclic sugars. In particular the conformational motion in the case of linear alcohols is on the same timescale as that of the overall diffusion of the solute molecule while for cyclic carbohydrates the conformational motion is much faster compared the overall diffusion. These differences in hydration, between molecules of similar hydroxyl number and distribution, must be due

to the variation in the rate of conformational motion exhibited by linear compared with cyclic systems. A classic study, by Shiio, of carbohydrate water interactions using ultrasonic interferometery concluded that the dominant effect in distinguishing the hydration behaviours of carbohydrates from each other were the number of hydroxyls and their relative positions to each other.[21] In a later study using a similar experimental technique, Galema and Høiland, proposed that carbohydrates with smaller hydration numbers have a stereochemistry more suited to fit into the three dimensional hydrogen bonded structure of water.[10]
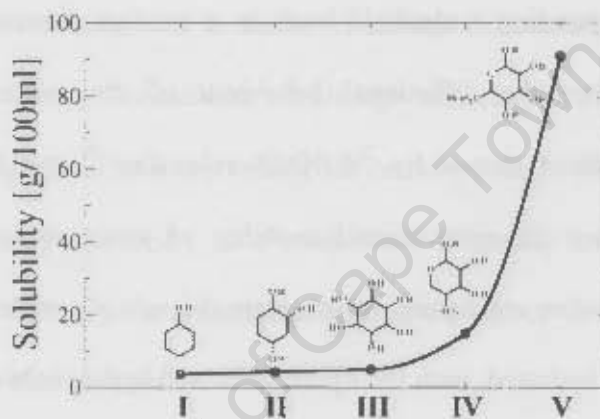
Figure 5.1 Schematic representation of the systems: **I** cyclohexanol, **II** 1,2-cyclohexanediol and 1,4-cyclohexanediol, **III** myo-inositol, **IV** β-D-xylopyranose and **V** β-D-glucopyranose.

While there has been a great focus on the immediate effect of polarity on hydration properties the conformational flexibility that underlies the internal molecular entropy has not been systematically investigated for carbohydrates. To uncover the nature of carbohydrate pyranose hydration we consider hexacyclic systems where the number of hydroxyls is progressively varied on a homogeneous ring skeleton (cyclohexane) followed by the introduction of an oxygen atom into the ring to make a pyranose frame. The cyclohexane based molecules that were studied are cyclohexanol, 1,2-cyclohexanediol, 1,4-cyclohexanediol and myo-inositol (in order of increasing number of hydroxyl

groups) while β-D-xylopyranose and β-D-glucopyranose are the pyranose based molecules that were studied.

The first clue that the hydration of hexa cyclic systems cannot simply be described in terms of the number of hydroxyl groups can be seen from their solubilities (Figure 5.1). There is a minor difference in solubility between cyclohexanol, 1,2-cyclohexanediol, 1,4-cyclohexanediol and 1,2,3,5/4,6-cyclohexanehexol (myo-inositol) where the only change is the increase in the number of hydroxyl groups. The big change is observed with the introduction of an ether linkage into the ring. The cycohexane based molecules (cyclohexanol through to myo inositol) are far less soluble than the pyranose based rings (xylopyranose and glucopyranose).

In this paper, we compare the hydration properties of these six flexible, non aromatic, cyclic molecules that are based on the hexane and pyranose rings. We explain the relationship between conformational properties and hydration for each molecule. We do this by measuring hydration numbers from ultrasound interferometery experiments and compare these with values calculated from MD simulations. To make sense of the hydration trend in relation to the molecular thermodynamics and relative conformational freedom we probe 1) the free energy, 2) the rotational distribution of the hydroxyls, 3) the conformational probability distribution of ring pucker and 4) the librational rate of ring pucker for each molecule. These measures and computations show that molecular flexibility is a fundamental property from which the relative hydration properties of hexacyclic systems can be extrapolated.

Measuring and Analysing Hydration

Ultrasound interferometer experiments are used to measure *dynamic* hydration number. This is the number of solvent molecules interacting with the solute for times long enough so that they diffuse together through the solvent. The hydration number (n_h) is calculated from

$$n_h = \frac{n_w}{n_s(1 - k_s/k_{s,0})} \tag{1}$$

which is a relation between the number of moles of water (n_w), number of moles of the solute (n_s), the isentropic coefficients of compressibility of water ($k_{s,0}$) and the solution (k_s). This relationship was first derived by Pasynski for electrolytes.[22] By passing repeated pulses of ultrasonic radiation through the solution followed by measuring the frequency of this radiation, the coefficients can be computed from the Laplace equation

$$k = \frac{1}{\rho u^2}$$

(2)

where ρ is the density of the aqueous solution and u is the ultrasonic speed. The assumption is that the compressibility of the hydration layer about the solute is zero. We present here hydration numbers that are for the solutes at infinite dilutions.

Water molecules located in popular configurations about the solute are radially and spatially identified from radial distribution (rdf) and spatial distribution functions (sdf) respectively. However, Impey et. al. recognized that hydration is dynamic and so sought to calculated the *persistent coordination* of water molecules about ions.[11] To calculate a hydration number, from a molecular dynamics (MD) computer simulation, that is comparable to that measured from the ultrasound experiment we expand on the expressions developed for ions.

First we calculate the number of waters in the first hydration shell (n_{solute}) that can be found from the rdf

$$g_{AB}(r) = \frac{\rho_B(r|r_A = 0)}{\rho_B} \tag{3}$$

where the first index A refers to a solute atom at the origin, and the second index B refers to a solvent atom B. r is the distance between the atoms, and ρ_B is the bulk density of B in the solvent. Two types of rdfs, $g_{CO}(r)$ and $g_{OO}(r)$ were calculated for the hexacyclic molecules in water. Coordination numbers were obtained by integrating $g_{AB}(r)$ out to the first minimum. The $g(r)$ radial distributions show for most of the rdfs in the hexacyclic molecular series that the first maximum appears around 3.5 Å for C-O$_w$ and 2.8 Å for O-O$_w$ interactions.[23]

Next we calculate the residence time, following Impey's work, using a Kronecker delta function $P_j(t,t_n; t^*)$. A value 1 is assigned if the solvent molecule j lies within the coordination shell of an atom on the solute for a time interval Δt without leaving the coordination shell for any period longer than t^*. A value of 0 is assigned if this condition is not met.[a]

The probability $p(t)$ of a solvent molecule surviving in the coordination shell for time t is then

$$p(t) = \frac{\left\langle P_j(t,t_j;t^*)\right\rangle_{j,t}}{\left\langle P_j(0,t_j;t^*)\right\rangle_{j,t}} \tag{4}$$

This probability decays exponentially with time and can be fitted by $p(t) \sim \exp(-t/\tau)$, where τ is the residence time that is inserted into

$$n_h = n_{solute}\, \exp^{(-\tau_{hyd}/\tau_{solw})} \tag{5}$$

where n_{solute} is the number water molecules coordinated to the solute atom A in its first hydration shell calculated from the rdf relation (equation 3). The hydration number (n_h) from the MD simulation is then calculated for each of the hexacyclic molecules.

Experimental Methods

All chemicals used were purchased from Sigma-Aldrich Inc. St. Louis, 3050 Spruce Street, MO 63103 USA. All the substances, with purities higher than 99% were used without further purification.

---

[a] Impey fixed the value of t^* for ions at 2ps. Here, comparison between different t^* have been made and eventually the results with t^*=2ps have been seen to better reproduce experimental results.

The water was doubly distilled. Solutions were prepared by weighting before measurements with a Afcoset-ER-120A with a precision of ±0.01 mg. The compositions of solutions are given in molarity (mol/L) throughout this paper. The ultrasonic speed of all the four solutions was measured using an Ultrasonic Interferometer Model M-81G set to an operating frequency of 3MHz. Density measurements were carried out at 27 °C through an Anton Paar digital Model DMA 35 with a precision of 0.01 gcm^{-3}. The temperature of the samples during the measurements was maintained to an accuracy of $\pm$0.1°C in an electronically controlled thermostatic Huber water bath with compatible control CC1.

While the rdfs are able to give information in one-dimension, i.e. the number of neighbours about each site, the sdfs provide a three-dimensional location of water probability density about a solute. When generating a probability density map of the water structuring a solute, all the rotational and translational diffusion throughout the simulation have to be removed. To achieve this, the instantaneous positions and orientations of the solute in each coordinate set, which make up the trajectory, are translated and rotated to obtain a best least-square overlap with a reference frame. The coordinate transformations performed during this reorientation procedure are applied to all the atoms in the system. If two or more trajectories are to be compared, then the original reference frame is used in order that the resulting density maps will have the same relative orientation. Care must be taken when choosing the atoms for the least squares fit; by using atoms that move around too much, an unsatisfactory overlay of solute molecules and a consequent possible blurring of the water densities will be obtained. Therefore, choosing only heavy atoms is the best option. Here we calculated the distribution of water around the solute, using only the oxygen atoms of the water molecules and these are shown in blue. A more complete description of the method used to calculate we have described the methods used to compute sdfs elsewhere. [6]

Computational Methods

We simulated infinite dilute conditions using a periodic cubic box of 24.8626 Å in length, containing 488 water molecules and a single solute molecule. Recently published CHARMM force field parameters were used for myo-inositol,[24] while the CSFF force field was used for all the other systems.[25] A study by Corzana et. al. investigated the comparative solution behaviour of several carbohydrate force fields primarily in the context of glycosidic linkages.[9] In that study the CSFF parameters was shown to produce reasonable hydration numbers compared with differential scanning calorimetry experiments for the methyl-α-D-maltoside molecule. The TIP3P model as implemented in CHARMM[26] was used to model the water molecules since this water model was used in the parameterization of both force fields. All simulations were performed as an isothermal-isobaric ensemble (NPT) using the CHARMM program,[26,27] with chemical bonds to hydrogen atoms kept fixed using SHAKE.[28] The systems were then heated to 300K over 2 ns and equilibrated for 2 ns. Each solute molecule was then analysed over separate 5ns trajectories.

The *Self Consistent Charge Density Functional Theory Tight Binding* (SCC-DFTB) has recently become a popular method for QM/MM molecular simulations[29] including carbohydrates.[30] We ran 5 ns of QM/MM SCC-DFTB dynamics for cyclohexanol, 1,2-cyclohexanediol, 1,4-cyclohexanediol, myo-inositol, β-D-xylose and β-D-glucose using CHARMM[26] v33b2. These were run with the mio-0-1[31] parameters and an improvement to hydrogen bonding interaction was included with the HBON keyword. QM/MM systems were run under the same conditions (24.8626 Å box-length, 488 water molecules with one independent solute, heated to 300K over 2 ns and equilibrated for 2 ns). However, in these simulations Ewald summations were used for the electrostatic interactions,[32] along with the TIP4P$_{FW}$ water model.[33] We use the SCC-DFTB QM/MM method to generate trajectories, as this is

the most accurate and computationally cost efficient simulation procedure available for the analyses of carbohydrate conformational motion specifically their ring puckering.[34,35]

Free Energy Perturbation

We used the Free Energy Perturbation (FEP) method to calculate the relative free energy difference between cyclohexane and each of the six cyclic systems. A non-physical path was chosen ensuring that the difference between successive perturbed states remained less than 3 kcal/mol (it varies and values are in the range of 1-2.5 kcal/mol).

Since the free energy depends only on the thermodynamic state, the results do not depend on the chosen path and the number of λ points but can be improved by increasing the overlap between the cyclohexane structure (A) and the structure B, i.e. by choosing many λ points. The free energy difference (ΔG) between states A and B is given by[36]

$$\Delta A = G(\lambda_B) - G(\lambda_A) = -k_B T \ln \frac{Z(\lambda_B)}{Z(\lambda_A)} \tag{6}$$

with partition function (Z) for NVT ensemble which is used in our simulations. This expression can be re-written in the form of

$$\Delta G = -k_B T \ln \left\langle e^{-[H(\lambda) - H(\lambda_A)]/k_B T} \right\rangle_{\lambda_A} \tag{7}$$

Here $\langle \ \rangle$ represents the ensemble average taken over the configurations representative of state A (cyclohexane) whereas H, k_B and T represent the Hamiltonian of the system, the Boltzmann constant and the temperature respectively. The free energy difference between two states can be given as the sum of the free energy differences between the intermediate steps

$$\Delta G = -k_B T \ln \sum_{i=0}^{n-1} \left\langle e^{-[H(\lambda_{i+1}) - H(\lambda_i)]/k_B T} \right\rangle_{\lambda_i} \tag{8}$$

where, $H(\lambda_{i+1})$ and $H(\lambda_i)$ are the Hamiltonians of the system at $\lambda=i+1$ and $\lambda=i$ respectively. The relative change in the binding free energy between two related (similar) compounds can be computed by using the FEP method in the thermodynamic cycle approach. All the systems for which FEPs have been performed start from the same structure, cyclohexane, thus referred to as A($\lambda=0$), while cyclohexanol, 1,2-cyclohexanediol, 1,4-cyclohexanediol, myo-inositol, xylopyranose and glucopyranose are referred as B(1). The simulations where performed at $\lambda = 0, 0.1, 0.2, 0.3, 0.4, \ldots 1$ and the consistency of the results, were confirmed by repeating the simulations in both directions $\lambda_i \rightarrow \lambda_{i+1}$ and $\lambda_{i+1} \rightarrow \lambda_i$ except at the two end points, so performing a so called double-wide sampling.[37]

Ring Pucker

The Cremer-Pople method used to define puckering of a six membered ring uses coordinates that are defined in terms of the distances z_j that are functions of 3 parameters (q_2, ϕ_2 and q_3).[38] *Converting these to a spherical coordinate frame* (Q, θ, ϕ)

$$
\begin{aligned}
q_2 \cos\phi_2 &= Q\sin\theta\cos\phi \\
q_2 \sin\phi_2 &= Q\sin\theta\sin\phi \\
q_3 &= Q\cos\theta
\end{aligned}
\tag{9}
$$

is practically more useful since the specific ring puckers can be located on specific points in a spherical surface. Here ϕ represents the azimuthal angle and lies in the range $[0,360°]$, θ puckering amplitude of range $[0, 180°]$ and Q is the radial coordinate ($Q^2 \equiv \sum_{j=1}^{N} z_j^2 = \sum_m q_m^2$; $Q > 0$) representing the total puckering amplitude. Combinations of θ and ϕ describe canonical conformations that a hexacyclic molecule can adopt, namely the two chairs, $^4C_1$ and $^1C_4$, the six boats (B) and six twists/skew shapes (S), twelve half-chairs (H) and twelve envelopes (E).

Hydroxyl Rotational and Ring Pucker Librational rates

The time-dependent correlations in spontaneous fluctuations of various conformational properties provide a measure of the flexibility of the molecular system. The conformations of importance here are the hydroxyl rotations and the ring pucker functions. The correlations for the conformational property (A) where $\delta A(t)$ is the instantaneous or spontaneous fluctuation at time zero for the time correlation function (tcf) such that

$$C(t) = \langle \delta A(t) \delta A(0) \rangle = \langle A(t) A(0) \rangle - \langle A \rangle^2 \qquad (10)$$

which can be normalized

$$C(t) = \frac{\langle \delta A(t) \delta A(0) \rangle}{\langle \delta A(0)^2 \rangle} . \qquad (11)$$

The correlation between two conformations, A and B, is found from the *cross-correlation function*:

$$C(t) = \frac{\langle \delta A(t) \delta B(0) \rangle}{\langle \delta A(0) \delta B(0) \rangle} \qquad (12)$$

Assuming that the decay of the tcf is exponential, it will fit a general form $C \exp(-t/\tau)$, where the time τ is the *correlation time*. Therefore the decay of the tcf gives an indication of the rate at which the fluctuations in conformational motion becomes uncorrelated i.e., the rate at which the conformational property of the molecule relaxes to an equilibrium.

Discussion and Results

Interatomic rdfs between non-hydrogen atoms (carbon and oxygen) in the solute molecules and the water oxygen atoms provide a radially averaged number of water molecules that are located about these solute atoms. The number of waters coordinated to the entire solute was calculated by summing the contributions from the carbon-water oxygen rdfs (Table 1 shows results calculated from the classical MD trajectory). This includes waters structured within a probability shell about the hydrophobic and hydrophilic sites. These coordinated waters derived from the static structural

distribution functions are large in number and represent the average hydration shell structure. They are however, not dynamically associated with the solute or its directly calculable hydration properties. Another measure of the general static hydration structure about a solute is to use the sdfs calculated relative to an average ring conformation. This gives an anisotropic description of the most probable water positions about the solute. In Figure 5.2, the spatial distribution functions (sdfs) for the six molecules are shown calculated from the classical MD trajectory. In the simplest case where there is only one hydroxyl (cyclohexanol in Figure 5.2 a), the most probable water positions are located in the three (t, g^+ and g^-) sterically favoured hydroxyl positions. When adding a second hydroxyl group on the opposite end of the cyclohexane ring, the same three high probility positions are seen about each of the two hydroxyls in the 1,4-cyclohexanediol sdf (Figure 5.2 b). However, when placing the two hydroxyls adjacent to each other as in 1,2-cyclohexanediol the high probability waters do not show up clearly in the sterically favoured hydroxyl positions, instead a new half moon shaped probability density appears between the two hydroxyls (Figure 5.2 c). Previously we had shown that the intermolecular hydrogen bonds that form between functional groups on pyranose rings and the surrounding water are very close in energy to the intramolecular hydrogen bonds formed between these functional groups.[3] The intramolecular hydrogen bond strength between two hydroxyl groups is marginally stronger (0.5kcal/mol) than the intermolecular hydroxyl-water hydrogen bond strength, leading to the two types of hydrogen bonds competing with each other.[4] When there are OH-groups on every ring atom, as is the case with myo-inositol, where all hydroxyls are in equatorial positions except the one on carbon 2 where the OH is axial, most of the high probability density volumes are distributed in a half moon shape (Figure 5.2 d). While the increases in the water probability density volumes for these four molecules correspond to the increase in their hydroxyls, there is no correlation

between the increase in water their probability densities (Figure 5.2 a-d) and their relative solubility (Figure 5.1).

Further evidence supporting this observation is bourne out when we analyse the probability densities
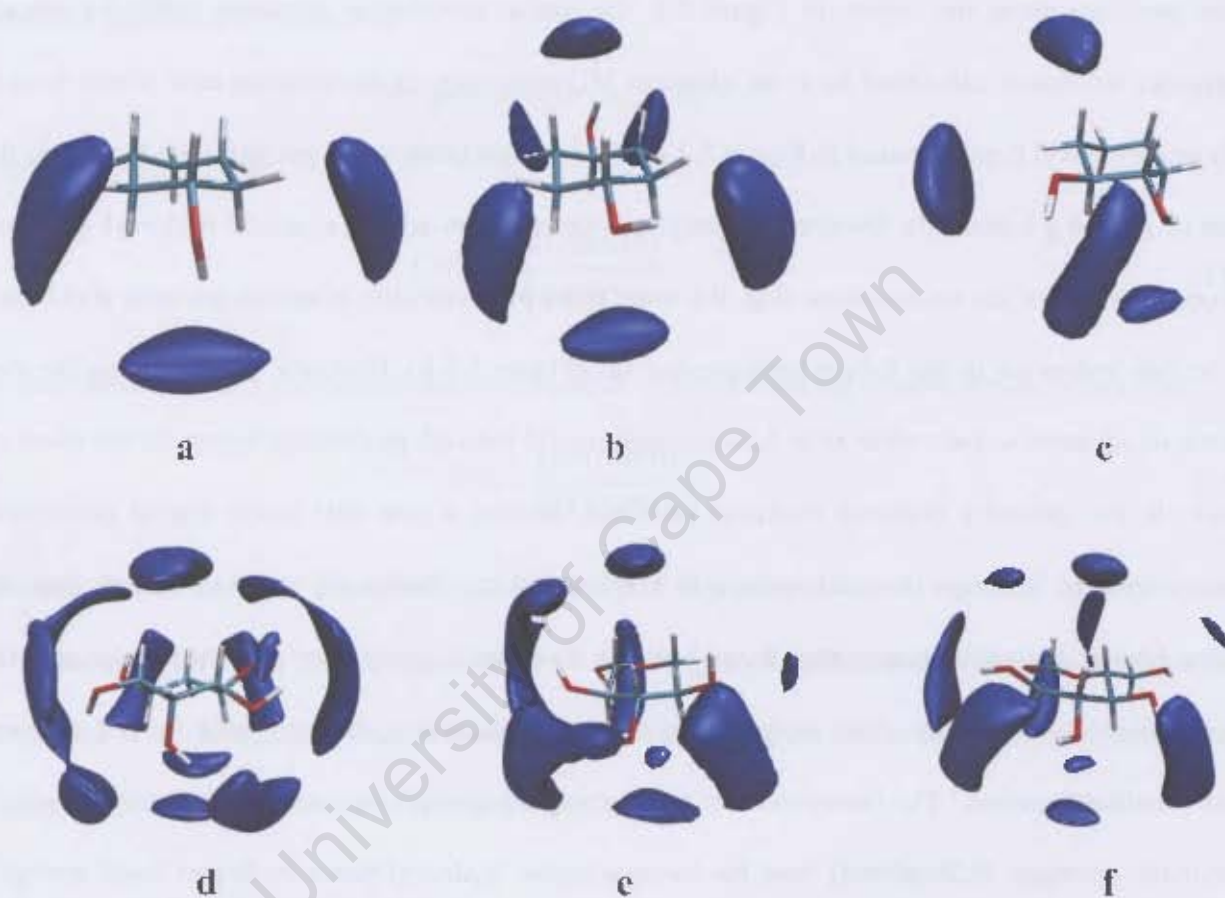
Figure 5.2. SDFs of the six aqueous systems, contoured at 50% above bulk density: a) cyclohexanol; b) 1,4-cyclohexanediol; c) 1,2-cyclohexanediol; d) myo-inositol; e) β-xylopyranose; f) β-glucopyranose.

of xylopyranose (Figure 5.2 e) and glucopyranose (Figure 5.2 f). Our results presented here provide the same observations of the anisotropic water structure about these carbohydrate molecules as have been previously reported by Brady et. al.[39] Here our attention is drawn to the similarity in the overall

water probability densities of myo-inositol, xylopyranose and glucopyranose while their solubilities are dramatically different. *This leads us to conclude that insight into the hydration of these cyclic molecules must be found in the dynamic properties of the solute and solvent not in the static distributions such as rdfs and sdfs.*

Firstly, we measure from ultrasound experiments hydration numbers for the molecular series. For each aqueous system, seven different concentrations (except in case of cyclohexanol, due to its low solubility) in the range 0.1-0.7 M [mol/L] were prepared and 150 measurements were performed on each concentration. The hydration numbers at infinite dilution have been obtained by extrapolating the linear-fit of the n_h(M) data down to infinite dilution. Plots of the experimental results for k_s and n_h versus M of the aqueous solutions of hydroxy-cyclic compounds, at 300K, are reported in Figure 5.3 a and 3 b respectively. Even though the ultrasound measurements took place over a very wide range of molarity we did observe a monotonically decreasing linear change of compressibility with composition (Figure 5.3 a), which is in agreement with previous reports.[10]

On facevalue a straightforward relationship exists between the hydration properties of the entire series (Figure 5.3 b), that is an increase in number of OH groups delivers an increase in the n_h. However, the increase is not linearly proportional to the increase in the hydration property. When we normalize the data relative to n_h for a single hydroxyl i.e., cyclohexanol (Table 2), we observe that the n_h for myo-inositol which has 6 hydroxyls is only twice that of cyclohexanol which has one. This relatively small increase in n_h for a system having all its carbons hydroxylated illustrates that simply adding multiple hydrophyllic groups capable of forming hydrogen bonds with water does not result in a linear increase in hydration. The n_h for xylopyranose increases 3.4 times that of cyclohexanol which is greater than the increase observed for myo-inositol even though it possesses only four hydroxyls.
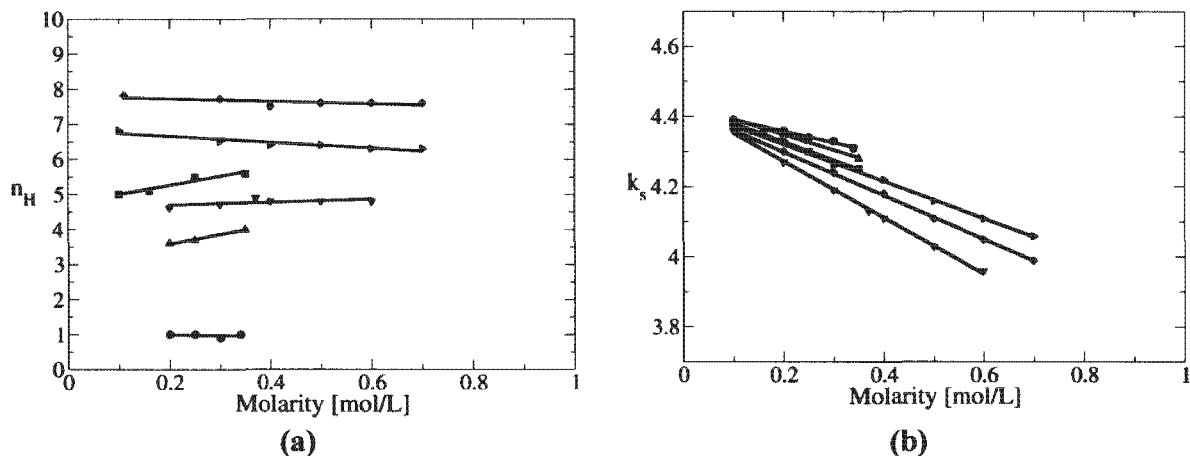
Figure 5.3a. Compressibility k_s [m^2/N] vs Molarity [mol/L]. The compressibility for water alone is $4.42 \cdot 10^{-10}$ m^2/N: (●) cyclohexanol, (■) 1,2-cyclohexanediol, (▲) 1,4-cyclohexanediol, (►) xylopyranose, (♦) glucose, (▼) myo-inositol. **5.3b.** Hydration numbers of the four molecules plotted vs molarity [mol/L]. Concentrations for cyclohexanol, 1,2-cyclohexanediol and 1,4-cyclohexanediol are limited, due to their low solubility in water: (●) cyclohexanol, (■) 1,2-cyclohexanediol, (▲) 1,4-cyclohexanediol, (►) xylopiranose, (♦) glucose, (▼) myo-inositol.

The greater change is therefore due to the introduction of an ether group into the ring. A further dramatic increase in hydration number continues for glucose where the structural change is the addition of a hydroxymethyl group.

Computer simulations 5ns in length using a classical potential and parameters from CSFF[25] as well as hybrid quantum classical (QM/MM) simulations using the SCC-DFTB implementation in CHARMM[40] were performed for each molecule in the series. The n_h values were calculated from these simulations and then normalized relative to that of cyclohexanol (Table 2). The hydration number ratios from the classical MD agree with experimental results within 1-σ (cyclohexanol, 1,2-diol, myo-inositol) and 2-σ (1,4-diol, xylopyranose, glucopyranose). The QM/MM results compare with the experimental values within 2-σ (cyclohexanol, 1,2-diol and 1,4-diol) and 3-σ (myo-inositol,

xylopyranose and glucopyranose). While magnitude of the experimental and simulation values differ, the trends are similar. They confirm that the n_h of a molecule has a minor dependence on the number of hydrophilic sites while there is a significant correlation with the introduction of a heteroatom to the ring. From this we conclude that a valid analysis of the underlying reasons for this trend can be deduced from the simulation trajectories.

The hydration free energies for each molecule in the series were calculated (using classical force fields) relative to cyclohexane where they show a negative ΔG_h so indicating greater solubility (Table 3). Although cyclohexanol is mostly hydrophobic with only a single hydroxyl it improves on the completely hydrophobic cyclohexane's free energy of hydration by -7.10 kcal/mol. By adding a second OH group to the cyclohexane frame, not adjacent to the first hydroxyl (1,4-cyclohexanediol), the ΔG_h improves by more than 4 kcal/mol compared with cyclohexanol. In the case of 1,2-cyclohexanediol where the hydroxyls are adjacent and capable of forming intramolecular hydrogen bonds, the increase in hydration free energy is less, i.e., ~3 kcal/mol. The least dramatic improvement in free energy is for the fully hydroxylated cyclohexane ring (myo-inositol) where ΔG_h = -12.70 kcal/mol increasing by ~1 kcal/mol on the 1,4-cyclohexanediol value. The free energy pattern for the cases where hydroxyls are incrementally added to cyclohexane rings supports the hydration measurements and follows the solubility obervations seen in Figure 1.

The n_h measurements and calculations show similar although not as dramatic an increase, as observed for solubility, when progressing from cyclohexane based molecules to the pyranose based molecules. The free energy data supports the same pattern where ΔG_h = -12.70 kcal/mol for xylopyranose is more than 3 kcal/mol stronger than myo-inositol even though it has fewer hydroxyl groups. This increase in free energy of hydration must therefore be due to the introduction of an ether group into the cyclohexane ring especially since the number of hydrophilic groups has decreased in

the pyranose system. Finally the most dramatic increase in the free energy of hydration i.e., 6 kcal/mol is due to the addition of a primary alcohol to the xylopyranose structure resulting in the glucopyranose molecule. This simple addition of a single hydroxymethyl group gives a three-fold increase in ΔG_h mimicking the dramatic increase in solubility seen in Figure 1.

The comparitive diffusion rates (Table 4) supports the n_h and ΔG_h trends such that cyclohexanol diffuses the fastest and glucose diffuses the slowest in water. Since these systems are of similar shape and size the slower diffusion of glucose implies that it drags a larger sheath of waters along with it compared with molecules that have greater numbers of hydroxyl groups (i.e., myo-inositol). Waters are able to form long term hydrogen bonds with hydroxyls that diffuse slower than the solvent diffusion rate. The diffusional rate of the solute hydroxyls depends on the solute's translational and rotational diffusion as well as the conformational motion leading to further hydroxyl positional diffusion. Solute molecules that are several times larger than water diffuse slower than the water diffusion rate. Therefore this will not be the cause of prematurely interrupted intermolecular hydrogen bonding between the solute and water. All evidence therefore points to a hydration model that does not depend on the number of hydrophilic groups but on the conformational properties of the system. The two sources of conformational motion that affect the rotational and positional diffusion of the hydroxyls are the rotation of the hydroxyl and the puckering of the hexacyclic ring (Figure 5.4). *The greater the rotational and positional diffusion of the hydroxyls, the shorter the lifetime of the intermolecular hydrogen bonds that they make with the surrounding water molecules.[2] This is the major factor that determines hydration number.*

To address the relative rotational behaviour of secondary and primary alcohols on cyclohexanol, 1,2-cyclohexanediol, 1,4-cyclohexanediol, myo-inositol, β-D-xylopyranose and β-D-glucopyranose we calculated tcfs for the hydroxyl groups' rotations from the force field simulations as these have

been accurately parameterized against experimental data for this property. From the tcfs we then calculated relaxation times (τ) which are listed in Table 5. We labelled hydroxyl groups in sequence around the ring so hydroxyl i is adjacent to $i+1$ and for non pyranose rings $i=6$ is adjacent to $i=1$. It can be seen that hydroxyl groups that are adjacent to each other rotate more slowly because of the bridging water hydrogen bonds[4] that form the half moon shaped high probability water densities. This accounts for 1,2-cyclohexanediol diffusing slower than 1,4-cyclohexanediol. However, the doubling of hydroxyl groups only marginally slows down the diols compared with cyclohexanol. In particular for myo-inositol where the cyclohexane ring is fully hydroxylated the diffusion rate is only 23% slower than that of cyclohexanol. The even slower diffusion of xylopyranose indicates that a slowdown in conformational motion of the hydroxyls must be the cause of an increase in the lifetime of the hydration about the ring.

The hydroxyl rotation of xylopyranose is similar to the cyclohexane based molecules (Table 5) The major change from cyclohexane to pyranose ring systems is due to the slowdown in the positional diffusion of the hydroxyl groups. Since this is not sourced from the rotational frequency of the hydroxyls the change in their positional motion must be due to the ring to which they are tethered (Figure 5.4 b).
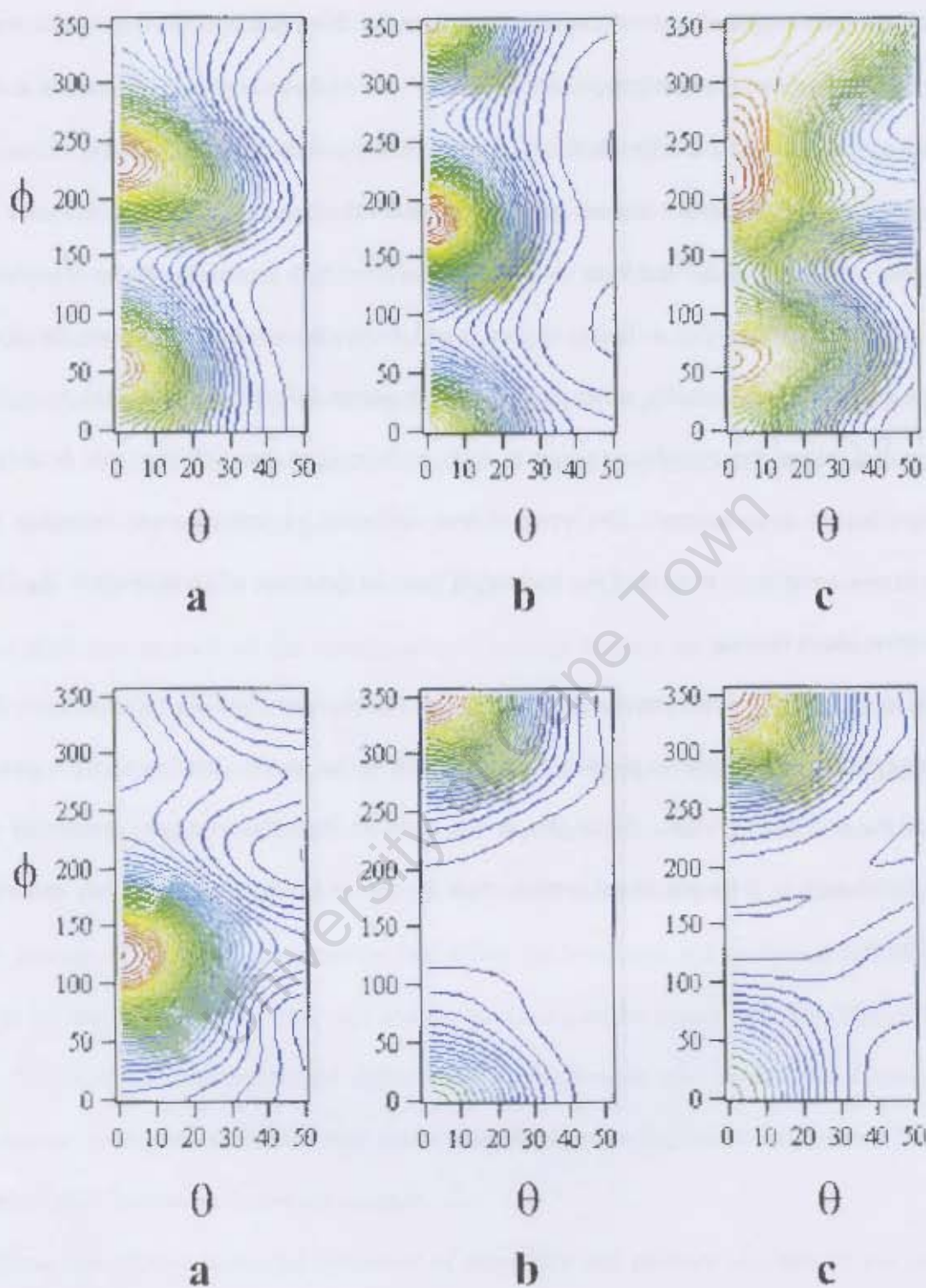
Figure 5.4. Pucker parameter probability distributions for a) cyclohexanol; b) 1,4-cyclohexanediol; c) 1,2-cyclohexanediol; d) myo-inositol; e) β-xylopyranose; f) β-glucopyranose.

An analysis of the extent of the ring puckering for cyclohexane and pyranose ring systems was done using a Cremer-Pople spherical polar coordinate definition[38] for hexacyclic ring pucker as described above. Here we only used the SCC-DFTB trajectories as these have been shown to be the most reliable for this property.[35] The most stable pyranose conformations are ${}^4C_1$ and ${}^1C_4$ chairs.[41] Our puckering results are in agreement with this so that all the rings studied here adopt the expected ${}^4C_1$ chair conformation (i.e., $\theta \sim 0°$ and $\phi \in (0, 360)$. The probability distribution of cyclohexane ring pucker has a wide range for ϕ (Figure 5 a-d) while in the case of the pyranose rings the distribution is more confined (Figure 5 e and f). We calculated the frequency of ring conformational motion using the pucker tcfs and then derived the corresponding relaxation times for both the zenithal (τ_θ) and azimuthal angles (τ_ϕ). The difference in ring pucker flexibility between the cyclohexanes and pyranoses is very clear (Table 6). The pyranose rings pucker more slowly compared with compounds that have a cyclohexane ring base. The slowdown from myo-inositol (τ_ϕ=4.5) to xylopyranose (τ_ϕ=6.6) is significant and accounts for water associating longer with the slower moving hydroxyls on the monosaccharide leading to the observed increase in hydration from n_h =3.7 to n_h = 6.4. The glucose ring pucker is about the same as that of xylopyranose (τ_ϕ= 6.8). However, it is the only molecule with a slowly rotating hydroxymethyl group ($\tau \sim 80$). Previously, we had shown that the hydroxymethyl group participates in numerous hydrogen bonds with water. In particular when in the gt conformation water molecules are trapped between the primary alcohol and the ring oxygen, leading to water forming bridging hydrogen bonds.[42] The combination of these relatively rigid conformational features found in glucose is the reason for its high hydration (n_h = 7.2) and correlates strongly with its high solubility in water.

Conclusions

We have shown that the number of hydroxyls on a flexible non-aromatic ring is not linearly proportional to its relative hydration. The H_2O molecules shared between adjacent OH groups tend to form complex bridging hydrogen bonds that are observed in the rdf and sdf probability densities. However, the residence time of the water molecules in the volumes close to the solute is more fundamental to the definition of hydration than is the probability density of finding the water in a hydrogen bonding location. This is because the energetic difference between intermolecular water be be the hydroxyl hydrogen bonds and intramolecular hydroxyl-hydroxyl hydrogen bonds is less than 1 kT. Therefore the competition between these two hydrogen bond types leads to frequent exchanges between them when the hydroxyl group has fast rotational and positional diffusion.

The most significant contribution to sustained intermolecular (water-hydroxyl) hydrogen bonds and therefore increased n_h, is the positional diffusion of the hydroxyls. This is closely related to the frequency with which the ring, to which it is tethered, puckers. Here frequency of pyranose ring pucker is significantly slower than that of cyclohexane ring pucker. The hydroxyls therefore librate slower on the former ring and so increases the possibility of hydrogen bonding with surrounding water molecules for longer times. This rationale for hydration is bourne out when we measure the water residence times about hydroxyls bonded to a cyclohexane frame compared with water residence times about hydroxyls bonded to pyranose rings. Conformational rigidity is therefore the reason for improved hydration in these cyclic systems bearing hydrophilic groups. Moreover, the presence of a conformationally rigid ring and a rigid hydroxymethyl group that slowly rotates accounts for the increase in hydration number of glucose compared with that of xylopyranose. These measurements and calculations show that conformational rigidity is responsible for an enhanced interaction between

water and monosaccahride molecules that corresponds well with the observed solubility trends for hexacyclic molecules.

Acknowledgements

References

(1)   Jensen, K. P.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2006**, *2*, 1499.

(2)   Astley, T.; Birch, G. G.; Drew, M. G. B.; Rodger, P. M. *J. Phys. Chem. A.* **1999**, *103*, 5080.

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

(4)   Naidoo, K. J.; Chen, Y.-J. *Mol. Phys.* **2003**, *101*, 2687.

(5)   Franks, F. *Pure and Applied Chemistry* **1987**, *59*, 1189; Brady, J. W.; Schmidt, R. K. *J. Phys. Chem.* **1993**, *97*, 958.

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

(7)   Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 292.

(8)   Corzana, F.; Motawia, M. F.; Du Penhoat, C., H.; Perez, S.; Tschampel, S. M.; Woods, R. J.; Engelsen, S. *J. Comput. Chem.* **2004**, *25*, 573; Engelsen, S. B.; Monteirob, C.; Herve de Penhoatb, C.; Perez, S. *Biophysical Chem.* **2001**, *93*, 103.

(9)   Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Cuello, G.; Brady, J. W. *J. Chem. Phys.* **2006**, *125*, 224505(1.

(10)   Galema, S. A.; Hoiland, H. *J. Phys. Chem.* **1991**, *95*, 5321.

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

(12)   Buchner, R.; Capewell, S. G.; Hefter, G.; May, P. M. *J. Phys. Chem. B* **1999**, *103*, 1185; Kaatze, U.; Hushcha, T. O.; Eggers, F. *J. Soln. Chem.* **2000**, *29*, 299.

(13)   Bockris, J. O.; Reddy, A. K. N. *Modern Electrochemistry 1: Ionics*; Plenum: New York, 1998; Vol. 1.

(14)   Hajduk, P. J.; Horita, D. A.; Lerner, L. E. *J. Am. Chem. Soc.* **1993**, *115*, 9196; DiBari, M.; Cavatorta, F.; Deriu, A.; Albanese, G. *Biophys. J.* **2001**, *81*, 1190; Brady, J. W. *J. Am. Chem. Soc.* **1989**, *111*, 5155.

(15)   Ishihara, Y.; Okouchi, S.; Uedaira, H. *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 3337.

(16)   Uedaira, H. *Cell. Mol. Biol.* **2001**, *47*, 823.

(17)   Graziano, G. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3567.

(18)   Qvist, J.; Halle, B. *J. Am. Chem. Soc.* **2008**, *130*, 10345.

(19)   Rezus, Y. L. A.; Bakker, H. J. *Phys. Rev. Lett.* **2007**, *99*.

(20)   Chavez, A. L.; Gordon, G. B. *Chem. Senses* **1997**, *22*, 149.

(21)   Shiio, H. *J. Am. Chem. Soc.* **1958**, *80*, 70.

(22)   Burakowsky, A.; Glinski, J. *Chem. Phys.* **2007**, *332*, 336.

(23)   Vishnyakov, A.; Lyubartsev, A. P.; Laaksonen, A. *J. Phys. Chem* **2001**, *105*, 1702.

(24)   Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. *J. Chem. Theory Comp.* **2009**, *5*, 2353.

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

(26)   Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comp. Chem.* **2009**, *30*, 1545.

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

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

(29)   Zhou, H.; Tajkhorshid, E.; Frauenheim, T.; Suhai, S.; Elstner, M. *Chem. Phys.* **2002**, *277*, 91; Xu, D.; Guo, H.; Cui, Q. *J. Am. Chem. Soc.* **2007**, *129*, 10814; Woodcock, H. L.; Hodoscek, M.; Brooks, B. R. *J. Phys. Chem. B* **2007**, 5720.

(30)   Barnett, C. B.; Wilkinson, K. A.; Naidoo, K. J. *J. Am. Chem. Soc.* **2010**, *132*, 12800.

(31)   Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B.* **1998**, *58*, 7260.

(32)   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.

(33)   Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665.

(34)   Barnett, C. B.; Naidoo, K. J. "Calculating Ring Pucker Free Energy Surfaces From Reaction Coordinate Forces"; Theory and Applications of Computational Chemistry 2008, 2009, Shanghai (China); Barnett, C. B.; Naidoo, K. J. *Mol. Phys.* **2009**, *107*, 1243.

(35)   Barnett, C. B.; Naidoo, K. J. *J. Phys. Chem. B* **2010**, *114*, 17142.

(36)   Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420.

(37)   Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050.

(38)   Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *96*, 1354

(39)   Schmidt, R. K.; Karplus, M.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 541; Liu, Q.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 12276; Lui, Q.; Brady, J. W. *J. Phys. Chem. B* **1997**, *101*, 1317.

(40)   Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B.* **2001**, *105*, 569.

(41)   Angyal, S. J. *Angew. Chem. Int. ed.* **1969**, *8*, 157.

(42)   Barnett, C. B.; Naidoo, K. J. *J. Phys. Chem. B* **2008**, *112*, 15450.

TABLE SECTION

TABLE 1: Number of coordination waters (n_{solute}) around each solute, obtained by applying the RDF to each heavy atom (C and O) of the molecules .

Compounds	$n_{solute}[g(r)_{min}]$	ratios
cyclohexanol	33	1
1,2-diol	38	1.2
1,4-diol	37	1.1
myo-inositol	67	2
xylopyranose	70	2.1
glucopyranose	89	2.7

TABLE 2: Hydration numbers from ultrasound measurements and computer simulations.[c]

Compounds	Raw Ultrasound n_h	Ratios Ultrasound n_h	Classical n_h	QM/MM n_h
cyclohexanol	1.9±1.6	1	1	1
1,2-diol [a]	2.7±1.3	1.4±2.0	1.9	0.89
1,4-diol	3.8±1.5	2±2.1	1.7	0.98
myo-inositol	3.7±0.6	1.9±1.6	3.4	1.1
xylopyranose	6.4±1.0	3.4±1.8	4.8	1.4
glucopyranose [b]	7.2±0.7	3.9±1.7	6.3	2.6

[a] 1,2-cyclohexanediol may undergo polymerization in water affecting the ultrasound measurement

[b] The β-glucose hydration number agrees with studies in ref [10]

[c] Errors on each measurement have been calculated from the propagation of errors (RMS) checking for a χ^2 distribution.

TABLE 3: Free Energies for systems are perturbed from cyclohexane (cyclo) calculated from classical MD simulations.

	$\Delta\Delta G$ [kcal/mol]
$\Delta\Delta G_{cyclo \rightarrow cyclohexanol}$	-7.10
$\Delta\Delta G_{cyclo \rightarrow 1,2diol}$	-10.3
$\Delta\Delta G_{cyclo \rightarrow 1,4diol}$	-11.4
$\Delta\Delta G_{cyclo \rightarrow myo\text{-}inositol}$	-12.7
$\Delta\Delta G_{cyclo \rightarrow xylopyranose}$	-15.8
$\Delta\Delta G_{cyclo \rightarrow glucopyranose}$	-21.7

TABLE 4: Linear diffusion coefficients for hexacyclic molecules calculated from classical MD simulations.

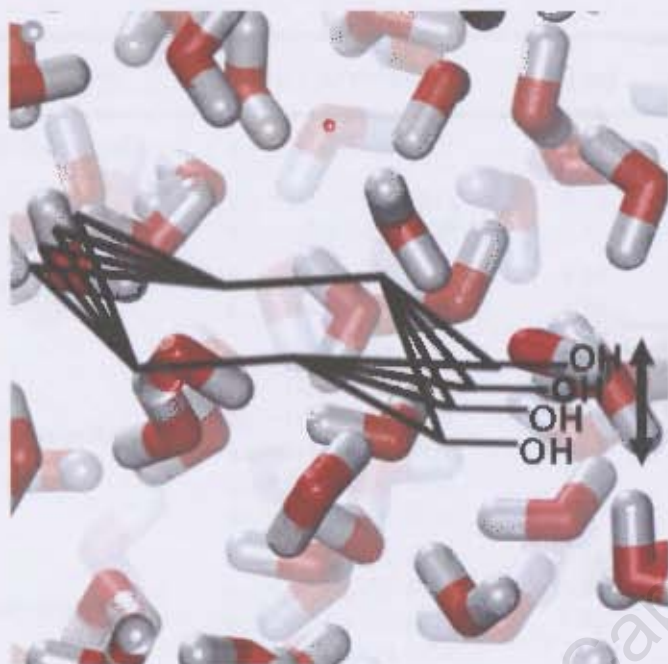| Compounds	$D / cm^2 \cdot s^{-1}$	Normalized D
cyclohexanol	$1.51 \pm 0.34 \cdot 10^{-5}$	1
1,2-diol	$1.38 \pm 0.02 \cdot 10^{-5}$	0.913
1,4-diol	$1.47 \pm 0.02 \cdot 10^{-5}$	0.974
myo-inositol	$1.17 \pm 0.01 \cdot 10^{-5}$	0.774
xylopyranose	$1.13 \pm 0.003 \cdot 10^{-5}$	0.748
glucopyranose	$0.922 \pm 0.01 \cdot 10^{-5}$	0.610

TABLE 5: Auto and Cross Correlation Functions for hydroxyl and hydroxymethyl rotations calculated from classical force field simulations.

Compounds	Auto Correlation	τ [ps]	Cross Correlation	τ [ps]
cyclohexanol	ϕ_{11} = H1-C1-O1-HO1	1.34	--	--
1,2-diol	ϕ_{11} = H1-C1-O1-HO1 ϕ_{22} = H2-C2-O2-HO2	4.73 3.11	ϕ_{12}	3.71
1,4-diol	ϕ_{11} = H1-C1-O1-HO1 ϕ_{22} = H4-C4-O4-HO4	1.95 1.56	ϕ_{12}	0.98
myo-inositol	ϕ_{11} = H1-C1-O1-HO1 ϕ_{22} = H2-C2-O2-HO2 ϕ_{33} = H3-C3-O3-HO3 ϕ_{44} = H4-C4-O4-HO4 ϕ_{55} = H5-C5-O5-HO5 ϕ_{66} = H1-C1-O1-HO1	6.72 9.75 8.28 10.5 5.93 5.94	ϕ_{12} ϕ_{23} ϕ_{34} ϕ_{45} ϕ_{56} ϕ_{61}	10.8 5.70 5.70 4.50 7.60 10.9
xylopyranose	ϕ_{11} = H1-C1-O1-HO1 ϕ_{22} = H2-C2-O2-HO2 ϕ_{33} = H3-C3-O3-HO3 ϕ_{44} = H4-C4-O4-HO4	4.70 5.64 6.40 5.14	ϕ_{12} ϕ_{23} ϕ_{34} ϕ_{41}	8.90 8.88 8.89 14.9
glucopyranose	ϕ_{11} = H1-C1-O1-HO1 ϕ_{22} = H2-C2-O2-HO2 ϕ_{33} = H3-C3-O3-HO3 ϕ_{44} = H4-C4-O4-HO4 ϕ_{55} = C5-C6-O6-HO6 ϕ_{66} = H61-C6-O6-HO6 ϕ_{77} = O5-C5-C6-O6 ϕ_{88} = C4-C5-C6-O6	4.45 4.60 6.47 12.2 7.80 8.10 82.5 75.8	ϕ_{12} ϕ_{23} ϕ_{34} ϕ_{45} ϕ_{56} ϕ_{68}	6.76 12.7 13.4 36.7 10.9 69.4

TABLE 6: Relaxation time of θ and φ angles for puckering analysis calculated from the SCC-DFTB trajectories.

Compounds	τ_θ [ps]	τ_ϕ [ps]
cyclohexanol	$0.072 \pm 3.2 \cdot 10^{-4}$	2.6 ± 0.037
1,2-diol	2.4 ± 0.13	3.9 ± 0.18
1,4-diol	3.8 ± 0.13	3.9 ± 0.17
myo-inositol	4.8 ± 0.21	4.5 ± 0.23
xylopyranose	4.6 ± 0.15	6.6 ± 0.86
glucopyranose	6.7 ± 0.49	6.8 ± 2.7

TOC for manucript "The Extent of Conformational Rigidity Determines Hydration in Non Aromatic HexaCyclic Systems"

Chapter 6
Glycosidase-inhibitors

In this chapter, an important class of compounds is presented, whose main function is the inhibition of carbohydrate-degrading enzymes. More explicitly, glycosidase-inhibitors are described. These molecules have been found to convert carbohydrates into simple sugars, i.e. monosaccharides, which can then be absorbed through the intestine. Hence, glycosidase-inhibitors can reduce the impact of carbohydrates on blood sugar and to act as glycemic controller over hyperglycemia, mainly in *diabetes mellitus type 2*. Therefore, it is clear their importance as anti-diabetic drugs.

From a structural point of view, they show many similarities with carbohydrates, as they are six-membered non-aromatic heterocyclic rings, in position 6 they have nitrogen, whereas carbohydrates have oxygen. **Figure 6.1** represents nojirimycin (a) and glucopyranose (b). Nojirimycin is one of the N-sugars studied in this thesis.

a   b

Figure 6.1 Nojirimycin (a) and β-D-glucopyranose (b). They differ for the atom in position 6, where nojirimycin has nitrogen (blue), while glucopyranose has oxygen (red).

Particular attention has been here devoted to understanding similarities and differences in hydration and, more particularly, in conformational behaviour within the two classes of sugars.

The dynamic hydration and conformational properties of glycosidase-inhibitors have been investigated through the same simulation tools (PDF, SDF, Residence Time, Ring Puckering analysis), which are extensively described in **Chapter 3**, and the obtained preliminary results are here given.
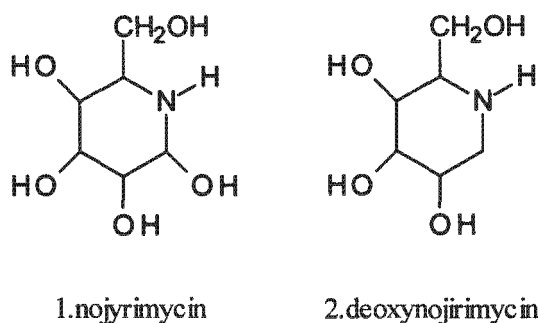
1.nojyrimycin   2.deoxynojirimycin

Figure 6.2 Structures of the two glycosidase-inhibitors compounds studied in this thesis: 1.Nojirimycin; 2. Deoxynojirimycin. They differ for one OH group on C1.

6.1 Glycosidases-inhibitors

This class of compounds have been found to inhibit carbohydrate-degrading enzymes, because of their structural resemblance to the sugar moiety of the natural substrate [1, 2]. The carbohydrate-degrading enzymes are involved in a wide range of important biological processes (intestinal digestion, postranslational processing of the sugar chain of glycoproteins, control mechanisms, lysosomal catabolism of glycoconjugates and some viral infections) while they also show very powerful therapeutic function in diabetes and lysosomal storage disorders treatments.

The first natural glucose mimic of the series to be discovered was nojirimycin (NJ, structure *1* in both **Figures 6.2** and **6.3**) [1, 3]. Produced by *Streptomyces* species, this compound shows a strong inhibition activity of both α- and β-glycosidases, having a greater inhibitory effect on β-glycosidase. NJ is believed to exist in solution as a mixture of two anomers (α and β) [4]. NJ has been observed to be stable for at least 4 hours at 4°C, where after it undergoes hydrolyses [4]. For this highly unstable nature, it can be stored as a bisulfite adduct or alternatively it can be reduced to 1-deoxynojirimycin (DNJ, structure *2* in both **Figures 6.2** and **6.3**). DNJ was isolated from the roots of mulberry trees (melanoline) and was further found to be produced by *Bacillus* and *Streptomyces*. It has potent inhibition activity for the majority of α-glycosidases [1].

NJ and DNJ are only two examples of this class of inhibitors that have been found to be active and which have been reviewed by Asano *et al.* [5].
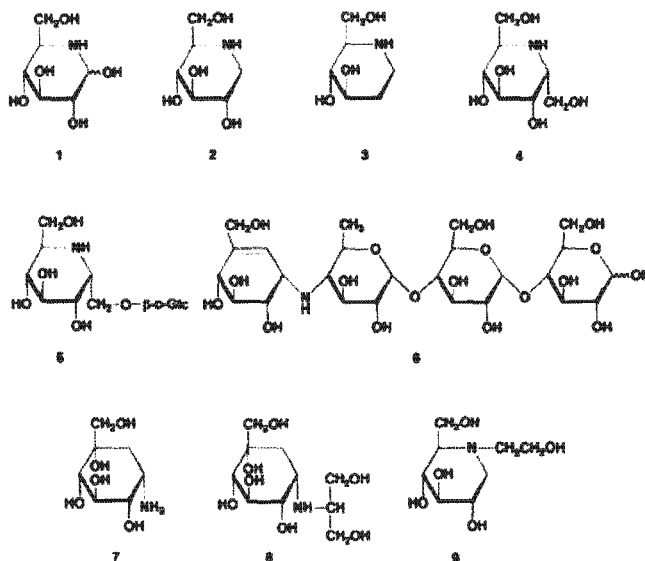
Figure 6.3 Snapshot from reference [1], which shows a series of glycosidases-inhibitors. Structures *1* and *2* are nojirimycin (NJ) and 1-deoxynojirimycin (DNJ), respectively; *3* is fagomine (or dideoxynojirimycin), *4,* α-homo-nojirimycin (HJN); *5,* 7-O-β-D-glucopyranosyl-HJN (glc-HNJ); *6,* acarbose; *7,* valiolamine; *8,* voglibose; *9,* miglitol.

A representation of some of these molecules is reported in **Figure 6.3**, which has been taken from reference 1 [1].

The reduction of DNJ produces 1,2-dideoxynojirimycin (or fagomine, structure *3* in **Figure 6.3**)which has shown interesting inhibition activity not observed with DNJ: when present the release of immunoreactive insulin (IRI) increases in a concentration-dependant-manner.

The importance of studying this particular class of compounds is largely from a medical point of view. Sugar mimics with inhibitory activities toward carbohydrate-degrading enzymes have potential as anti-diabetics, anti-obesities, antivirals, and therapeutic agents for some genetic disorders. Only a few of these inhibitors are currently marketed. They are *acarbose* (which under the trade name of Glucobay), *voglibose* (Basen), and *miglitol* (Glyset) which are used for the treatment of type 2 diabetes, which are on the market [2]. Their structures are reported in **Figure 6.3** (structures *6, 8* and *9*, respectively). Understanding of these compounds may lead to an improved understanding of their inhibitory activities which, in turn, could make them more suitable for treating current disorders.

6.1.1 The Role of Inhibitors

In mammals, glycogen exists both as a cytosolic and a lysosomal form, and is broken down by different pathways: while the cytosolic form is cleaved by glycogen phosphorylase (GP) and the debranching enzyme amylo-1,6-glucosidase (1,6-GL), the lysosomal form is degraded by a single enzyme, acid (lysosomal) α-glucosidase.

The liver is the predominant source of blood glucose. It is generally recognized that the hepatic glucose output in type 2 diabetes is elevated and thus significantly contributes to hyperglycemia.

A possible way to suppress hepatic glucose production and lower blood glucose in type 2 diabetes may be through inhibition of GP.

DNJ has been shown to have no inhibition toward GP but it was a potent inhibitor for 1,6-GL [1, 2].

6.2 Parameterization Details.

As reported in **Chapter 2, Section 2.4.1**, a crucial ingredient in performing simulations is a suitable initial structure. In the cases of both NJ and DNJ, their X-ray structures (PDB codes: *DEZRAB* [6] and *YAXYAX* [7], respectively) have been used as starting point for their respective simulations. A note on NJ: as aforementioned (**Section 6.1**), NJ is unstable at room temperature and therefore it has to be stored as bisulphite adduct. The PDB *DEZRAB* refers indeed to the substituted NJ, which is characterised by the presence of SO_3^- on C1. Consequently, a first step foresaw an adaptation of the structure. More explicitly, the bisulphite group has been changed with an OH group. This is an allowed and generally used procedure, in case the starting structure refers to a very similar compound and only very small changes need to be done.

Once the starting geometries have been set, the following step included an accurate check on the force field (in this case, CSFF) parameters. Unfortunately, the CSFF lacks of iminosugars parameters, and more specifically it lacks of nitrogen atom type. It follows that a parameterisation of the latter has been required and consequently it led to perform an implementation of the N-sugars functions in the CSFF.

The parameterisation process has taken the advantage of the availability of a complete parameters set for a very similar compound, pyrrolidine (**Figure 6.4-a**). This molecule is included in the general force field (CGenFF) for CHARMM, developed by Mackerell and co-workers [8]. This led to choose the "*by analogy*" parameterisation procedure (see **Chapter 2, Section 2.4.5**).

The adaptation of the CGenFF parameters for the CSFF foresaw the following steps:

• A deoxynojirimycin structure has been built by using a molecule builder (ArgusLab [9], in this case).

• On that structure, an optimisation step through *Gaussian* [10], at both HF/6-31G(g) and MP2/6-31G(g) levels of theory, has been performed.

• Frequency analysis on the resulting geometry has been performed, in order to determine analytically the second derivatives of the potential energy surfaces. This is needed to confirm that the geometry is at its minimum rather than in a transition state (in the latter case, at least one negative frequency is obtained).

• Normal mode analysis with the optimised geometry, for both QM levels of theory, has been performed.

• CHARMM *Vibran* routine for normal mode analysis has been run.

• QM frequencies were compared with MD values (obtained in CHARMM), after applying a scale factor (0.943) to the MP2/6-31G(g) values, in order to be comparable with CHARMM [8].

• Multiple steps of comparison between MD and QM frequencies have been done. The aim is to obtain similar values for MD and QM normal modes by modelling values for angles and dihedrals. A general target is to get the vibrational frequencies to be within 5% (on average) of their MP2 values [8].

• A comparison of the obtained normal modes is reported in **TABLE 6.5**, at the end of the text. Although it is not easy to make the right comparison between different levels of theories and experimental IR frequencies, here an attempt is given. Usually, the high frequencies around 2850-3000 cm^{-1} are assigned to the stretching of aliphatic C-H bonds. Around 3000-3500 cm^{-1} there are the characteristic absorption bands of N-H. Close to this region, O-H also absorbs, giving possible overlaps which can be difficult to be assigned. A 3371 cm^{-1} absorption band has been assigned to the N-H stretching. At 1653, the N-H bending can be observed (a note: this information is absent in MP2/6-31G(g)). The peaks at 1407 cm^{-1}, 1108 cm^{-1}, and 1027 cm^{-1} have been assigned to the C-N stretching, C-O bending C-O stretching, respectively [11].

Similar parameterisation for NJ was not required. DNJ and NJ differ only in an OH group on C1. Therefore, it is possible to assume that the values for DNJ are suitable for NJ, as well as for the entire series of iminosugars which are structurally close to DNJ.
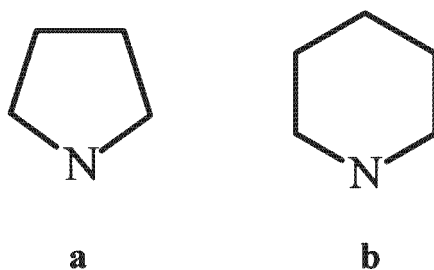
Figure 6.4 The "*by analogy*" parameterisation procedure adopted in this thesis started from pyrrolidine (a), for which a complete parameterization set has been derived by Mackerell and co-workers, as reported in reference [8]. (b) represents a general iminosugar structure.

6.3 Results and Discussion

MD and QM/MM simulation procedures for NJ and DNJ are reported in **Chapter 5** (*Simulation Protocols*). As for the other six hexacyclic systems (cyclohexanol, 1,2- and 1,4-cyclohexanediol, myo-inositol, β-D-xylopyranose and β-D-glucopyranose), the hydration analysis for DNJ and NJ started from the calculation of "static" water in their first coordination shell, by integrating under the first peak of their respective PDF. These results are reported in the fourth columns of **Tables 6.1** and **6.2**, for NJ and DNJ, respectively.

Further, by calculating the residence time at the *cutoff* value, the hydration numbers results have been obtained, by applying equation **3.18**.

Because NJ and DNJ differ from each other for one OH group in position 1 (**Figure 6.2**), one expects an increment in the hydration number for NJ, due to its additional hydrophilic group. This is in agreement with calculations: NJ is characterised by $n_H = 30.0$ versus 28.5 for DNJ.

Further, a comparison between glucopyranose and NJ shows that the O-sugar has the highest hydration number between the two, namely 32.8 water molecules in its first coordination shell (versus 30.0 for NJ). These values show that the presence of the nitrogen in the ring is affecting the hydration number. On the other hand, this effect is not dramatic which suggests that hydration numbers alone are not the best tracers of the mechanisms responsible for the inhibition activity of NJ.

TABLE 6.1 MD Residence Time (RT) and n_H results for each atoms for DNJ. Water residence time is 4ps. $t^*=2.0$ps.

deoxynojirimycin	RT_{rcut} [ps]	solute cutoff [Å]	n_{solute} [g(r)$_{min}$]	n_H
C1	3.82	4.4	7.5	2.6
C2	3.33	4.3	6.0	1.8
C3	3.39	4.3	5.8	1.8
C4	3.13	4.4	5.8	1.6
C5	5.83	5.7	10	10
N5	3.39	3.4	2.0	0.6
C6	5.15	5.0	14	6.4
O2	2.88	3.5	3.6	0.9
O3	3.38	3.8	4.7	1.4
O4	2.72	3.7	3.6	0.8
O6	2.20	3.4	3.1	0.5

TABLE 6.2 MD Residence Time (RT) and n_H results for each atoms for NJ.

nojirimycin	RT_{rcut} [ps]	solute cutoff [Å]	n_{solute} [g(r)$_{min}$]	n_H
C1	4.27	4.4	7.1	2.9
C2	3.62	4.3	5.8	2.0
C3	3.99	4.3	6.5	2.1
C4	3.51	4.4	4.9	1.8
C5	5.38	5.4	16	7.8
N5	4.75	3.4	2.4	0.9
C6	5.96	5.0	15	7.1
O1	4.84	4.3	8.0	3.5
O2	3.28	3.5	4.0	1.1
O3	3.34	3.8	3.8	1.4
O4	3.00	3.7	3.7	0.9
O6	2.40	3.4	3.2	0.6

In order to explain in more details the different hydration results of these compounds, ring puckering analysis – and related relaxation times for the θ and ϕ angles - has been performed, which revealed a peculiar behaviour regarding NJ molecule.

Figure 6.5 reports the contour plots for the puckering results - from QM/MM trajectory - for DNJ (a) and NJ (b). Both of them adopt the expected $^4C_1$ chair conformation, as they both show $\theta \approx 0°$ and $\phi \in [0,360]$ degrees, according to the Cremer and Pople notation for a $^4C_1$ chair (**Chapter 3, Section 3.4**).
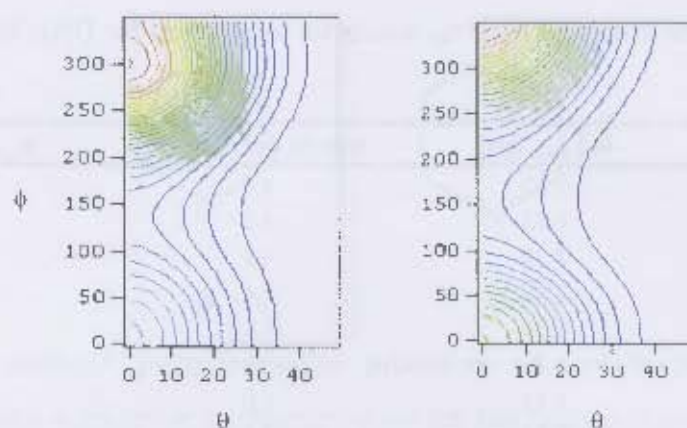
113

Figure 6.5 Contour plots for the puckering results for DNJ, left panel, and NJ, right panel. According to the Cremer and Pople notation, both DNJ and NJ adopt the $^4C_1$ chair conformation ($\theta \approx 0, \phi \in [0,360]$).

On the other side, by observing the MD trajectory for NJ, it can be noted that after about 3.5 ns of simulation, its conformation starts to change: it goes through a boat conformation, namely a $^{1,4}B$, to eventually going back to the $^4C_1$ conformation. These changes are reported in **Figure 6.6**, where the different conformations adopted by the ring during the simulation can be identified by using the *PaperChain* visualisation algorithm, freely available in Visual Molecular Dynamic (VMD)[3] software. *PaperChain* highlights the ring of the molecule structure with a polygon, which is coloured according to the ring pucker [12]. In **Figure 6.7** is represented the colouring notation for *PaperChain* (only for the puckering conformations of interest), as reported in reference [12].

These puckering analyses have been coupled with the relaxation times for θ and ϕ angles, reported in **Table 6.4**. From these results, some conclusions can be made. Although both molecules are found in the expected $^4C_1$ chair conformation (**Figure 6.5**), however DNJ remains fixed in that conformation for a longer time (τ_θ=3.8ps and t_ϕ=6.9ps) in respect to NJ (τ_θ=3.4ps and t_ϕ=5.4ps). The latter is more flexible and consequently a temporary transition of conformation is possible. Further, if comparing the two iminosugars with glucopyranose's relaxation times (τ_θ=6.7 and t_ϕ=6.8), one can observe that the O-sugar is more rigid in its conformation. The relatively high flexibility of NJ may be responsible for observing NJ in a boat conformation during its MD simulation.

---

[3] VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign

The n_{H} results for NJ and glucopyranose discussed above can have now a clearer explanation in light of puckering analysis and TCF for θ and φ angles. The presence of the nitrogen in position 6 gives more flexibility to the molecule with respect to having the oxygen in the same position (for glucopyranose). NJ spends less time stacked in a specific conformation, thus having less time to form H-bonding with the surrounding water. Consequently, although the two molecules (nojirimycin and glucopyranose) have equal number of functional groups of the same nature and in the same position of the ring, they end up in a different number of water molecules in their respective first coordination shell.
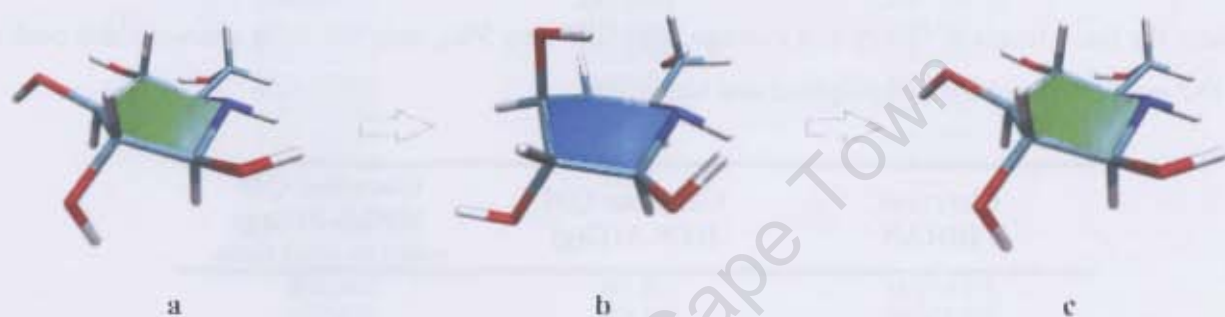

a   b   c

Figure 6.6 The different conformations adopted by NJ, $^{4}C_{1}$ (a), $^{1,4}B$ (b) and back to $^{4}C_{1}$ (c), highlighted by using *PaperChain* algorithm. The colouring notation is reported in **Figure 6.7**.

TABLE 6.4 Relaxation times for θ and φ angles, for NJ and DNJ compared with glucopyranose.

molecules	τ_{θ} [ps]	τ_{ϕ} [ps]
nojirimycin	3.4	5.4
deoxynojirimycin	3.8	6.9
glucopyranose	6.7	6.8

Therefore, from these preliminary results, it follows that the presence of the nitrogen in the ring gives a higher flexibility to NJ and DNJ structures, if compared with glucopyranose. This could realistically represent one of the reasons for iminosugars to act as inhibitors towards specific substrates.

Further studies of similar compounds (ideally, of the complete series of iminosugars reported in **Figure 6.3**), and in particular the investigation of the influence of pH - and consequent protonation of the heteroatom - will reveal more insight into this class of compound and hopefully it will give important information on how it is possible to enhance their activity.

Conformer	Cremer-Pople Coordinates			Original Carbohydra Colours			Colour	Revised Carbohydra Colours			Colour
	φ	θ	Q	R	G	B		R	G	B	
$^1C_4$	0.360	180	0.57	0.0	0.57	0.0		0.0	1.0	0.13	
$^4C_1$	0.360	0	0.57	0.0	0.57	0.0		0.0	1.0	0.13	
^{14}B	240	90	0.76	0.76	0.0	0.0		0.0	0.0	1.0	
B_{14}	60	90	0.76	0.76	0.0	0.0		0.0	0.0	1.0	

Figure 6.7 The colouring method for some puckering conformations, as reported in reference 11 [12].The colouring notation used in *PaperChain* algorithm follows the general Cremer and Pople method for ring puckering.

TABLE 6.5. Comparison between IR frequencies values for DNJ, from CHARMM, HF/6-31G(g) and MP2/6-31G(d) [8]. Although all the frequencies are reported for giving a complete comparison between the three levels of theory (on average, they differ by 5%), only the most characteristic peaks for DNJ molecule have been highlighted and identified.

Charmm VIBRAN	Gaussian QM HF/6-31G(g)	Gaussian QM MP2/6-31G(g) scaled by 0.943 factor
351.818745	328.1981	336.2688
379.271507	384.4295	368.5194
427.728978	421.3781	379.7208
459.209030	424.1779	422.5993
502.986301	471.8699	434.5541
529.708277	532.1044	483.0463
543.507123	567.9952	541.5073
588.692857	600.6829	568.1463
614.275496	626.2801	568.6338
649.757076	660.3339	652.5663
666.577032	726.9950	699.9183
792.199335	784.9375	761.4949
822.627586	888.0132	800.6243
861.102765	938.2188	844.5119
922.992211	945.8173	849.8076
969.759461	987.8406	889.7689
993.192171	**1042.6522**	930.2965
1035.374213	1077.0284	957.107
1048.079514	**1114.4578**	977.3423
1100.766340	1173.6628	**1027.543**
1112.499556	1196.9957	1052.524
1121.054005	1212.0300	1066.036
1127.848776	1227.3885	1075.585
1153.613548	1229.2591	1081.852
1167.680475	1242.8617	1089.75
1222.568298	1295.2499	**1142.062**
1263.195113	1311.4281	1146.331
1287.321501	1333.1091	1180.448
1291.991709	1355.7671	1201.52
1313.460015	1378.4782	1219.202
1335.078362	**1401.3805**	1236.066
1343.005936	1429.4823	1258.789
1356.608855	1473.6660	1295.211

1388.615335	1482.7954	1304.587
1397.891629	1501.6498	1320.922
1416.151788	1518.1223	1333.374
1436.212714	1533.6942	1338.607
1475.806558	1563.1330	1364.445
1503.687219	1572.4285	1393.615
1533.055132	1580.6724	1395.541
1597.909389	1600.6776	**1407.727**
1627.622639	1612.2264	1423.266
1647.112512	1647.3163	1431.312
1653.135290	**1669.7577**	1480.863
1697.916224	1677.7486	1487.569
2898.884606	3187.9514	2885.844
2907.038644	3225.1492	2913.964
2930.010989	3227.2830	2927.599
2931.564993	3236.6157	2935.407
2934.271242	3253.8423	2938.369
2937.053331	3265.6532	2954.248
2954.284285	3269.2606	2967.104
2964.212986	**3278.4761**	2988.052
3361.781219	3782.8786	3340.000
3370.973665	4049.9631	**3359.478**
3381.238394	4053.0974	3441.236
3387.608128	4089.7721	3462.702
3396.365303	4112.7264	3552.603

References

1. Asano, N., *Sugar-mimicking glycosidase inhibitors: bioactivity and application.* Cellular and Molecular Life Sciences, 2009. **66**: p. 1479-1492.

2. Kuriyama, C., et al., *In vitro inhibition of glycogen-degrading enzymes and glycosidases by six-membered sugar mimics and their evaluation in cell cultures.* Bioorganic and Medicinal Chemistry, 2008. **16**: p. 7330-7336.

3. Asano, N., et al., *Nitrogen-in-the-Ring Pyranoses and Furanoses: Structural Basis of Inhibition of Mammalian Glycosidases.* Journal of Medical Chemistry, 1994. **37**: p. 3701-3706.

4. Reese, E.T., F.W. Parrish, and M. Ettlinger, *Nojirimycin and D-glucono-1,5-lactone as Inhibitors of Carbohydrases.* Carbohydrate Research, 1971. **18**: p. 381-388.

5. Asano, N., *Glycosidase Inhibitors:Update and Perspectives on Practical Use.* Glycobiology, 2003. **13**(10): p. 93R-104R.

6. Kodama, Y., et al., *Molecular Structure and Glycosidase-Inhibitory Activity of Nojirimycin Bisulfite Adduct.* Journal of Antibiotics, 1985. **38**(1).

7. Hempel, A., et al., *Glucosidase Inhibitors: Structures of Deoxynojirimycin and Castanospermine.* Journal of Medical Chemistry, 1993. **36**: p. 4082-4086.

8. Vanommeslaeghe, K., et al., *CHARMM General Force Field:A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields* Journal of Computational Chemistry, 2010. **31**(671-690).

9. Thompson, M.A., *ArgusLab 4.0,* http://www.ArgusLab.com, Editor, Planaria Software LLC: Seattle.

10. Frisch, M.J.T., G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A., *Gaussian 03, Revision C.02.* 2004, Gaussian, Inc.: Wallingford, CT.

11. Islam, B., et al., *Novel Anti-Adherence Activity of Mulberry Leaves: Inhibition of Streptococcus mutans Biofilm by 1-deoxynojirimycin Isolted from Morus alba.* Journal of Antimicrobial Chemotherapy, 2008. **62**: p. 751-757.

12. Cross, S., et al., *Visualisation of cyclic and multi-branched nolecules with VMD.* Journal of Molecular Graphics and Modelling, 2009. **28**: p. 131-139.