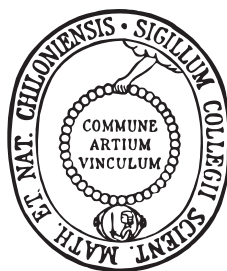


Molecular Dynamics Studies of the Glycocalyx: carbohydrates in physiological water, and their interactions with FimH

Doctoral Thesis



submitted in partial fulfillment
of the requirements for the doctoral degree
" Dr. rer. nat."
to the Faculty of Mathematics and Natural Science
of the Christian-Albrechts-University of Kiel

Submitted by
Mats Eriksson
Kiel 2009

Referent: Prof. Dr. Bernd Hartke

Koreferent: Prof. Dr. Thisbe K. Lindhorst

Tag der mündlichen Prüfung: 03-11-2009

Zum Druck genehmigt: 03-11-2009

Der Dekan:

Kurzzusammenfassung

Die molekulare Bindung verschiedener chemischer Substanzen an Kohlenhydrat-Moleküle bildet die Grundlage für den Austausch von Informationen zwischen Zellen und ist an der Kontrolle einer Vielzahl von biologischen Prozessen, wie beispielsweise der Immunabwehr, der Befruchtung oder der Abwehr von Krebszellen beteiligt. Das genaue Verständnis und die Möglichkeiten zur Beeinflussung dieser Reaktionen sind von großem Interesse für die Entwicklung zukünftiger pharmazeutischer und medizinischer Anwendungen. Mittels klassischer Moleküldynamik und unter Verwendung speziell implementierter Analyse-Methoden wurden verschiedene auf eukaryotischen Zelloberflächen vorkommende Saccharid-Systeme in unterschiedlichen Lösungsmittel-Umgebungen simuliert. Hierbei wurden deren Reaktionen in einer wässrigen, in einer hoch-ionischen und in einer kohlenhydrat-reichen Umgebung bezüglich einiger ungeklärten Fragestellungen untersucht. Es konnte gezeigt werden, dass für Wassermoleküle eine Position in der Nähe zu Kohlenhydrat-Molekülen im Vergleich zu einer rein wässrigen Umgebung thermodynamisch ungünstig ist, und dass deren Freisetzung eine Triebkraft zur Ausbildung von Kohlenhydrat-Kohlenhydrat und Kohlenhydrat-Protein Wechselwirkungen ist. Des Weiteren wurden die spezifischen Wechselwirkungen zwischen Kohlenhydrat-Molekülen und einer Reihe von biologisch relevanten Kationen detailliert untersucht, wodurch eine starke Korrelation zwischen den Solvationseigenschaften der Kationen und ihren Bindungseigenschaften zu Sacchariden gezeigt werden konnte. Die erhaltenen Ergebnisse haben eine Bedeutung für das allgemeine Verständnis ionenspezifischer Phänomene in einer Vielzahl von biologischen Systemen. Im letzten Kapitel der Arbeit werden die strukturellen Zusammenhänge für die mannosenspezifische Bindung des Lektins FimH analysiert. Hierbei wurden die für die Erkennung mannose-haltiger Saccharide wichtigen Aminosäuren in FimH identifiziert, und die Anwendungen dieser Entdeckungen für die Entwicklung von Anti-Adhäsionsmedikamenten gegen von *E. coli* Typ-1-Fimbrien-vermittelte Infektionen des Harn-Traktes diskutiert.

Abstract

Molecular binding events involving carbohydrates are responsible for information exchange between cells and control numerous biological functions, including immune responses, fertilization and defense against cancer. Understanding and controlling these binding events is of great interest for the development of future pharmaceutical and medical applications. Using classical molecular dynamics simulations of different saccharide systems in different media, together with a series of specially implemented analysis tools, we address several open questions about the chemistry behind molecular interactions taking place in the aqueous, highly ionic and carbohydrate-rich environment of eukaryotic cell surfaces. It is shown that water in the vicinity of carbohydrates is thermodynamically unfavorable compared to bulk water, and that the release of carbohydrate hydration water can provide a driving force to carbohydrate-carbohydrate and carbohydrate-protein molecular association processes. The occurrence of ion-specific interactions between saccharides and a series of biologically relevant cations is further investigated. A clear correlation between hydration properties of the cations and their respective binding affinity to saccharides can be seen. These results have implications for understanding general ion-specific phenomena taking place in all biological systems. The structural basis of mannose-specific binding of the lectin FimH is elucidated in the last section of this thesis. Key residues in FimH, responsible for the recognition of mannose-containing saccharides, are identified, and the application of these findings to the design of anti-adhesion drugs against type 1 fimbriated *E. coli* uropathogenic infection is discussed.

Contents

1	Introduction	1
2	Techniques and Methods	5
2.1	The Molecular dynamics method	5
2.1.1	Molecular mechanics and force fields	6
2.1.2	Equations of motion	8
2.1.3	Simulation methods	10
2.1.4	Measurement of macroscopical system properties	14
2.1.5	Basic equilibration and simulation protocol	16
2.2	Analysis methods	18
2.2.1	Distribution functions	18
2.2.2	Hydrogen bond analysis	20
2.2.3	Time correlation functions	20
2.2.4	Voronoi tessellation	24
2.2.5	Water order parameter	26
2.2.6	MM-PB(GB)SA	27
2.2.7	Methods for calculating biomolecular entropies	33
2.2.8	Two-phase model entropy calculations	42
3	Hydration studies of carbohydrates	47
3.1	Introduction	47
3.2	Water models	49
3.3	Systems and simulation protocols	53
3.4	Hydration of the N-linked high mannose-type oligosaccharide	54
3.4.1	Structural properties of saccharide hydration water	54
3.4.2	Dynamic properties of saccharide hydration water	61
3.4.3	Thermodynamics of saccharide hydration water	66

3.5	Dependence of hydration properties on carbohydrate density	74
3.6	Saccharide hydration with SPC/E water	81
3.7	Discussion	87
4	Carbohydrate-ion interactions	89
4.1	System and simulation protocol	91
4.2	Ion hydration	93
4.3	Saccharide ion affinities	99
4.3.1	Man ₆ GlcNAc ₂ -Asn-cation interaction	99
4.3.2	Dependence of ion affinities on saccharide density	103
4.4	Ion specific effects on saccharide conformations	108
4.5	Discussion	111
5	FimH-carbohydrate interactions	113
5.1	Introduction	113
5.2	System and simulation protocol	114
5.3	Binding free energy calculations	116
5.4	Structure of the FimH CRD	119
5.5	FimH-Man α Me complex	121
5.5.1	Hydrogen bond analysis	121
5.5.2	Binding free energy analysis	124
5.5.3	The role of structural water molecules	130
5.5.4	Per-residue free energy decomposition	132
5.6	Fimh-Man α PNP complex	135
5.6.1	Force field development	135
5.6.2	Hydrogen bond analysis	137
5.6.3	Binding free energy analysis	138
5.6.4	Per-residue free energy decomposition	141
5.7	FimH-Man ₂ GlcNAc complex	144
5.7.1	Hydrogen bond analysis	144
5.7.2	Binding free energy analysis	146
5.7.3	Conformation of Man ₂ GlcNAc in the complex	149
5.7.4	Per-residue free energy decomposition	151
5.8	FimH-Man ₃ GlcNAc ₂ complex	153
5.8.1	Hydrogen bond analysis	154
5.8.2	Binding free energy analysis	155

5.8.3	Per-residue free energy decomposition	159
5.8.4	Conformation of Man ₃ GlcNAc ₂ in the complex	160
5.9	Discussion	163
	Bibliography	167
	Declaration	182
	Acknowledgments	184
	Curriculum Vitae	186

1 Introduction

During the last decades carbohydrates have been recognized to play an important role in cell interaction, giving rise to the field of glycobiology, i.e. the study of the role of carbohydrates in biological events. Carbohydrates are by now, in addition to the three major classes of biological macromolecules, proteins, nucleic acids and lipids, recognized as a fourth group of macromolecules controlling the chemistry of life. The surface of all eukaryotic cells is covered by a layer of complex carbohydrates called the glycocalyx. This macromolecular system is structurally very complex and its detailed structure differs with the type of the organism, the cell type and even changes during the life cycle of the cell [1]. Within this structural diversity, there are, however, characteristic glycosylation patterns common to all eucaryotic cells. In the direct vicinity of the cell surface exists a layer of saccharides covalently bonded, either directly to the membrane lipids (glycolipids) or to membrane proteins (glycoproteins). The membrane bound oligosaccharide moieties are typically short and highly branched oligosaccharides of high structural diversity. Additional glycoconjugates, such as proteoglycans, are attached to this initial layer of saccharides, giving rise to a thick carbohydrate coating covering the cell surface. It is by now known that this nano-dimensioned layer of glycoconjugates is involved in a myriad of important functions for the cell that extend beyond serving as a simple buffering layer, protecting the cell membrane against mechanical and chemical stress. An important function of cell surface carbohydrates is to store and mediate information exchange between different cells. Variations in the structure of the glycocalyx in different cells are used by the immune system to recognize and destroy invading organisms [2]. Changes in the structure of cell surface carbohydrates of cancerous cells can, in a similar manner, trigger the appropriate immune response. In these processes, carbohydrate-recognizing proteins (lectins) or glycoproteins (selectins) are responsible for the recognition and binding of specific saccharide units, or arrays of saccharides in the glycocalyx. Using the same mechanism, bacteria utilize extra-

cellular lectins to recognize and bind specific saccharides on the host cell surface, in the first stage of the colonization process [3]. Also pure carbohydrate-carbohydrate interactions have been found to play an important role in some cell adhesion processes [4]. There is an obvious pharmaceutical interest to understand the mechanisms behind these processes. Carbohydrate-based drugs can be designed to effectively block the carbohydrate recognition domains of lectins, thereby blocking the ability of the bacterium to colonize the host cell. Such drugs may be an alternative to traditional antibiotic treatments, with the attractive feature that any mutations causing a loss of the affinity for the drug also should result in loss of binding affinity to the natural receptor [5]. Drugs mimicking the carbohydrate recognition of lectins can further be used as drug-carriers to direct a therapeutic agent to its desired target [6]. However, to this date, the biochemistry and biophysics of cell surface saccharide recognition and binding is not completely understood [1]. Compared to protein-protein recognition processes, protein-carbohydrate and carbohydrate-carbohydrate mediated adhesion processes are much more complex, since these typically involve not a single binding domain, but rather interactions with an extended part of the cell surface saccharides. In this manner, a high binding affinity can be achieved although individual carbohydrate-protein and carbohydrate-carbohydrate interactions are reactively weak [7, 8]. Thus, understanding of molecular association processes taking place with the saccharides of the glycocalyx necessitates understanding of single saccharide interactions, as well as, the cooperative effect of the the saccharide array that covers the cell surface.

Although glycobiology is a relatively young discipline, both structure and function of the glycocalyx, and that of its oligo- and polysaccharide building blocks, have started to be unraveled through a combination of experimental studies and computer simulations in the last decades. The structural basis of lectin-carbohydrate recognition has been elucidated through X-ray and NMR spectroscopy experiments, determining both the structure of different lectins and that of lectins in complex with their saccharide ligands [9–12]. Microcalorimetry and surface plasmon resonance measurements have revealed binding affinities and thermodynamics of lectin-carbohydrate interactions [13, 14]. The role of multivalency in ligand-carbohydrate recognition has further been investigated through development of synthetic polyvalent ligands mimicking glycocalyx-like surfaces [15–17] and self-assembled monolayers representing full two-dimensional saccharide surfaces [18], as well as their binding affinities to lectins [19].

Since the late 1970s, computer studies have been used to investigate structure and dynamics of carbohydrates. Ab initio and semiempirical methods have been used to

study internal conformational behavior of monosaccharides, such as preferred ring and anomeric conformers [20, 21]. Early molecular mechanics studies focused on conformational behavior of disaccharides using Ramachandran plots of the dihedral angles of the glycosidic linkage [22]. These studies were performed on mono- or disaccharides in the gas phase, without accounting for the natural aqueous environment of these molecules. With increasing computational resources, explicit solvent molecules could be included in the simulations. In 1990 Edge *et al.* [23] performed a 0.5 ns long molecular dynamics study of a disaccharide with explicit water. From this study it was shown that molecular fluctuations were dampened significantly compared to gas phase simulations and that accounting for carbohydrate-water interaction is important for modeling saccharide dynamics. Since then, several important contributions have been made to the field of computer simulations of carbohydrates. In 1995 Woods *et al.* [24] presented the GLYCAM93 force field, allowing for simulations of larger oligosaccharides and glycoproteins. The constant increase in computational power has opened a whole new field for computational studies of carbohydrate systems, such as hydration studies of carbohydrates [25–27], conformational dynamics of larger saccharides [28, 29] and lectin-carbohydrate interactions [30–32]. Still, modeling of saccharides is a challenging task for computer scientists. Compared to protein and DNA, larger saccharides are very flexible and do typically not have a fixed three-dimensional structure. From a modeling point of view, this translates to a need of very long simulation times to model conformational behavior of saccharides [33]. The structural diversity of larger saccharides, i.e. the large number of building blocks and possible linkages between them, poses a challenge for both force field developers and scientist trying to understand the structure and function of the glycocalyx. Nevertheless, computational studies of macromolecular biological systems hold great potential as a valuable tool for understanding physical properties of biomolecules and their interactions *in vivo*. With the development of both interaction potentials with increasing accuracy and efficient simulation techniques, computer simulations provide an excellent complement to traditional experimental studies. In many cases, computational studies can provide detailed information on an atomic scale about molecular interactions, not accessible from experiment. In this manner, computer simulations can be used to explain, predict and even provoke new experimental studies.

Following this idea, the work in this thesis aims at describing molecular interactions, involving carbohydrates that, in many cases, cannot be measured by experiment. Using molecular dynamics simulations, together with force fields developed specially for

modeling conformational behavior of saccharides in the condensed phase, several different types of molecular interactions are investigated, with the overall goal of providing a greater understanding of specific interactions occurring with the saccharides of the glycocalyx. In Chapter 3, a comprehensive study of the interaction of water with different carbohydrate systems, ranging from monosaccharides to glycocalyx models, is presented. This study is of general interest for all molecular association processes taking place in the aqueous environment of extracellular carbohydrates. Ion-specific effects, taking place with carbohydrates in an ionic aqueous environment, are investigated in Chapter 4. Here, interactions between a selected set of biologically relevant cations and saccharides found in the glycocalyx are studied in detail. Finally, in Chapter 5, the structure/thermodynamic relationship between a series of carbohydrate ligands and the lectin FimH is studied. The theory behind the molecular dynamics method and that of other computational tools used in this work is presented in the following chapter.

2 Techniques and Methods

2.1 The Molecular dynamics method

Simulations of biomolecules, or biomolecular systems, typically require modeling of large-scale systems. The number of atoms of a biomolecule alone, such as a protein or a biological membrane, is typically in the order of thousands or tens of thousands of atoms. When including explicit water in the simulation, the size of the system increases drastically. Modeling of such large systems requires a computationally feasible method that still is able to capture the physics of interactions between atoms in the system. Classical molecular dynamics (MD) simulations have become an important technique in computational chemistry for modeling large systems, which, due to size restrictions, cannot be treated at higher levels of theory. The constant increase in computer power together with the development of well parallelized molecular dynamics programs, such as AMBER [34], Gromacs [35], NAMD [36] and CHARMM [37], has led to a great increase in both system sizes and simulation times available for computational studies. The first MD simulation, done over 30 years ago, was made on a system of 500 atoms for a time interval of 9.2 picoseconds [38]. Today systems sizes as large as one million atoms can be simulated for several nanoseconds [39]. This has opened the field for studying biological processes on relevant time-scales and made MD simulations an excellent complement to experimental studies. By following the time evolution of a molecular system over time, time-dependent processes can directly be followed in the simulation. Furthermore, MD-simulations can be used as a tool to sample the phase space of the system and provide means to understand and predict macroscopic phenomena. The basic theory of molecular dynamics simulations is presented in the following subsections of this chapter. The theory behind a series of post-processing tools used to analyze the simulation results is presented in Section 2.2.

2.1.1 Molecular mechanics and force fields

The foundation of classical molecular dynamics simulations is the representation of the potential energy hypersurface in a functional form of the nuclear coordinates of all atoms in the system. Using the Born-Oppenheimer approximation, motions of the electrons and the nuclei can be treated separately. In force field methods, electrons are treated implicitly by a single (usually ground state) potential energy surface, on which the nuclei move around as classical particles obeying Newton's equations of motion. In order to allow modeling of large systems, the potential energy surface is calculated from a simple parametric function of the nuclear coordinates, with fitted empirical parameters depending on the atom type and its surroundings. This potential energy function, i.e. the parametric function and the fitted parameters, is called the force field. The form of the parametric function is arbitrary and can differ in different force fields. The total potential, in the force fields used in this work, is calculated as a sum over potential terms of bonded and non-bonded interactions,

$$U_{total} = U_{bond} + U_{angle} + U_{torsion} + U_{vdW} + U_{Coulomb}. \quad (2.1)$$

The first three terms are potential functions for bond distances, bond angles and torsional angles of covalently bonded atoms in a molecule,

$$U_{bond} = \sum_{bonds,i} k_i^{bond} (r_i - r_{0,i})^2, \quad (2.2)$$

$$U_{angle} = \sum_{angles,i} k_i^{angle} (\theta_i - \theta_{0,i})^2, \quad (2.3)$$

$$U_{torsion} = \frac{1}{2} \sum_{dihedrals,i} k_i^{dihedral} (1 + \cos(n_i \phi_i - \gamma_i)) \quad (2.4)$$

The bond stretching and bond angle bending motions are modeled as simple harmonic oscillators around some equilibrium bond length $r_{0,i}$ or bond angle $\theta_{0,i}$. The harmonic approximation is sufficient for modeling conformational behavior at ambient tempera-

tures [40], but can clearly not be used to model bond breaking processes. Thus, atom connectivities are typically only determined once and remain constant for the duration of the simulation. The torsional potential needs to account for the fact that the energy profile of rotations around a torsional angle may have several minima, separated by barriers of different heights. Utilizing the fact that the potential for rotations around a torsional angle needs to be periodic, i.e. $U_{torsion}(\psi=0^\circ)=U_{torsion}(\psi=360^\circ)$, the potential function of the torsion angle ψ is expanded in a Fourier series. Each bonded series of four atoms have at least one set of torsional parameters, and more complex potentials can be deconvoluted into several Fourier series terms.

Interactions between atoms separated by more than three atoms in a molecule, or interactions between different molecules, are treated by the non-bonded interaction potentials U_{vdW} and $U_{Coulomb}$ in Eq. 2.1:

$$U_{vdW} = \sum_i \sum_{j>i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.5)$$

$$U_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.6)$$

U_{vdW} is the van der Waals energy. This potential describes attractive interactions between non-bonded atoms due to correlations of instantaneous fluctuations in the electron density of neighboring atoms, as well as the repulsive forces acting on atoms close enough for electron densities to overlap. The van der Waals interaction between two non-bonded atoms i and j at distance r_{ij} is commonly modeled by the pairwise Lennard-Jones potential in Eq. 2.5, where ϵ_{ij} and σ_{ij} are parameters determining the minimum energy distance and depth of the minimum respectively, calculated from program specific combination rules using ϵ and σ parameters optimized for each atom type. The last force field term, $U_{Coulomb}$, models inter- or intramolecular interactions between permanent molecular dipoles. A common approximation in molecular mechanics force fields is to neglect polarization and treat charge distributions as fixed point charges centered on the nuclei of the atoms. The interaction energy between charged non-bonded atoms can then be calculated by the effective pairwise Coulomb potential in Eq. 2.6. The biomolecular force fields used in this work are all, so called, second generation force fields, that model atomic interactions by an effective two-body additive potential

with atomic point charges determined by quantum mechanical calculations. The main limitation of these force fields is the neglect of multipole moments, fluctuating charges and polarization in the point charge approximation [40]. Long-ranged Coulomb interactions dominates the overall energy in polar systems, and therefore require accurate treatment. Although next generation force fields, including more realistic descriptions of electron density, are under development, the potential function given by Eq. 2.2-2.6 is still mainly used in common biomolecular molecular dynamics programs.

Within the limits enforced by the form of the potential function, the accuracy of the force field is fully controlled by the undetermined parameters in Equations 2.2-2.6. Parameterization of the force field is typically done differently for the bonded and non-bonded potential functions and involves optimization with respect to both experiment and quantum mechanical calculations. Modeling biomolecules in the condensed phase requires correct description of, not only solute-solute interactions, but also solute-solvent interactions. The biomolecular force fields used in this study are based on the Cornell *et al.* force field [41] parameterized especially for the condensed phase. Second generation biomolecular force fields of this type can reproduce experimental free energies of solvation with deviations around 1 kcal/mol, which is within the accuracy of the experimental measurements [42]. The methods used to parameterize the force fields used in this study typically differ for each force field and are not discussed further here. For more information, the reader is referred to the original papers of the respective force field [24, 43–45].

2.1.2 Equations of motion

Given the force field potential energy function, the coordinates and velocities of a system of N classical particles at a given time can be obtained by solving Newton's equations of motion,

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial}{\partial \mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N), i = 1, 2, \dots, N, \quad (2.7)$$

where m_i is the mass of atom i , \mathbf{r}_i its position and U is the potential energy that depends on all N atoms in the system. Since no analytical solution exists for the coupled differential equations in Eq. 2.7 for system containing more than two atoms, numerical

algorithms are needed to model systems of realistic sizes. The Verlet algorithm is one, out of many methods, for integrating the equations of motions in Eq. 2.7 to obtain the path followed by each atom in the system through space. Using the Taylor expansion of the coordinate of an atom, around the time t ,

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{f}(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}\ddot{\mathbf{r}} + O\Delta t^4, \quad (2.8)$$

and similarly,

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{\mathbf{f}(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}\ddot{\mathbf{r}} + O\Delta t^4. \quad (2.9)$$

Combining the two equations above gives the formula for calculating the position of the atom at a later time step Δt with an error in the order of Δt^4 :

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{f}(t)}{m}\Delta t^2 \quad (2.10)$$

In the first time step, the $\mathbf{r}(t - \Delta t)$ term is not known and an alternative formula has to be used:

$$\mathbf{r}(\Delta t) \approx \mathbf{r}(0) + \mathbf{v}(0)\Delta t + \frac{\mathbf{f}(0)}{2m}\Delta t^2 \quad (2.11)$$

The velocities are not explicitly needed for calculating the new positions, but are often needed in simulations to calculate physical properties of the system, such as kinetic energy and temperature. By subtracting Eq. 2.8 and 2.9 the velocity of the atom can be calculated as:

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t}, \quad (2.12)$$

Thus, given initial coordinates and velocities, the time evolution of the system can directly be calculated using the above formulas. Due to its simplicity, speed and good energy conservation properties, the Verlet algorithm, or the related velocity-Verlet algorithm, is used as standard propagator in molecular dynamics programs, such as AMBER and NAMD.

2.1.3 Simulation methods

Treatment of non-bonded interactions

The computationally most expensive term in the force field equation is the calculation of non-bonded interactions. In principle, non-bonded interactions need to be calculated between a given atom and all other atoms in the simulation box. The computational cost of such a calculation has order of N^2 time complexity, where N is the number of atoms in the simulation box. A full evaluation of non-bonded interactions of all atoms in the system would put strong limitations on system sizes available for molecular dynamics studies. One way to circumvent this problem is to introduce cutoffs for non-bonded interactions. Figure 2.1 shows an example simulation box in two dimension with a few of its periodic images. Using cutoffs, Coulomb and van der Waals interactions are only calculated for neighboring atoms within the circle in Figure 2.1, with a radius determined by the cutoff distance. For van der Waals interactions, which decay rapidly with distance, this is a reasonable approximation, however for long ranged Coulomb interactions, truncation can result in poor modeling of charged and polar systems [46, 47]. A typical problem with simulations using cutoffs for Coulomb interactions is poor energy conservation, due to the neglect of long range forces in the system. In fact, for typical system sizes used in molecular simulations, Coulomb forces from atoms beyond the borders of the simulation box can be important to ensure good energy conservation and natural trajectories. Ewald summation [48] can be applied to computer simulations, providing an efficient method for calculating long range contributions to the potential energy of a periodic system. The periodic representation in Figure 2.1 is, for a real three dimensional system reproduced, infinitely in each dimension. For such a system, the Coulomb contribution to the potential energy from interaction with atom i , with all atoms in the original simulation box and all periodic copies can be written as:

$$U_{Coulomb} = \frac{1}{2} \sum_{\mathbf{n}} \left(\sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{4\pi\epsilon_0} |\mathbf{r}_{ij} + \mathbf{n}|^{-1} \right), \quad (2.13)$$

where $\mathbf{n} = (n_x L_x, n_y L_y, n_z L_z)$. Here (n_x, n_y, n_z) is a set of integers and (L_x, L_y, L_z) are the dimensions of the simulation box.

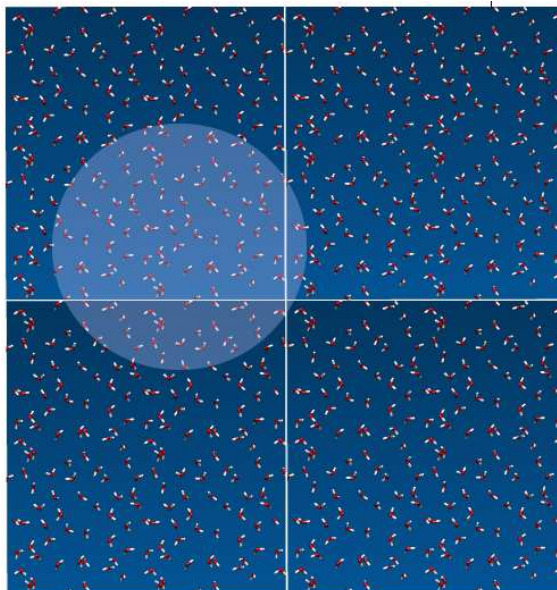


Figure 2.1: Periodic boundary conditions in two dimensions of a an example pure water system.

The prime of the first summation indicates that when $\mathbf{n} = \mathbf{0}$, then $i=j$ is omitted in the summation over the atom indices. It can be shown that Eq. 2.13 is, due to the long range of Coulomb interactions, a poorly converging sum and therefore not suitable for direct use in a molecular dynamics program. In the Ewald method, the summation is split up into a short range and a long range part. The short range part, accounting for interactions with atoms within a given cutoff distance, is calculated in real space, whereas the long range part is summed in the Fourier space, where the sum is converging rapidly. With modern grid based methods, such as PME [49] utilizing the Fast Fourier Transform algorithm, the computer time scales as $O(N \log N)$ with respect to the number charges in the system. Since Ewald summation introduces correlation between an atom and its periodic images, it is necessary to have a large enough simulation box, especially for charged systems. The cutoff used to separate the real space and Fourier space parts can not be larger than half the length of the smallest box dimension since a given atom will interact with itself when $\mathbf{n} = \mathbf{0}$ (see Figure 2.1). The accuracy of the PME calculation can be controlled by setting the spacing of the grid, onto which the atomic partial charges are mapped in the simulation box. In these simulations, grid spacings of approximately 1 \AA in each dimension have been used. This is a reasonable trade-off between computational cost and accuracy [50].

Thermostats and barostats

The use of methods like PME has increased the energy conservation of molecular simulations, but still small drifts in energy can occur due to numerical inaccuracies, for instance, in the integration algorithm. When attempting long simulations it is therefore necessary to do simulations with a thermostat that ensures that the total energy is conserved in the calculated trajectories. Furthermore, certain studies may require that the simulations are performed in a given ensemble, for which both thermostats and barostats are needed. Here follows a brief description of the methods for temperature and pressure control used in the simulations in this work as implemented in the NAMD molecular dynamics package.

The Berendsen thermostat [51] is a velocity rescaling approach in which atomic velocities, v_i , are rescaled each time as λv_i , where

$$\lambda = 1 + \frac{\Delta t}{2\tau_B} \left(\frac{T_0}{T} - 1 \right), \quad (2.14)$$

with the instantaneous temperature, T , given by [52]:

$$T(t) = \sum_{i=1}^N \frac{m_i v_i^2(t)}{k_B N_f}. \quad (2.15)$$

Here, N_f is the number of degrees of freedom of the system. The instantaneous temperature T is connected to a heat bath with a fixed reference temperature T_0 , so that the rate of change of the temperature is proportional to the difference in temperature of the heat bath and the system. The Berendsen thermostat does not force a system to a given temperature, instead it pushes the temperature towards the desired temperature, still allowing temperature fluctuations characteristic for the canonical ensemble. The coupling to the heat bath can be controlled by the time constant τ_B . Typical values of τ_B are in the range of 0.1-5 ps, where smaller τ_B values gives a tighter coupling to the heat bath and as τ_B approaches infinity, the scaling factor λ is unity and the thermostat has no effect on the system. Thus, temperature fluctuations can effectively be controlled by the value of τ_B . The Berendsen thermostat can easily be included in a given propagation algorithm and can be used together with holonomic restrains

(see below). The main advantage of the Berendsen thermostat is that it does not affect the short time dynamics of the system, and therefore gives trajectories with minimal deviations from NVE trajectories [53]. This becomes important when calculating time-dependent properties such as spectral densities and self-diffusion coefficients. The drawback of the Berendsen thermostat is that it does not generate correct temperature fluctuations for the canonical ensemble [54]. Since the Berendsen thermostat does not ensure that the temperature is evenly distributed in the system, but relies in atomic collisions to remain an evenly distributed temperature in the system, it is often desirable to use another thermostat in the equilibrium phase of a simulation (see Section 2.1.5).

NPT simulations were, in this work, performed using a combination of the Berendsen thermostat and Berendsen pressure bath coupling algorithms. The pressure bath coupling is based on the same principles as the coupling/scaling principle of the Berendsen thermostat. Instead of velocities, dimensions of the simulation box are scaled to push the instantaneous pressure towards the given reference value.

Constrained dynamics and Multiple Time Step methods

The choice of the time step, Δt in Eq. 2.10, is made as a tradeoff between accuracy of the calculated trajectory and the computational cost of the simulation. As a general rule, the time step should not be larger than the fastest motion in the system. For molecular systems, the fastest motions are hydrogen-heavy atom vibrations, putting an upper limit of 1 fs (1×10^{-15} s) for the integration time step. These vibrations are so high in frequency that they are not well treated by classical mechanics. To increase the stability of the simulation and allow for larger integration time steps, it is desirable to remove these degrees of freedom by constraining covalent bonds involving a hydrogen atom. In practice this can be done by applying a set of, so called, holonomic constraints to Newton's second rule in Eq. 2.7. Using the Lagrangian formalism Eq. 2.7 can be written as

$$m_i \ddot{\mathbf{r}}_i = \frac{\partial}{\partial \mathbf{r}_i} U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) + \mathbf{G}_i, \quad i = 1, 2, \dots, N, \quad (2.16)$$

where \mathbf{G}_i is the force constraint given by a set of Lagrangian multipliers involving atom

i. This is the approach used by the SHAKE algorithm [55], used in all simulations in this work. Atomic velocities are in a first step calculated under unconstrained conditions. In a second step, velocities are readjusted in a manner that the applied constraints are fulfilled. This can be done by iteratively solving a linear system of equations [56], or by improved methods [57], which can be incorporated into the Verlet integration scheme with much lower computational cost and without sacrificing accuracy of the method. Simulations where stretching motions of heavy atom-hydrogen are frozen allow the use of a time step as large as 2 fs.

Integration efficiency can further be improved using so called Multiple Time Step approaches [58], in which the force evaluations are done on different time intervals depending on the distance separating two atoms. These algorithms are based on the observation that the change of pair interactions with time, generally is smaller as the distance between two atoms increases. In the NAMD molecular dynamics program, force evaluation is done using three different time steps, depending on the distance separating two atoms in the periodic system. The surrounding of a given atom is divided into three regions. In the first region are atoms directly bonded to the central atom. The second region is atoms within the cutoff distance for van der Waals and real space Coulomb interactions. Long range electrostatic interactions from molecules beyond the cutoff distance are considered in the third region. Using this approach, time steps as large as 4 fs can be used for pairwise interactions with atoms in the third region, still maintaining a good energy conservation in the simulation [36].

2.1.4 Measurement of macroscopical system properties

The result of a molecular dynamics simulation is a collection of velocities and positions of all atoms of the modeled system. This microscopical information can be used directly to calculate time-dependent properties of the system, but can further be extended to calculate macroscopical properties using theory from statistical thermodynamics. This is important for connecting theory with experiment, especially for studying liquids, where microscopic properties are difficult to measure from experiment. Given a force field parameterized to reproduce macroscopic properties, computer simulations can aid the understanding of a given macroscopic phenomena by providing full detail of the system at an atomic scale.

In statistical thermodynamics, the ensemble average of a given observable, $A(\mathbf{r}^N, \mathbf{p}^N)$,

as a function of all possible the positions and momenta of all atoms in the system is given by,

$$\langle A \rangle_E = \int \int d\mathbf{p}^N d\mathbf{r}^N A(\mathbf{r}^N, \mathbf{p}^N) P(\mathbf{r}^N, \mathbf{p}^N), \quad (2.17)$$

where $P(\mathbf{r}^N, \mathbf{p}^N)$ is the probability distribution:

$$P(\mathbf{r}^N, \mathbf{p}^N) = \frac{\exp(-H(\mathbf{r}^N, \mathbf{p}^N)/k_B T)}{\int \int d\mathbf{p}^N d\mathbf{r}^N \exp(-H(\mathbf{r}^N, \mathbf{p}^N)/k_B T)} \quad (2.18)$$

Here, H is the Hamiltonian of the system. The ensemble probability distribution is calculated from a large number of replicas of the system under the same macroscopical state (here, the same number of atoms, temperature and volume). In fact the number of replicas are so large that the every microscopical state corresponding to the macroscopical state is represented in the ensemble. One of the fundamental postulates of statistical thermodynamics is the ergodic hypothesis, stating that the ensemble average in Eq. 2.17 equals the time average, $\langle A \rangle_T$, of the same phase-space property, given that the sampling period is long enough:

$$\langle A \rangle_E = \langle A \rangle_T = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{\tau=0}^{\tau} A(\mathbf{r}^N, \mathbf{p}^N) dt \quad (2.19)$$

This theorem connects ensemble-based statistical thermodynamic theory to experiment and simulations, where sampling of a given property typically is done by successive measurements over a period of time. Given that a simulation is long enough that the system will pass through the relevant parts of the phase space, averages calculated from the simulation are equal to ensemble averages which, in turn, are directly comparable to experimental averages. Thus, simulation time is of uttermost importance when performing molecular dynamics simulations. Problematic, from a modeling point of view, are simulations of systems with a potential energy landscape composed of several local minima, separated by large energy barriers. Averages calculated for such systems may be poor estimations of the true averages, since the system may be trapped in a small subset of local minima for the duration of the simulation. Sampling problems of this kind can be difficult to detect, since calculated averages may still be constant

with increased sampling times and are best treated by special simulation techniques [59, 60]. Insufficient sampling of the phase space can, on top of the already mentioned force field approximations, be another source of error in standard MD-simulations.

Given the partition function (the integral over the whole phase space in Eq. 2.18), equilibrium thermodynamic properties of the system, such as free energies, heat capacities and entropies, can be calculated using standard statistical thermodynamics formulas. A direct evaluation of the partition function of a given ensemble from a simulation is, however, not possible due to the dimensionality of the problem for typical system sizes. In Sections 2.2.6-2.2.8, a series of methods for calculating thermodynamic system properties directly from molecular dynamics simulations are discussed further.

2.1.5 Basic equilibration and simulation protocol

MD-simulations of several different systems are presented in the work. Although the systems change, the same equilibration/simulation protocol is always used. Before production runs are performed, the initial system must be brought into equilibrium. The equilibration protocol used here can be summarized by the following steps:

1. Minimization with restraints on the solute molecules
2. Minimization of the whole system
3. Heating of the system in the NVT ensemble with restraints on the solute molecules
4. Equilibration of the density in the NPT ensemble
5. Production run

Before any dynamics of the system is performed, the systems is minimized in two steps. The first minimization aims at just removing bad solute-water or water-water contacts from the initial setup, and is done with weak restraints on the solute. The second minimization allows for relaxation of the initial solute structure as well. The first equilibrium step is a heating stage, where the temperature of the system is brought to the desired value. This is achieved by a short MD-run in the NVT-ensemble, using positional restraints on the solute to avoid any large conformational changes at this stage. The second equilibration run is a fully unrestrained MD-run the NPT-ensemble, where the density of the system is equilibrated. The simulation time for the last step

is 200 ps in these simulations, which is long enough to have stable fluctuations of the macroscopic variables around their equilibrium values. In the whole equilibration phase, the Langevin thermostat [61] is used to ensure that the temperature is evenly distributed throughout the system. The production run variables vary somewhat from simulation to simulation and are presented in the text of respective chapter.

2.2 Analysis methods

Post-processing of the calculated molecular dynamics trajectories is done in this work using series of analysis tools implemented as extensions of the PTRAJ analysis tool of the AMBER molecular dynamics package. The following chapter describes the underlying theory of the most important analysis methods used in later chapters.

2.2.1 Distribution functions

The distribution function describes fluctuations in density around a particle and provides structural information of a system. Distribution functions can be determined both from computer simulations and from neutron scattering or X-ray scattering experiments and therefore serve as an important bridge between simulation and experiment for fluids. The distribution function is defined as the probability of finding a pair of molecules at a given distance r , relative to the probability expected for a completely random distribution at the same density. If the available volume around an atom i , at any distance r , is spherical and isotropic, the distribution is fully described by the so-called radial distribution function $g(r)$, which can be calculated from computer simulations as:

$$g(r) = \frac{1}{n_{rand}} \frac{1}{N_{fr}} \sum_{n=1}^{N_{fr}} \sum_{j=1}^{N_{at}} f(\mathbf{r}_i, \mathbf{r}_j) \quad (2.20)$$

$$f(\mathbf{r}_i, \mathbf{r}_j) = \begin{cases} 1 & \text{if } r < |\mathbf{r}_i - \mathbf{r}_j| \leq r + \delta r \\ 0 & \text{otherwise} \end{cases} \quad (2.21)$$

$$n_{rand} = \frac{4\pi\rho}{3} ((r + \delta r)^3 - r^3) \approx 4\pi\rho r^2 \delta r \quad (2.22)$$

Here N_{fr} is the number of frames of the trajectory over which the distribution function is calculated and averaged over. \mathbf{r}_i is the position vector of atom i and the second sum runs over all other atoms in the system (with position vectors \mathbf{r}_j). The binary function $f(\mathbf{r}_i, \mathbf{r}_j)$ is unity only when another atom is present in the spherical shell around atom

i, at distance r , with thickness δr . The normalization factor n_{rand} (Eq. 2.22) gives the number of atoms in the same volume at density ρ at a random distribution of all other atoms and ensures that $g(r)$ is unity for a completely random distribution. The radial distribution function can further be used to investigate the local structure around a given atom. The position of the first peak gives the average distance to the nearest neighbors of the central atom. The height of the first peak is related to the average number of nearest neighbors. The coordination number can be obtained by integrating the radial distribution function from zero to the distance of its first minimum, R_{min} :

$$n_{coord} = \rho \int_0^{R_{min}} 4\pi r^2 g(r) dr \quad (2.23)$$

Radial distribution functions are not suitable for investigating structural trends around larger non-spherical molecules. The available volume at a distance r , from a given atom in a large molecule, is not well approximated by a spherical shell of volume $4\pi r^2 \delta r$ since parts of this volume will be occupied by other atoms in the molecule. An alternative is to use the proximal distribution function proposed by by Ashbaugh *et. al.* [62]. The proximal distribution function $g_{prox}(r)$ is calculated as:

$$g_{prox}(r) = \frac{\langle n_i(r, \delta r) \rangle}{\rho \langle V_s(r, \delta r) \rangle} \quad (2.24)$$

Here $\langle n_i(r, \delta r) \rangle$ is the average number of atoms within a shell of thickness δr at distance r from the *closest* atom in the central molecule. ρ_i is the bulk density of the molecule or atom for which the distribution function is calculated and $\langle V_s(r, \Delta r) \rangle$ is the average volume of the shell. The volume of a given shell around the central molecule will change as it undergoes conformational changes and is therefore averaged over the whole trajectory. The instantaneous volume of the shell around the molecule, at distance r and thickness δr , is calculated from the surface area of the union of spheres with radii r , centered on each atom j :

$$V_s(r, \delta r) = \sum_j \Omega_j(r) r^2 \delta r, \quad (2.25)$$

where $\Omega_j(r) r^2$, with $0 \leq \Omega_j(r) \leq 4\pi$, is the accessible surface area of the sphere of radius r centered at atom j , determined by a Monte Carlo sampling.

2.2.2 Hydrogen bond analysis

Hydrogen bonds, i.e. electrostatic interaction between a hydrogen atom covalently bonded to an electronegative atom (the donor) and another non-bonded electronegative atom (the acceptor), are ubiquitous in biological systems and play an important role in biochemical processes. Hydrogen bonding properties can be investigated in full detail using computer simulations. Simulations have shown that hydrogen bond interaction energies strongly depend on the geometry of the involved atoms and a wide distribution of interaction energies can be seen for a typical polar system [63].

Two different approaches for identifying hydrogen bonds can be found in the literature. A hydrogen bond is formed between a hydrogen bond donor and a hydrogen bond acceptor when the interaction energy between the two atoms is above some given threshold [64] or based on some geometric criteria of the donor, the donor hydrogen and the acceptor atoms. In this work, the latter approach is used. A hydrogen bond between a donor and an acceptor atom, at a given time step of the trajectory, is present when (see Figure 2.2):

1. The distance r is smaller than a cutoff distance r_{cut} .
2. The angle θ is smaller than θ_{cut} .

The constants r_{cut} and θ_{cut} are somewhat arbitrary and depend on the type of hydrogen bonds investigated. For water-water hydrogen bonds r_{cut} is typically set to be the first minimum of the water oxygen-oxygen radial distribution function. r_{cut} is set to be 3.6 Å and a value of 30° is used for θ_{cut} . Martí *et al.* [65] showed from simulations, using the SPC water model, that hydrogen bonds fulfilling the distance criteria (1), have only small probability to have hydrogen bond angles greater than 30° at a range of different temperatures.

2.2.3 Time correlation functions

A series of different time correlation functions are used in this work for investigating time-dependent equilibrium system properties. Although details of how these correlation functions are calculated differ, the basic approach is the same. A general time

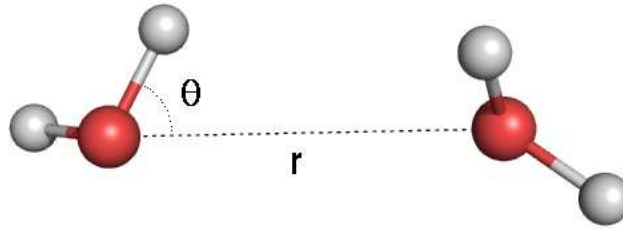


Figure 2.2: Geometric criteria used to identify hydrogen bonds

autocorrelation function, $C(t)$ of a given quantity $A(\mathbf{r}^N, \mathbf{p}^N)$ as a function of the whole phase-space, is given by

$$C(t) = \langle A(t)A(\tau + t) \rangle \quad (2.26)$$

where the brackets denote an ensemble average and τ is a given time origin of the trajectory. By dividing $C(t)$ with $\langle A(0)A(0) \rangle$ the correlation function is normalized to take values between -1 and 1. A value of $C(t)$ close to unity indicates a high degree of correlation with time, whereas zero indicates that the measured property is fully uncorrelated at time $\tau+t$. Negative values of the correlation function show that the measured property is anti-correlated at a given time. A general trend of the autocorrelations functions in the liquid phase is that they take values close to unity at short times and decay asymptotically to zero at long time intervals, as averages are evened out by random thermal motions.

In order to increase the sampling of a given time dependent process, measurements are in many of the following autocorrelation functions, calculated using the overlapped data collection approach [66]. Since $C(t)$ do not depend on the time origin, τ , the correlation function can be calculated using several overlapping measurements (see Figure 2.3). Ideally, the overlap should be done over time intervals where the correlation between

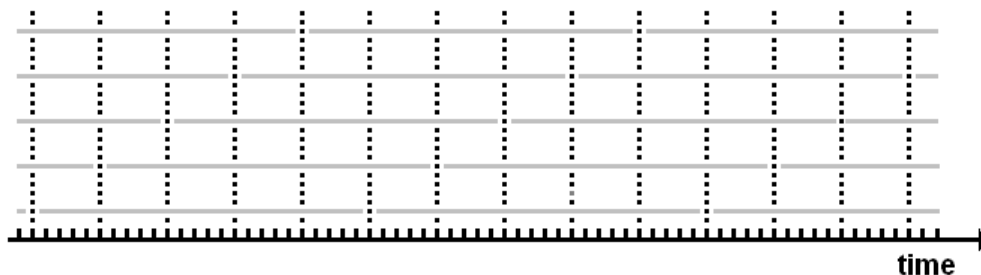


Figure 2.3: Pictorial description of the overlapped data method for calculating time correlation functions.

the measurements has converged to a small value.

Normalized autocorrelation functions can be fitted to a given analytical expression to determine characteristic decay times. Furthermore, average times of a given time-dependent process can be calculated by integrating the autocorrelation function or the fitted function from zero to infinity:

$$\tau = \int_0^{\infty} C(t) dt. \quad (2.27)$$

Rotational autocorrelation function

Molecular rotational dynamics can be investigated using the second-order rotational autocorrelation function [67],

$$C_{\mu}^2(t) = \langle P_2[\mu(t)\mu(0)] \rangle, \quad (2.28)$$

where P_2 is the second order Legendre polynomial and μ is a unit vector fixed in the molecule. For water molecules μ is chosen as the dipole vector, i.e. the vector bisecting the H-O-H angle. The second-order orientational relaxation time, τ_2 , can be calculated from $C_{\mu}^2(t)$ using Eq. 2.27. In the case of water, it describes the average time needed

for the dipole vector to become uncorrelated to its initial direction, and can be used to quantify the rotational freedom of water molecules.

Velocity autocorrelation function

The velocity autocorrelation function and its frequency spectrum can be used to probe translational dynamics of liquids. For a molecular liquid the unnormalized center of mass velocity autocorrelation function is given by [68]:

$$C_V^{cm}(t) = m \langle \mathbf{v}_i^{cm}(t) \mathbf{v}_i^{cm}(0) \rangle, \quad (2.29)$$

where \mathbf{v}_i^{cm} is the center of mass velocity of molecule i and m is the mass of the molecular species under investigation. Further discussions about the velocity autocorrelation function and its frequency spectrum can be found in Section 2.2.8.

Hydrogen bond lifetime autocorrelation function

Dynamics of the water hydrogen bond network can be measured by the hydrogen bond lifetime, or survival probability, autocorrelation function $C_{HB}(t)$ [69, 70]:

$$C_{HB}(t) = \frac{\langle \eta_{ij}(t) \eta_{ij}(0) \rangle}{\langle \eta_{ij}(0)^2 \rangle}, \quad (2.30)$$

where $\eta_{ij}(t)$ is a binary function defined as:

$$\eta_{ij}(t) = \begin{cases} 1 & \text{if molecules } i,j \text{ are hydrogen bonded at both time } t \text{ and } 0, \\ & \text{without being broken for a period longer than } \tau^* \\ 0 & \text{otherwise} \end{cases} \quad (2.31)$$

Hydrogen bonds are, as described in Section 2.2.2, defined using geometric criteria

imposed on the involved atoms. When $\tau^* \neq 0$, the intermittent hydrogen bond lifetime is measured, allowing hydrogen bonds to break and reform during a time interval controlled by τ^* . Setting τ^* to zero, continuous hydrogen bond lifetimes are measured.

Survival time autocorrelation function

The survival time autocorrelation function is calculated in a similar manner as the hydrogen bond lifetime autocorrelation function and can be used to measure residence times of a molecule in a given volume of the simulation box. Here, $p_{R,j}(t)$ is a binary function associated with each atom or molecule j under investigation. $p_{R,j}(t)$ is unity whenever the atom or molecule under consideration was in the layer R at a given time origin and remains in the layer for the time period t , without leaving the layer at any intermediate time step, and zero otherwise. Average residence times can be calculated directly from $C_R(t)$ (see Eq. 2.27), which gives the probability that, for instance, a water molecule remains in a given volume around an ion for a time period t .

$$C_R(t) = \frac{1}{N_J} \sum_{j=1}^{N_J} \frac{\langle p_{R,j}(0) p_{R,j}(t) \rangle}{\langle p_{R,j}(0) \rangle^2} \quad (2.32)$$

The survival time autocorrelation function and corresponding average survival times (see Eq. 2.27) can be used to probe the dynamics of non-hydrogen bonded molecular interactions, e.g. dynamics of the hydration shell of a hydrated ion.

2.2.4 Voronoi tessellation

Voronoi tessellation is the decomposition of space, here the simulation box, into polyhedra centered around atoms, in such a way that each polyhedron contains only a single atom and that any point in the polyhedron is closer to this atom than to any other atom in the simulation box. Voronoi polyhedra are equivalent to the Wigner-Seitz cells of crystal solids and are therefore suitable for characterization of coordination geometries and local structure of fluids. By definition, Voronoi polyhedra are also suitable for investigating packing and density fluctuations in molecular systems. The algorithm used to decompose the simulation box into Voronoi polyhedra is from Rapaport [66].

In summary, this algorithm works by assigning a large tetrahedron around a given atom spanning the whole simulation box. The shape of the Voronoi polyhedron is then successively pruned by removing parts of the initial polyhedron, cut by bisecting planes between the central atom and its closest neighbors. An example of a Voronoi polyhedron constructed around a sodium ion can be seen in Figure 2.4.

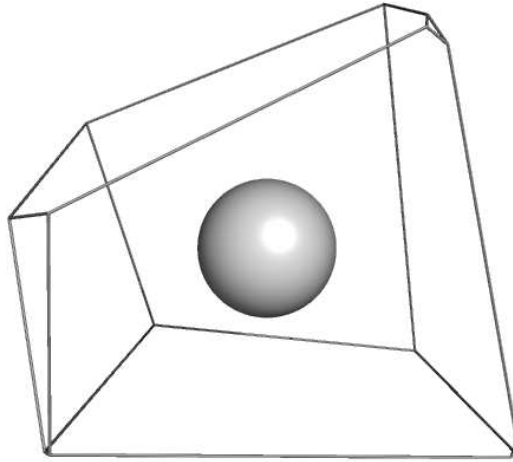


Figure 2.4: Example Voronoi polyhedron of a solvated sodium ion

The asphericity parameter of Ruocco *et al.* [71] characterizes the shape of the Voronoi polyhedron using the volume and the surface area of the polyhedron. The area of face k , of a polyhedron, is calculated as the sum over areas of all triangles into which it can be decomposed:

$$A_k = \frac{1}{2} \sum_{i=2}^{n_v^{(k)}-1} |(r_{i+1}^{(k)} - r_1^{(k)}) \times (r_i^{(k)} - r_1^{(k)})|, \quad (2.33)$$

where $r_j^{(k)}$ is the coordinate vector of vertex j in face k . The total area is then simply the sum over all faces:

$$S = \sum_{k=1}^{N_f} A_k \quad (2.34)$$

The polyhedron volume can be calculated using the areas of all faces and their respective distance d_f to the polyhedron center:

$$V = \frac{1}{3} \sum_{k=1}^{N_f} d_f \cdot A_f. \quad (2.35)$$

Given the volume and surface area, the dimensionless asphericity parameter is defined as:

$$\eta = \frac{S^3}{36\pi V^2}. \quad (2.36)$$

Using reference values for standard regular polyhedra; 1, 1.21, 1.32, 1.65, 1.91 and 3.31 for a sphere, an icosahedron, a dodecahedron, an octahedron, a cube and a tetrahedron respectively, coordination geometries can be determined of a given atom in the simulation box.

In this work Voronoi tessellation is used to calculate local density fluctuations and to investigate the local structure around a given molecule. In many cases it is of interest to investigate properties of the subset of molecules that are in direct contact with a given central molecule. Voronoi tessellation can be used for this purpose to identify direct neighbors as the subset of molecules sharing a polyhedron face with the Voronoi polyhedron of the central molecule.

2.2.5 Water order parameter

The structural order of water can be determined by the tetrahedrality of the water hydrogen network. Each water molecule can form four hydrogen bonds, two donor bonds involving the two hydrogen atoms and two acceptor bonds through interaction with hydrogen atoms of neighboring water molecules. The optimal arrangement of a water molecule i and its four closest neighbors takes the form of a tetrahedron, with

water molecule i in the center and the four neighbors located at the vertices of the tetrahedron. This arrangement corresponds to that found in hexagonal ice. In the liquid phase however, the hydrogen bond network is constantly rearranging and the local structure of water can take many forms.

Chau and Hardwick [72] defined the tetrahedral order parameter for a water molecule i as:

$$q_i = 1 - \frac{3}{8} \sum_{j>k} \left(\cos \psi_{ijk} + \frac{1}{3} \right)^2, \quad (2.37)$$

where ψ_{ijk} is the angle of the oxygen atom of the central water molecule i , and neighboring water oxygens j and k . The summation runs over all six possible angles spanned by central water oxygen i and neighboring water oxygens j and k . In a perfect tetrahedron ψ takes a values of 109.47° and $\cos(\psi)$ equals $-\frac{1}{3}$. Here, the normalization of Errington and Debenedetti [73] is used. In a perfect tetrahedron, the second term vanishes and q is unity. For a completely random distribution of the neighboring water molecules the average value of q is 0.

2.2.6 MM-PB(GB)SA

Free energy analysis using the MM-PB(GB)SA software, implemented in the AMBER molecular dynamics package, is used to analyze the free energy change associated with complex formation of a ligand and a receptor. In this section, the underlying theory of the methods used within the MM-PB(GB)SA approach is discussed.

For a given receptor A and ligand B , the association process to form complex AB can be written as the two state process:



The standard free energy change, ΔG^0 , associated with this process can either be calculated using the equilibrium dissociation constant K_d or directly from changes in entropy and enthalpy of the system in the complex state relative to the free ligand and receptor:

$$\Delta G^0 = -k_b T \ln K_d = \Delta H - T \Delta S^0, \quad (2.39)$$

where the dissociation constant K_d is measured as $[A]_{eq}[B]_{eq}/[AB]_{eq}$. Changes in entropy and enthalpy of the system can be calculated from computer simulations, either using potential of mean force approaches [74, 75], or so-called endpoint approaches, where ΔG^0 is calculated from absolute enthalpies and entropies of the two states:

$$\Delta G^0 = H_{complex} - (H_{receptor} + H_{ligand}) - T(S_{complex}^0 - (S_{receptor}^0 + S_{ligand}^0)) \quad (2.40)$$

In order to predict affinities of biological molecular association processes, changes in enthalpy and entropy of both solute molecules and the surrounding solvent need to be included. A common approximation is to neglect the $P\Delta V$ term in ΔH and simply use the change in potential energy ΔU in the free energy calculation. Using a pairwise potential, the potential energy of a solvated biomolecular system can be decomposed into pairwise contributions from solute-solute, solvent-solvent and solute-solvent interaction energies:

$$U(\mathbf{r}_u, \mathbf{r}_v) = U(\mathbf{r}_u) + U(\mathbf{r}_v) + U(\mathbf{r}_u; \mathbf{r}_v) \quad (2.41)$$

Here \mathbf{r}_u and \mathbf{r}_v denote solute and solvent degrees of freedom respectively. Although the total potential energy of the system, in principle, could be calculated directly from a MD simulation, fluctuations of the total potential energy of a large biomolecular system would be too large to converge into reliable averages in the simulation times available for such systems [76]. The simulation box of a typical explicit water biomolecular system consists mainly of water molecules and the major contribution to the total potential energy comes from water-water interactions, $U(\mathbf{r}_v)$. For computational reasons, it is therefore attractive to treat the free energy contribution from solute-solvent and solvent-solvent interactions implicitly and only calculate free energy contributions from solute-solute interactions explicitly. The MM-PB(GB)SA approach uses a continuum solvation model for calculating binding free energies in solution. Us-

ing the thermodynamic cycle illustrated in Figure 2.5, the energy change on binding in solution, $\Delta G_{bind,solv}$, can be calculated using the following path:

$$\Delta G_{bind,solv}^0 = \Delta G_{bind,vac}^0 + \Delta G_{solv,complex} - (\Delta G_{solv,receptor} + \Delta G_{solv,ligand}), \quad (2.42)$$

where $\Delta G_{bind,vac}$ refers to the change in free energy associated with the complex formation in vacuum, i.e. without taking any effects of changes in water-water or water-solute interactions into account:

$$\Delta G_{bind,vac}^0 \approx \Delta U_{bind,vac} - T\Delta S_{bind,vac}^0. \quad (2.43)$$

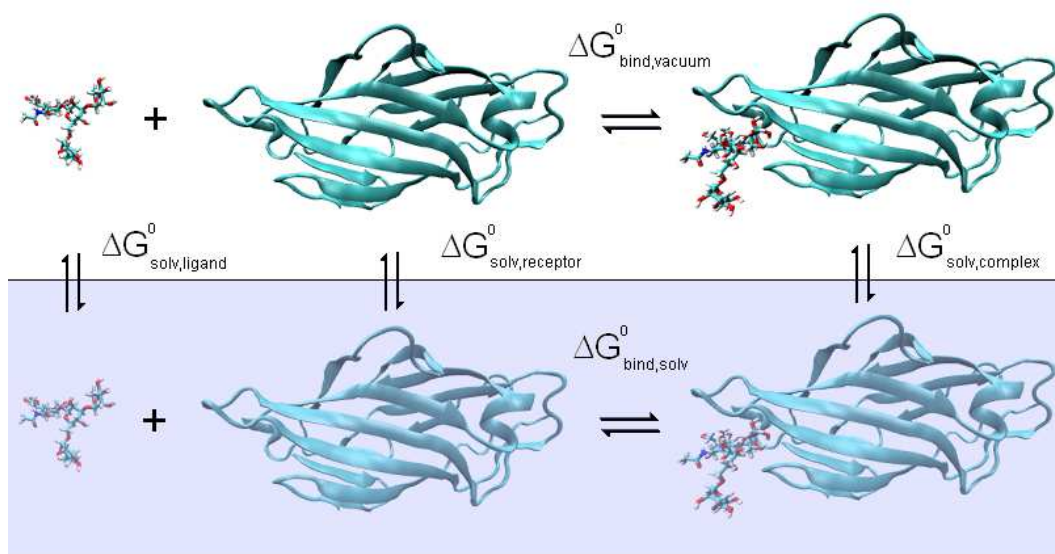


Figure 2.5: The thermodynamic cycle used to calculate $\Delta G_{bind,solv}^0$ in the MM-PB(GB)SA approach

The potential energy in vacuum can be calculated from the bonded and non-bonded terms from the solute force field and $\Delta U_{bind,vac}$ is simply the difference in solute potential energies for the complex and the free receptor and ligand molecules. Methods for estimating changes in solute entropies, $T\Delta S_{bind,vac}$, are discussed in Section 2.2.7.

The ΔG_{solv} term is the free energy change for the process of transferring a given solute

molecule from vacuum to an aqueous environment. In the MM-PB(GB)SA method, this free energy change is calculated using an implicit solvation model. In order to understand the connection between implicit solvation methods and the free energy of solvation, it is necessary to discuss some basic statistical thermodynamics theory of a solute in solution. Comprehensive reviews on implicit solvent theory can be found in the papers of Roux *et al.* [77] and Wang *et al.* [78].

The goal of implicit solvation methods is to be able to model conformational behavior of a solute molecule in solvation without having to include explicit water molecules in the simulation. For a general system of a solute molecule immersed in water with the total potential energy function, $U(\mathbf{r}_u, \mathbf{r}_v)$ in Eq. 2.41, the canonical configurational partition function is given by:

$$Z(N, V, T) = \int d\mathbf{r}_u d\mathbf{r}_v e^{-\beta U(\mathbf{r}_u, \mathbf{r}_v)}, \quad (2.44)$$

where $\beta = \frac{1}{k_b T}$. The ensemble average of any property as a function of the solute conformation can be calculated (see Equations 2.17 and 2.18) by integrating over the whole (solvent and solute) configurational space. The key to implicit solvation is the potential of mean force, $W(\mathbf{r}_u)$ of a given solute conformation \mathbf{r}_u . This potential of mean force captures, in an average way, the effect of the solvent and can be calculated from Eq. 2.44 by integrating over all possible configurations of the solvent:

$$Z(N, V, T) = \int d\mathbf{r}_u e^{-\beta W(\mathbf{r}_u)}, \quad (2.45)$$

where,

$$W(\mathbf{r}_u) = -k_b T \ln \int d\mathbf{r}_v e^{-\beta U(\mathbf{r}_u, \mathbf{r}_v)} = U(\mathbf{r}_u) - k_b T \ln \int d\mathbf{r}_v e^{-\beta(U(\mathbf{r}_v) + U(\mathbf{r}_u, \mathbf{r}_v))}. \quad (2.46)$$

Using Eq. 2.45 and 2.46, a reduced probability function only depending on solute coordinates is obtained. In the last equality in Eq. 2.46, the pure solute potential has been moved out of the integral since it is not a function of \mathbf{r}_v and the potential of mean force is calculated only over the solvent potential and the coupled solute-

solvent potentials. The total interaction potential can be written as a sum of the solute potential plus the potential of mean force term; $U(\mathbf{r}_u) + W(\mathbf{r}_u)$. Direct evaluation of the integral in Eq. 2.46 is not suitable from a computer simulation due to the vast conformational space of solvent molecules. $W(\mathbf{r}_u)$ can, however, be calculated approximately, using the properties of the potential of mean force function. It can be shown that the gradient of $W(\mathbf{r}_u)$ with respect to a given solute atomic Cartesian coordinate equals the average force exerted on the atom, averaged over all coordinates of the solvent [77]. It can further be shown that the absolute value of $W(\mathbf{r}_u)$ is not important and therefore can be calculated with respect to a reference state, where solute-solvent interactions are absent. Thus, $\Delta W(\mathbf{r}_u)$ can be calculated as the reversible work, or free energy change, associated with the process of transferring the solute, in a given configuration \mathbf{r}_u , from vacuum to solution, which is equivalent to ΔG_{sol} in the thermodynamic cycle in Figure 2.5. This process can be done using a thermodynamic integration approach, in which solute-solvent interactions are gradually turned on in a step-wise manner. This approach would not be computationally feasible for large biomolecules. Instead an end-point approach is used, where only the two end states are considered in the free energy calculation.

The solvation free energy is, in practice, calculated using another thermodynamic cycle. In a first step the molecule is stripped of charges in vacuum. The non-polar molecule is then transferred from vacuum to the aqueous phase and in a third step the non-polar solute is recharged in the solvent. The sum of free energy changes in these three steps will give the total solvation free energy change [79]:

$$\Delta G_{sol} = \Delta G_{ele,vac} + \Delta G_{np} + \Delta G_{ele,solv}. \quad (2.47)$$

ΔG_{np} refers to the free energy contribution from the process of transferring the uncharged solute molecule into water. This free energy term captures the energetics of creating a cavity in the hydrogen bond network of water, large enough to accommodate the solute. This is in total an unfavorable contribution to solvation free energy due to the energetic cost of disrupting water-water hydrogens bonds, outweighing the small favorable energetic contribution from solute-solvent van der Waals interactions. In the MM-PB(GB)SA approach, ΔG_{np} is estimated using a surface area (SA) model where non-polar free energy contribution is assumed to be proportional to the number of water molecules in the hydration shell of the solute, and hence, also to the solute

surface area. A linear relationship between the solvent accessible surface area and the solvation free energy has been verified for a series of linear alkenes of different sizes [79, 80]. The nonpolar vacuum to water transfer free energy is approximated by,

$$\Delta G_{np}(\mathbf{r}_u) = \gamma SASA(\mathbf{r}_u) + b, \quad (2.48)$$

where γ and b are fitted parameters of the slope and the intercept of surface area versus experimental solvation free energies plots.

The ΔG_{ele} term is the electrostatic contribution which describes the work needed to gradually assign charges to the atoms in water. This term captures contributions from electrostatic interactions between water and the solute and also the work needed to generate the solvent reaction field induced by the solute charge distribution. The work of charging the solute can be calculated by solving the Poisson equation:

$$\nabla \cdot \epsilon(\mathbf{r}) \nabla \phi(\mathbf{r}) = -\rho(\mathbf{r}) \quad (2.49)$$

Solving the differential equation in 2.49, using the charge density ρ and the dielectric constant ϵ , gives the electrostatic potential ϕ at a given point in the simulation box. The charge density at a position, \mathbf{r} , in the simulation box is given by the atomic point charges in the force field. The position-dependent dielectric constant, $\epsilon(\mathbf{r})$, is set to the room temperature value of water ($\epsilon=80$) outside the solute and to a value, typically between 1-4, inside the solute. The boundary of the solute-solvent interface is determined by the contact area of the van der Waals surface of the molecule with a spherical probe the size of approximately a water molecule rolled over the whole solute [81]. In practice, the differential equation in Eq. 2.49 is solved numerically by mapping charges and dielectric constants to grid points separated by a given spacing throughout the simulation box. Using a finite difference approach, the Poisson equation is solved numerically, to give the electrostatic potential at any given position in the simulation box. The work needed to put the solute molecule in conformation \mathbf{r}_u in the solvent reaction field, can be calculated as [77, 79],

$$\Delta G_{ele} = \frac{1}{2} \sum_i q_i (\phi_i^{\epsilon=80} - \phi_i^{\epsilon=1}), \quad (2.50)$$

where the summation index i runs over all atoms in the solute. ΔG_{ele} includes the free energy contributions from the first and third step in the above-mentioned thermodynamic cycle. Thus, ΔG_{ele} and ΔG_{np} together gives the solvation free energy ΔG_{sol} in Eq. 2.47.

An alternative, less computationally expensive, method for estimating the electrostatic free energy contribution is the generalized Born approach [82]. In this model, ΔG_{ele} is given by an analytical pairwise formula that is designed to reproduce electrostatic free energies calculated with the Poisson equation:

$$\Delta G_{ele} = -\frac{1}{2} \left(\frac{1}{\epsilon_s} - \frac{1}{\epsilon_w} \right) \sum_{i,j} \frac{q_i q_j}{f_{ij}^{gb}(r_{ij})}, \quad (2.51)$$

q_i and q_j are here the solute atomic charges and ϵ_s and ϵ_w are the dielectric constants of the solute and the solvent respectively. f_{ij}^{gb} is a function of the effective atomic radii, R_i and R_j , and the interatomic distance, r_{ij} . The generalized Born approach is, as suggested by the name, a generalization of the Born formula for calculating solvation free energies of ions, to systems of many atoms in any given configuration. A direct application of the Born formula to molecules is not suitable, since many atoms in the molecule are not completely surrounded by high dielectric water, but rather in contact with other atoms in the molecule. The key idea in Generalized Born approaches is to calculate effective atomic radii, describing to which degree a given atom is buried in the low dielectric medium of the solute, in a manner that the Born formula can be applied to each atom individually. There are several different General Born models available, mainly differing in the way the atomic radii are assigned. With proper assignments of effective atomic radii, the Generalized Born model can be used to calculate solvation free energies that are in good agreement with results from the Poisson approach, at a fraction of the computational cost [83].

2.2.7 Methods for calculating biomolecular entropies

The free energy analysis using the MM-PB(GB)SA approach is missing one important contribution to the free energy change, namely the entropy change of the solute molecules in the complex. Typical ΔG_{bind}^0 values of complex formation fall in the range of 5-12 kcal/mol. The free energy change is, however, due to enthalpy-entropy

compensation effects, much smaller than the changes in entropy and enthalpy alone. Non-covalent binding of a ligand to a receptor result in a limitation of the free space available to the involved solute molecules, which can result in relatively large and unfavorable entropy penalties.

Assuming thermodynamic additivity, the total entropy change of a solute species can be expressed as a sum of translational, rotational, vibrational and conformational entropies, which can be calculated separately:

$$\Delta S_{solute} = \Delta S_{rotations} + \Delta S_{translations} + \Delta S_{vibrations} + \Delta S_{conformations} \quad (2.52)$$

$\Delta S_{rotations}$ and $\Delta S_{translations}$ reflect the loss in rigid body rotational and translational motions of the ligand in the bound state compared to the free state. These two terms will always be negative and give a positive contribution to the overall free energy change. $\Delta S_{vibrations}$ and $\Delta S_{conformations}$ reflect changes in internal motions of both receptor and ligand in the complex relative to the free states. These two terms are typically also negative due to reduction of the configurational space available in the complex. Calculations of entropy changes associated with molecular association processes from molecular dynamics simulations are, however, not a trivial task. The entropy is, by definition, a quantity measuring the accessibility of the phase space of a collection of atoms. When calculating entropy changes, absolute entropies of the solute in the two states are needed and hence, the whole configurational space accessible to the molecule in both states need to be sampled during the simulation. Development of methods for estimating biomolecular entropies from computer simulations is still an active field of research and a vast variety of methods can be found in the literature. The following subsections describe the theory behind a series of methods for estimating biomolecular entropies used in this work.

Entropy of rigid body motions

The standard method for calculating absolute rotational and translational entropies is to use the analytical, one molecule, translational and rotational ideal gas partition functions. Under the assumption that a solute molecule do not interact with the solvent or any other molecules in the system, the one molecule translational partition function is given by [84]:

$$q_{trans}(V, T) = \frac{V}{\Lambda^3}, \quad (2.53)$$

where $\Lambda = (h^2/2\pi M k_b T)^{1/2}$. h is here Planck's constant and M , V and T is the total mass of the solute, the volume available for the solute and the temperature of the system. Using the standard formula for calculating the entropy from the canonical partition function, the entropy of a molecule in the gas-phase can be written as,

$$S_{trans} = k_b \ln \left[\left(\frac{2\pi M k_b T}{h^2} \right)^{3/2} \frac{V e^{5/2}}{N} \right], \quad (2.54)$$

where e is Euler's number and N is the number of ideal gas particles. Eq. 2.54 is also known as the Sackur-Tetrode equation. The translational entropy depend on the concentration of the molecule and is typically calculated for a standard state of solute concentration of 1 M (= 1 molecule/1660Å³).

The rigid body rotational entropy can be calculated in a similar manner, using the one molecule ideal gas rotational partition function [84]:

$$S_{rot} = k_b \ln \left[\frac{\sqrt{\pi}}{\sigma} \left(\frac{8\pi^2 I_A k_b T e}{h^2} \right) \left(\frac{8\pi^2 I_B k_b T e}{h^2} \right) \left(\frac{8\pi^2 I_C k_b T e}{h^2} \right) \right], \quad (2.55)$$

where I_A, I_B and I_C are the three molecular principal moments of inertia, and σ is a symmetry number, which typically is unity for large asymmetric biomolecules.

Although ideal gas rotational entropies can be a good approximation for molecules in solution, the ideal gas treatment fails to predict translational entropies even for monoatomic species in solution [85]. Ideal gas rigid body entropies are clearly a rather

rough estimation for large biomolecules in solution.

An alternative method to calculate the rigid body entropy loss of a ligand in a complex, relative to the unbound state, is provided by Swanson *et al.* [86]. This method is based on statistical thermodynamics of non-covalent binding of molecules in solution. The basic theory of this method is described in the following. For full derivation of the method, the reader is referred to the original paper. The standard free energy change for molecular association of molecules A and B to form complex AB can be calculated from the ratio of configurational integrals, $Z_{N,AB}$, $Z_{N,A}$, $Z_{N,B}$ and $Z_{N,O}$ of the complex, receptor, ligand and solvent alone [87]:

$$\Delta G_{bind,AB}^0 = -RT \ln \left(\frac{C^0}{8\pi^2} \right) \left(\frac{Z_{N,AB} Z_{N,O}}{Z_{N,A} Z_{N,B}} \right) + P^0 \langle \Delta V_{AB} \rangle \quad (2.56)$$

Here, $Z_{N,A}$ denotes the configurational integral in internal coordinates of molecule A plus the external coordinates of N solvent molecules. C^0 is the concentration of the solute molecule, which typically is set to 1 M (= 1 molecule/1660 Å³). Swanson *et al.* introduced a series of approximations for evaluating Eq. 2.56 directly from computer simulations. The first approximation is to treat solvent degrees of freedom implicitly by a potential of mean force method, as described in Section 2.2.6. Using an implicit solvation model, the configurational integrals in Eq. 2.56 can be solved only with respect to the internal degrees of freedom of the solute molecule:

$$Z = \int e^{-\beta U(\mathbf{r}_u) + W(\mathbf{r}_u)} d\mathbf{r}_u, \quad (2.57)$$

where $U(\mathbf{r}_u)$ is the potential energy of the solute and $W(\mathbf{r}_u)$ is the potential of mean force of a given solute conformation \mathbf{r}_u . Using the reduced configurational integrals, the standard free energy of binding can be written as:

$$\Delta G_{bind,AB}^0 = -RT \ln \left(\frac{C^0 Z_{AB}}{8\pi^2 Z_A Z_B} \right). \quad (2.58)$$

Here the pressure-volume work in Eq. 2.56 is neglected and the configurational integrals are only evaluated over the internal degrees of freedom of the solute molecules.

Swanson *et al.* introduced a new set of coordinates, $\delta_n = (x_1, x_2, x_3, \xi_1, \xi_2, \xi_3)$, for the six

additional degrees of freedom of the the complex. These degrees of freedom represent the six translational and rotational degrees of freedom of the ligand transformed into internal degrees of freedom in the complex. These six degrees of freedom can be separated out from the configurational integrals to give,

$$\Delta G_{bind,AB}^0 = -RT \ln \left(\frac{C^0 z_B^{trans} z_B^{rot} Z_{AB'}}{8\pi^2 Z_A Z_B} \right), \quad (2.59)$$

where AB' denotes the remaining degrees of freedom of the complex. The configurational integrals of the rigid bonds motions of the ligand, in the complex are given by,

$$z_B^{trans} = \int e^{U(x_1, x_2, x_3) + W(x_1, x_2, x_3)} dx_1 dx_2 dx_3 \quad (2.60)$$

and

$$z_B^{rot} = \int e^{U(\xi_1, \xi_2, \xi_3) + W(\xi_1, \xi_2, \xi_3)} d\xi_1 d\xi_2 d\xi_3 \quad (2.61)$$

Eq. 2.59 is based on the approximation that ligand rigid body rotations and translations in the complex are fully decoupled. Further assuming that the configurational space of the ligand and receptor is the same in the free and bound states, Swanson *et al.* simplified Eq. 2.59 further to:

$$\Delta G_{bind} = -RT \ln \left(\frac{C^0 z_B^{trans} z_B^{rot}}{8\pi^2} \right) + (\langle E_{AB} \rangle - \langle E_A \rangle - \langle E_B \rangle), \quad (2.62)$$

where $\langle E_{AB} \rangle$, $\langle E_A \rangle$ and $\langle E_B \rangle$ are the potential energies plus the solvation free energy of the complex, ligand and receptor respectively, which can be calculated from molecular dynamics simulations together with implicit solvation methods (i.e. the PBSA or GBSA methods described in Section 2.2.6). Thus, the free energy contribution from loss of translational and rotational degrees of freedom in the complex, with respect to the free standard state, is given by the first term in Eq. 2.62. The main difficulty of this approach is to evaluate the configurational integrals in Eq. 2.60 and 2.61. Instead of attempting to evaluate the integrals numerically, Swanson *et al.* suggested to approximate the configurational integrals by a quasi-harmonic approach. The average

potential energy of an one-dimensional classical harmonic oscillator is related to the variance of the oscillator coordinate, $\langle \Delta x^2 \rangle$ by,

$$\langle U(\mathbf{r}_u) + W(\mathbf{r}_u) \rangle = \frac{1}{2}k \langle \Delta x^2 \rangle, \quad (2.63)$$

where the variances of the a given oscillator coordinate can be measured from an explicit water MD simulation. In order to treat oscillations of the three translational degrees of freedom as independent harmonic oscillators, variances are determined as eigenvalues of the covariance matrix of ligand center of mass movements in the complex, after overall translational and rotational movements of the complex are removed by superposing the coordinates of the receptor molecule onto a set reference coordinates. With the quasi-harmonic approximation the translational configurational integral in Eq. 2.60 can then be calculated as:

$$z_B^{trans} = \int e^{(-k_1 \Delta x_1^2 / 2k_B T)} dx_1 \int e^{(-k_2 \Delta x_2^2 / 2k_B T)} dx_2 \int e^{(-k_3 \Delta x_3^2 / 2k_B T)} dx_3, \quad (2.64)$$

where Δx_1 , Δx_2 and Δx_3 are the eigenvalues of the ligand center of mass covariance matrix. Solving the integral in Eq. 2.64 and using the equipartition theorem to connect the force constants of the harmonic oscillators to the variances ($k_i \cong k_B T / \Delta x_i^2$), the final formula to estimate the configurational integral of ligand center of mass movement in the complex is given by:

$$z_B^{trans} = (2\pi)^{3/2} (\langle \Delta x_1^2 \rangle \langle \Delta x_2^2 \rangle \langle \Delta x_3^2 \rangle)^{1/2} \quad (2.65)$$

Rotational configurational integrals of the ligand in the complex are calculated from Eq. 2.63-2.65, replacing Δx^2 with $\Delta \xi^2$. Here the covariance matrix is calculated from displacements of the ligand Euler angles in the reference frame of the receptor in the complex. Once the eigenvalues of the translational and rotational covariance matrices are determined, the integrals in Eq. 2.60 and 2.61 can be calculated under the quasi-harmonic approximation and finally inserted into Eq. 2.62, to give the binding free energy change ΔG_{bind}^0 (neglecting changes in internal degrees of freedom of both receptor and ligand in the complex).

Entropy of internal motions

In this section, methods for estimating $\Delta S_{vibrations}$ and $\Delta S_{conformations}$ are discussed. For a general large non-linear biomolecule, this is the change in entropy due to restrictions of the $3N-6$ internal degrees of freedom of the molecule in the complex relative to the free state. The internal dynamics of large biomolecules, such as proteins, can be described by motions with a wide frequency range, ranging from low frequency delocalized vibrational modes, where a large number of atoms are involved in collective motions [88], to high frequency motions, typically involving a smaller subset of atoms. Angle bending and bond stretching motions are examples of the latter category. Although there is evidence that the low frequency vibrational modes of proteins, responsible for the major part of the internal atomic displacements, are anharmonic in nature at physiological temperatures [89], a common approximation in computational chemistry is to consider all vibrational modes as simple harmonic oscillations around a local energy minimum.

Absolute vibrational entropies can be estimated using normal mode analysis, based on the harmonic approximation. Harmonic normal modes of vibration, of a given local minimum solute conformation, can be calculated from the Hessian matrix,

$$H_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j}, \quad (2.66)$$

where q_i are the mass-weighted Cartesian coordinates of the solute. The Hessian can be evaluated directly from analytical derivatives of the force field potential energy function. The normal modes of vibrations and corresponding vibrational frequencies can be calculated from the eigenvectors and eigenvalues from an orthogonal transform of the Hessian matrix. The absolute vibrational entropy of the molecule can then be calculated, by summing over all $3N-6$ vibrational modes, using the analytical formula of the entropy of a quantum harmonic oscillator,

$$S_{ho} = \frac{k_B \alpha}{e^\alpha - 1} - \ln(1 - e^{-\alpha}), \quad (2.67)$$

with $\alpha = \hbar\omega/k_B T$, where ω is the frequency of the oscillator and $\hbar = h/2\pi$. Normal mode analysis can be used to calculate vibrational entropy changes in molecular associ-

ation processes by evaluating the difference in absolute entropies between the complex and the free ligand and receptor.

Schlitter [90] derived an elegant method for calculating configurational entropies, i. e. both vibrational and conformational entropies, of biomolecules from molecular dynamics trajectories. Schlitter showed that the probability distribution that maximizes the entropy of a given variance is the Gaussian distribution. Connecting the variance of fluctuations along a given degree of freedom of a molecule to the entropy of a harmonic oscillator, therefore gives an upper limit of the real configurational entropy. Using the analytical expression for the entropy of an one-dimensional quantum harmonic oscillator in Eq. 2.67, Schlitter introduced the approximation,

$$S' = \frac{k_B}{2} \ln \left(1 + \frac{e^2}{\alpha^2} \right) \quad (2.68)$$

and showed that $S' > S_{ho}$. Thus, $S' > S_{ho} \geq S$, where S denotes the true entropy of the system. The frequency of a quantum harmonic oscillator can be connected to the classical variance, which can be measured from MD simulations using the equipartition theorem,

$$m\omega^2 \langle x^2 \rangle \cong k_b T, \quad (2.69)$$

where $\langle x^2 \rangle$ is the classical variance of the harmonic mode. The equality in Eq. 2.69 is valid for low frequency modes where $\hbar\omega \ll k_b T$. For very high frequency vibrational modes is Eq. 2.69 not valid, but since these high frequency vibrations only make small contributions to the total configurational entropy, this is not a serious approximation in this method. Combining Eq. 2.68 and 2.69, the upper limit of the absolute entropy of an one-dimensional harmonic oscillator can be written as:

$$S' = \frac{k}{2} \ln \left(\frac{kT e^2}{\hbar} m \langle \Delta x^2 \rangle + 1 \right), \quad (2.70)$$

Eq. 2.70 can easily be extended to many degrees of freedom. In order to account for correlated atomic motions in a molecule, the covariance matrix of atomic coordinate fluctuations,

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle, \quad (2.71)$$

can be diagonalized, yielding a new set of $3N-6$ uncorrelated internal modes. The brackets in Eq. 2.71 denote the average Cartesian position of a given atom. The eigenvalues of the covariance matrix give variances along each mode and the entropy estimate for the one-dimensional case can be applied separately to each uncorrelated mode. The total absolute entropy can be obtained by summing over entropy contributions from each mode, or in an alternative form,

$$S' = \frac{k}{2} \ln \det \left(\frac{kT e^2}{\hbar} \mathbf{M} \sigma + \mathbf{1} \right), \quad (2.72)$$

where \mathbf{M} is a matrix holding the atomic masses in the diagonal and $\mathbf{1}$ is the unity matrix. The determinant in Eq. 2.72 can be solved numerically using a matrix decomposition algorithm, such as LU decomposition. Eq. 2.70 show that by using the approximative form, S' , instead of the analytical formula of the quantum harmonic oscillator, there is no need to calculate the covariance matrix in internal coordinates when calculating the entropy of internal degrees of freedom. Using the full $3N \times 3N$ covariance matrix in Cartesian coordinates, the six rotational and translational degrees of freedom can be 'frozen out', for instance by superposing the coordinates to a reference frame. By diagonalizing the covariance matrix, the variance of these six modes will be zero, or very close to zero, and will not make any significant contribution to the calculated overall absolute entropy.

In this work, we restrict the use of Schlitter entropy calculations to the internal degrees of freedom, but also rotational and translational molecular degrees of freedom can be included in the covariance matrix in Eq. 2.71. The calculations of Schäfer *et al.* [91], using Schlitter's method on different systems, showed that the method performs well for systems where the harmonic approximation is valid. The largest errors were in this study found for an ideal gas, where calculated (translational) entropies differed 17% compared to the exact analytical value.

2.2.8 Two-phase model entropy calculations

The harmonic approximation used in many of the previously discussed methods for estimating entropies is not suitable for liquids. In the aforementioned study of Schäfer *et al.*, the largest errors using Schlitter's method were found for the entropy of an ideal gas, where a large part of the atomic motions can be described by low frequency diffusional motions. These motions are anharmonic in nature and not well described by the motions of a harmonic oscillator. Lin *et al.* [68] proposed the so called 2PT (two-phase-thermodynamic) method for calculating thermodynamic properties of systems in all phases. This method is based on the observation that the density of states (DoS) function of an atomic liquid typically is a superposition of the solid and gas phase DoS functions of the same system. Figure 2.6 shows vibrational DoS functions for an

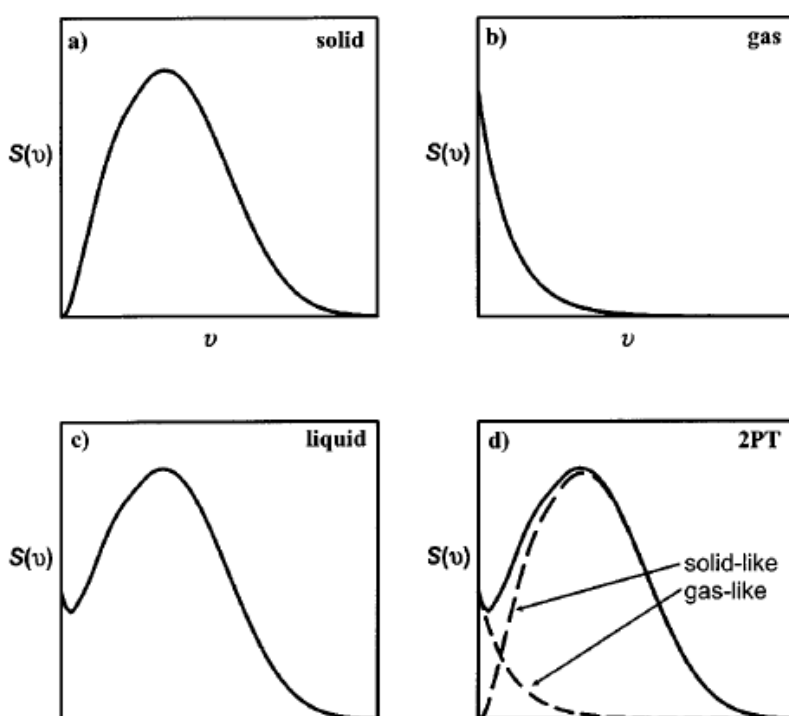


Figure 2.6: Figure from original 2PT paper [68]. Typical vibrational density of states distributions of a) a solid b) a gas and c) a liquid. Figure d) shows that the density of states function of a liquid can be described by a superposition of solid-like and gas-like density of states functions.

example system in the solid, liquid and gas phase. The 2PT-method is based on a decomposition of the DoS function of a liquid into contributions from solid-like vibra-

tional motions and diffusional gas-like motions. The solid-like part of the DoS function contain high frequency motions that can be modeled as independent harmonic oscillators, as previously discussed, whereas motions in the gas-like part are modeled by a hard sphere fluid. Using the decomposed DoS function the total entropy of the liquid can be calculated using the analytical formulas for the entropy of a harmonic oscillator and a hard sphere fluid respectively, weighted by the vibrational intensity at a given frequency. Although the method originally was developed for atomic fluids, it can easily be extended to water molecules [92, 93], thereby providing a method for calculating the absolute translational entropy of a set of water molecules.

The vibrational density of states of water motions, $D(\nu)$, can be calculated from the Fourier transform using the the molecular center of mass velocity autocorrelation function $C_v^{cm}(t)$ (see Eq. 2.29):

$$D(\nu) = \frac{2}{kT} \lim_{\tau \rightarrow \infty} \int_{-\tau}^{\tau} C(t) e^{-i2\pi\nu t} dt, \quad (2.73)$$

Lin *et al.* showed that integrating the DoS function over all frequencies gives the number of degrees of freedom of the system, which when calculating the center of mass velocity autocorrelation function for N water molecules, adds up to a total of 3N translational degrees of freedom. Within the 2PT method, the DoS function is decomposed into contributions from solid like vibrational modes $D^s(\nu)$ and and low frequency gas-like diffusive modes $D^g(\nu)$:

$$D(\nu) = D^g(\nu) + D^s(\nu) \quad (2.74)$$

Here, $D^g(\nu)$ corresponds to 3Nf degrees of freedom and $D^s(\nu)$ to the remaining 3N(1-f) degrees of freedom. The factor f is called the fluidity factor and determines the partition of the translational DoS function into solid and gas components. In the 2PT method, the fluidity factor is calculated as a function of temperature, density, mass and diffusivity of the system. For a system at low temperature and high density, the fluidity factor will approach zero as the diffusion coefficient approaches zero and the system is represented by the solid state. In the other limit, at high temperature and low density, the fluidity factor approaches unity and the system will be represented as a dilute gas. Lin *et al.* defined the fluidity factor as the ratio of the self-diffusivity

of the system with respect to the hard-sphere diffusivity in the zero pressure limit. The full derivation of the formula to calculate the fluidity factor is lengthy and therefore not repeated here. The final equation of the fluidity factor is given by:

$$2\Delta^{-\frac{9}{2}}f^{\frac{15}{2}} - 6\Delta^{-3}f^5 - \Delta^{-\frac{3}{2}}f^{\frac{7}{2}} + 6\Delta^{-\frac{3}{2}}f^{\frac{5}{2}} + 2f - 2 = 0, \quad (2.75)$$

where the function Δ is proportional to the diffusivity of the system and takes the form:

$$\Delta(T, \rho, m, D(0)) = \frac{2D(0)}{9N} \left(\frac{\pi kT}{m} \right)^{\frac{1}{2}} \rho^{\frac{1}{3}} \left(\frac{6}{\pi} \right)^{\frac{2}{3}}, \quad (2.76)$$

where ρ is the density of the system and m is the mass of the particles. $D(0)$ is the density of vibrational states at $\nu=0$ and is related to the self-diffusion coefficient D_s by:

$$D(0) = \frac{12mND_s}{k_B T}. \quad (2.77)$$

Using Equations 2.75 and 2.76, the fluidity factor f is determined by system specific parameters that can be calculated directly from the simulation.

Once the fluidity factor is known, the density of states function for the gas component can be calculated from the velocity autocorrelation function of a hard sphere fluid using Eq. 2.73. The resulting hard sphere density of states function is given by:

$$D^g(\nu) = \frac{D(0)}{1 + \left[\frac{\pi D(0)\nu}{6fN} \right]^2}, \quad (2.78)$$

Here, N is again the total number of water molecules (for which the autocorrelation function $C(t)$ was calculated for). $D^s(\nu)$ can now be calculated by simply subtracting $D^g(\nu)$ from the total DoS function $D(\nu)$. Using the partitioned DoS functions, the entropy of the $3N$ translational degrees of freedom can be calculated by integrating over the whole frequency range, using the analytical formulas for the entropy at a

given frequency of a quantum harmonic oscillator and a hard sphere diffusive fluid respectively,

$$S_{2PT} = \int_0^\infty d\nu D^s(\nu) W_{ho}^s(\nu) + \int_0^\infty d\nu D^g(\nu) W_{hs}^g(\nu), \quad (2.79)$$

where W_{ho} is given by Eq. 2.67 and W_{hs} is:

$$W_{hs}^g(\nu) = \frac{1}{3} \frac{S_{hs}}{k}, \quad (2.80)$$

where S_{hs} is the excess entropy of a hard sphere fluid, calculated from the Carnahan-Sterling equation of state of hard spheres [94].

As previously mentioned, the 2PT-method can only be used to calculate the entropy associated with the 3N translational degrees of freedom of water molecules. To calculate total absolute entropies, rotational entropies, as well as bond vibration entropies need to be accounted for. Since the water force fields used in this study are rigid water models, the latter contribution is neglected. The ideal gas approximation (see above) is in this study used to estimate rotational entropies as suggested by Jana *et al.* [92].

3 Hydration studies of carbohydrates

3.1 Introduction

An important goal of drug design is to understand and control cellular functions, such as regulation, signal transduction and intermolecular transport processes. These processes are controlled, within the machinery of the cell, by specific reversible binding of ligands to biomolecular receptors. The specificity and affinity of these molecular association processes can, to some extent, be explained by pure receptor-ligand interactions, e.g. complementary geometries of ligand and receptor molecular surfaces at the binding interface and local interactions between the receptor and the ligand in the complexed state. Cellular receptor-ligand interactions take place in an aqueous ionic environment and it is by now recognized that both water molecules and ions can play an important role in these processes. Water molecules can act as extensions of the molecular structures and thereby further enhance the specificity and affinity of the association process [95–98]. This is evidenced by the presence of water molecules at the binding interface of many crystal structures of biomolecular complexes. Furthermore, the binding process is accompanied by partial desolvation of both receptor and ligand molecular surfaces, where water molecules are transferred from the vicinity of the solutes to bulk water environment. Microscopic energetic changes, when solute-water interactions are replaced by water-water interactions, are reflected in macroscopic thermodynamics and kinetics of the binding process. The contribution from the desolvation free energy change to the overall binding affinity strongly depends on the chemical nature of the solute and is difficult to predict without detailed information of microscopic solute-water interactions.

Previous studies of hydration of various mono- and disaccharides have shown that both structure [25, 27, 99–101] and dynamics [26, 102] of carbohydrate hydration water is perturbed compared to bulk water. It is not clear, however, how changes in water properties affect binding affinities of carbohydrate-lectin or carbohydrate-carbohydrate complexes. This is particularly interesting since relatively low affinity characterize both binding of lectins to carbohydrates, or carbohydrates to other carbohydrates [4, 103]. The presence of hydrophobic patches on the molecular surfaces of many biologically important carbohydrates, and stacking of carbohydrates on aromatic side-chains in saccharide-lectin complexes, has led to speculations about the role of hydrophobic hydration and solvent reorganizing effects in molecular association processes involving saccharides [25, 104, 105]. Desolvation of water around hydrophobic patches of the saccharide would, according to the hypothesis of Frank and Evans [106], give a free energy change contributing favorably to complexation. Polar and charged solutes are readily soluble in water due to Coulomb solute-water interactions, which compensate the enthalpy loss at the molecular surface due to disruption of the water-water hydrogen bond network. Non-polar and non-charged solutes, on the other hand, only interact with water weakly through favorable solute-water van der Waals interactions [107], which only partly compensates the disruption of the water-water hydrogen bond network. Further energetic compensation comes from structural ordering of the hydrogen bond network in the vicinity of hydrophobic surfaces, where water-water hydrogen bonds on average are more linear, i.e. more energetically favorable compared to bulk water hydrogen bonds [108]. In order to compensate for the lack of solute hydrogen bond donors and acceptors, hydration water forms hydrogen bonded cages around the hydrophobic solute. Clathrate-like arrangements of the water hydrogen bond network have been found around hydrophobic patches on protein [109, 110] and DNA [111] molecular surfaces. The ordering of water structure and the accompanying entropy loss is, according to Frank and Evans, the reason for the low solubility and spontaneous aggregation of hydrophobic solutes in aqueous environment. Similar arguments have been used to explain driving forces in protein folding and the stability of tertiary protein structure [112–114].

In this study we have moved beyond simple mono- and disaccharides and investigate hydration properties of larger oligosaccharide systems, modeling the cell surface glycocalyx. We investigate structural, dynamic and thermodynamic properties of hydration water to answer the question of how solvent reorganization effects influence binding affinities of molecular association processes involving carbohydrates.

3.2 Water models

Two different water force fields have been used to model water in this study. The TIP3P [44] and SPC/E [45] water models are both, so called, minimalistic transferable interaction potentials, with only three interaction centers located on the nuclei. Partial charges are assigned to each nucleus and an additional Lennard-Jones potential, located at water oxygen atoms, accounts for intermolecular van der Waals forces (see Figure 3.1 and Table 3.1). Both models are parameterized as rigid body models, i.e. with a fixed geometry where H-O-H angle bending motions and O-H stretching motions are neglected. In the parameterization process, partial charges and Lennard-Jones coefficients are optimized to reproduce important properties of liquid water, such as, density, radial distribution functions and heats of vaporization. It should be noted that although these models are based on strict pairwise interactions between water molecules, they are *effective* pair-wise interaction potentials, that is, many-body effects are incorporated in an average way in the parameterization process. The cooperative strengthening of the hydrogen bond network due to polarization effects is important for describing liquid water properties [115] and average induced dipole moments of water molecules are incorporated in these models by introducing larger permanent dipole moments than the experimentally observed ones for isolated water molecules [45].

Table 3.1: Water force field parameters

	$\sigma(\text{\AA})$ ^a	ϵ (kJ mol ⁻¹) ^a	q_1 (e) ^b	q_2 (e) ^c	l (Å) ^d	θ° ^e
TIP3P [44]	3.15061	0.6364	+0.4170	-0.8340	0.9572	104.52
SPC/E [45]	3.166	0.650	+0.4238	-0.8476	1.0000	109.47

^a van der Waals parameters for water oxygen.

^b Partial charge for water oxygen.

^c Partial charges for water hydrogens.

^d O-H bond length.

^e H-O-H bond angle.

The success of these water models in modeling condensed phase water has been reviewed in several previous papers [116–118]. These comparisons show that each model has its strengths and weaknesses when it comes to modeling energetics, structural and dynamic properties of water. Since calculated properties depend on the simulation protocol, e.g. cutoffs, treatment of long range electrostatic interactions, thermostats

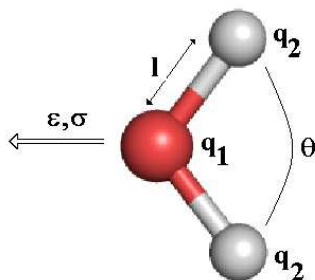


Figure 3.1: Force field parameters for the three-point potentials TIP3P and SPC/E

and barostats, both water models have been tested for pure water systems with the same simulation parameters used for simulations of saccharide systems presented in later chapters. This serves as a force fields validation and further provides a pure bulk water reference of various water properties as modeled by the two force fields. Pure water systems consisting of 8004 TIP3P or SPC/E water molecules are here modeled in a simulation box of initial dimensions of $63.54 \times 63.31 \times 63.40 \text{ \AA}$. The simulations were performed in the NPT-ensemble at 300 K and a pressure of 1 atm. Further details about the simulation protocol can be found in Chapter 3.3. Force field parameters for the two water models are presented in Table 3.1.

Various water properties calculated from these simulations, together with corresponding experimental values, are presented in Table 3.2. The most notable difference between the two force fields is the high diffusivity of TIP3P water. This artefact has been noticed before and is believed to be related to accounting for long-range electrostatic interaction in the periodic system, as opposed to a 9 \AA cutoff of both van der Waals and electrostatic interactions used in the derivation of this force field [122]. Although the potential energy per TIP3P water molecule is in fairly good agreement with experiment, the density of water at a pressure of 1 atm is too low. The SPC/E force field, on the other hand, models density and diffusivity in better agreement with experiment, but overestimates the average interaction energy per water molecule.

The oxygen-oxygen radial distribution functions in Figure 3.3a show structural differences in the modeled water hydrogen bond network of the two force fields. SPC/E water is in fairly good agreement with radial distribution functions extracted from neu-

Table 3.2: Properties of the TIP3P and SPC/E water models from pure water NPT-simulations at 300 K and 1 atm.

	E_p (kJ mol ⁻¹) ^a	ρ (g cm ⁻³) ^b	V (Å ³) ^c	D (*10 ⁻⁹ m ² s ⁻¹) ^d	q ^e	w_{hb} (ps) ^f
TIP3P	-39.8 (0.04)	0.977	30.6	6.1	0.56	1.06
SPC/E	-46.5 (0.01)	0.992	30.2	2.5	0.63	2.14
Exp ^g	-41.5	0.997	29.9	2.3	-	~ 1.0

^a Average potential energy per water molecule. Standard deviation in parenthesis. Experimental value from [45].

^b Average water density. Experimental value from [119]

^c Average volume per water molecule in the simulation box.

^d Average self-diffusion coefficient calculated for the mean squared displacements in figure 3.2 using the Einstein relation [56] $\lim_{t \rightarrow \infty} \langle |\mathbf{r}(t' + t) - \mathbf{r}(t')|^2 \rangle = 6Dt$. Experimental value from [120].

^e Average structural order parameter q (see Fig. 3.3).

^f Average hydrogen bond life-time. Experimental value from [121].

^g Experimental values at 25° and 1 atm.

tron scattering experiments [123], whereas the TIP3P force field fails to model structure beyond the first peak. The non-existent second and third peaks of the TIP3P radial distribution function, which correspond to second and third neighbors in the tetrahedral hydrogen bond network, show that TIP3P models the structure of water too weakly. The local order of the hydrogen bond network around any given water molecule is further investigated with the tetrahedral order parameter, q , of Chau and Harddwik [72] (see Chapter 2.2) for the two water models in Figure 3.3b. A clear shift towards lower q -values can be seen for TIP3P water. This means that the hydrogen bond network modeled by the TIP3P force field is, on average, less tetrahedral, or more destructured, compared to SPC/E water. Although there is no equivalent experimental measurement of the local order of water, it is interesting to note that the difference in local structure of the two water models is larger than that of SPC/E water at 0°C and 30°C [26]. The first peak of the radial distribution function gives information about the strength and average length of water-water hydrogen bonds. The height of the first peak of the water oxygen-oxygen radial distribution function is from X-ray and neutron scattering experiments determined to be in the range of 2.2-2.5 [124]. Here, TIP3P water is in better agreement with experiment. The structural differences of the two water models are also reflected in the hydrogen bond lifetimes (w_{hb} in Table 3.2). The average hydrogen bond life times are calculated from hydrogen bond autocorrelation functions (see Eq. 2.30) fitted to Kohlrausch-Williams-Watts functions, which are then integrated

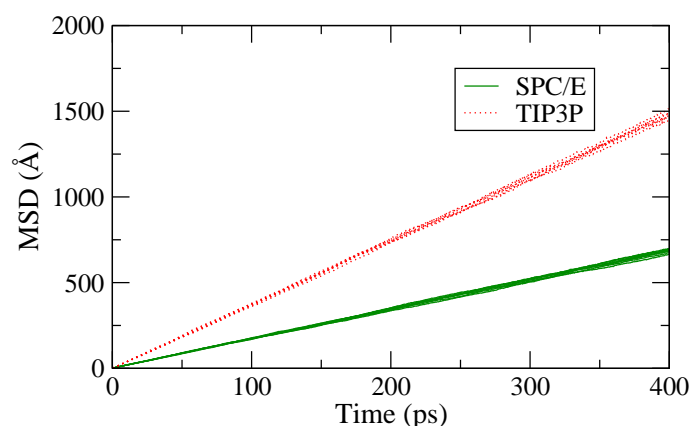


Figure 3.2: Average mean square displacements of water for 10 separate 400 ps blocks of the trajectory for the TIP3P and SPC/E water models.

from zero to infinity. The calculated average hydrogen bond life times depend on the criteria used to define a hydrogen bond, but the relative difference between the two water force fields show that the SPC/E force field models a more stable hydrogen bond network, with on average two times longer lived hydrogen bonds compared to TIP3P water. Using the hydrogen bond criteria defined in Chapter 2.2.2, calculated TIP3P hydrogen bond life times are closer to the experimental value of approximately 1 ps. These results show the difficulty of modeling both structural and dynamic properties of bulk water with simple three-point-potentials. More elaborate water models have been developed in the TIP4P [125] and TIP5P [126] water force fields and modifications thereof [127, 128]. These force fields utilize more interactions centers and model many liquid water properties in better agreement with experimental data. The effective polarization in all pair-wise potentials is a potential problem when modeling solute-water interactions. Dipole moments of hydration water of a given solute strongly depend on the nature of the solute and are likely to differ from bulk water dipole moments. Only polarizable water models can capture these effects. Although several polarizable water models are available, commonly used biomolecular force fields are developed together with the simpler three-point potentials. To be strict, these water models should be used together with the biomolecular force fields to have a consistent set of parameters for the system. Furthermore, several molecular dynamics programs only support the three-point potential water models. There are, however, several studies showing that the simple water models, together with the common biomolecular force fields, can

model hydration water properties in good agreement with experiment. Merzel and Smith [129] showed that distribution functions of TIP3P water around lysosome, modeled by CHARMM force fields, compare well with the scattering profile measured by Burling et al. [130]. Another comparison between molecular dynamics simulations and scattering experiments can be found in the study of Mason et al. [131]. This study concluded that the TIP3P force field models hydration water properties well enough to be used for predicting structural properties of water that cannot be probed experimentally.

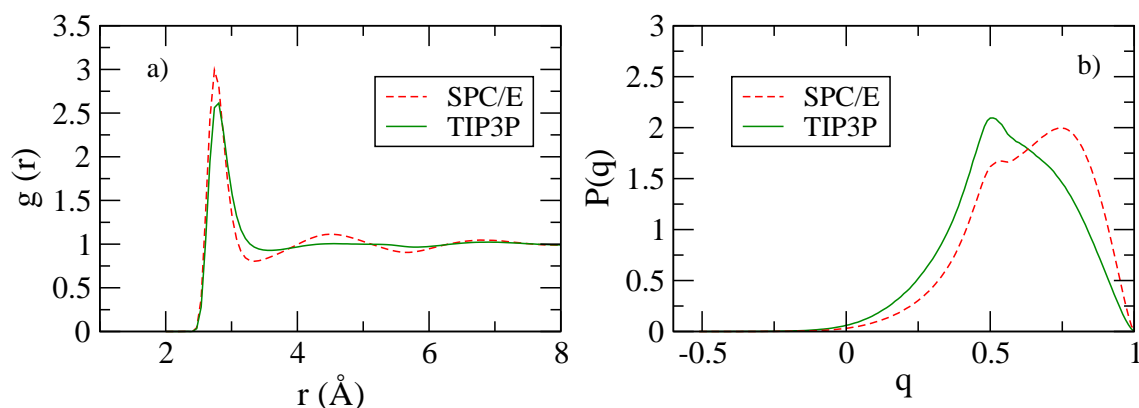


Figure 3.3: a) Oxygen-oxygen radial distribution functions for TIP3P (full line) and SPC/E (dashed line) water. b) Probability distributions of the tetrahedral order parameter q , for TIP3P (full line) and SPC/E (dashed line) water.

The GLYCAM force field, used to model carbohydrates in our simulations, is parameterized together with the TIP3P water model. Therefore, TIP3P is used as our primary water model in the next chapters. In order to test the influence of the water model on our results, reference simulations using the SPC/E water model are presented in Chapter 3.6.

3.3 Systems and simulation protocols

The results presented in the next section are calculated from classical-mechanical molecular dynamics simulations of the high-mannose type oligosaccharide, presented in Figure 3.5, with explicit water modeled by the GLYCAM04 and TIP3P force fields respectively. The simulation box, shown in Figure 3.4, consists of the oligosaccharide together

with a water box of initial dimensions 53.76 x 64.74 x 56.41 Å containing 5011 water molecules. The oligosaccharide is restrained to the x-y plane by means of positional restraints on the main-chain asparagine atoms in the *N*-link, thus allowing for no diffusional motion of the saccharide. With its periodic images, this system can be seen as a sparse saccharide array along the x-y plane with a distance of approximately 50 Å between a saccharide and its closest neighbor in the periodic system. The z-dimension of the box is extended another 10 Å to model bulk properties above and below the sugar array. Standard molecular dynamics simulations were performed using the NAMD 2.6 molecular dynamics program [36] in the NPT-ensemble. A constant temperature of 300 K was applied using a Berendsen thermostat [51] with a weak coupling coefficient of 5 ps. A constant pressure of 1.01325 bar was applied using a Berendsen pressure bath coupling with a rescaling frequency of 8 fs, a barostat relaxation time of 100 fs and a compressibility of $4.57 \times 10^{-5} \text{ bar}^{-1}$. A cutoff of 12 Å is used for van der Waals interactions, without using any switching functions. The simulation was performed using the PME method [132], with a grid spacing of approximately 1 Å in each dimension of the simulation box. 1–4 bonded electrostatic and vdW scaling was set to unity in accordance with the GLYCAM04 force field [133]. The SHAKE algorithm [55] was used to constrain heavy-atom hydrogen bonds, allowing for an integration time-step of 2 fs. A time-step of 4fs was used for long range electrostatic interactions. The results presented in the next section are calculated from a 8.8 ns trajectory. Snapshots of the trajectory were stored every 0.4 ps, giving a total of approximately 22000 snapshots over which the presented results are averaged.

3.4 Hydration of the N-linked high mannose-type oligosaccharide

3.4.1 Structural properties of saccharide hydration water

In this section, we investigate structural perturbations of the water hydrogen bond network due to the presence of the saccharide. Figure 3.6a shows the proximal distribution function of water oxygen atoms around the saccharide. The distribution function is normalized to bulk water density, using the method described in Chapter 2.2.1 to calcu-

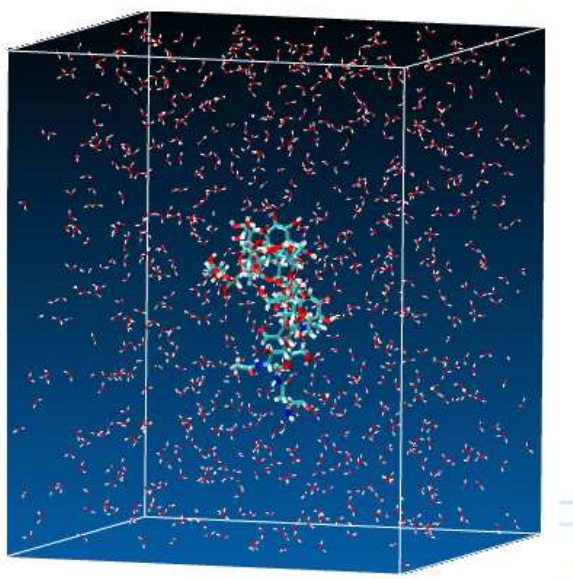


Figure 3.4: Simulation box of high mannose-type oligosaccharide-water system. Water density is reduced in the figure for a clearer representation

late the available volume around the saccharide at a given distance. Here, the distance refers to the distance between the water oxygen atom and the closest non-hydrogen saccharide atom. The first peak of the proximal distribution function in Figure 3.6a corresponds to the first peak of the water oxygen-oxygen radial distribution function in Figure 3.3a, and represents water molecules hydrogen bonded to the saccharide. A second smaller and wider peak can also be seen, followed by small oscillations in density up to a distance of approximately 7.5 \AA , where no further perturbations of the water structure can be seen. The origin of the second peak may not be clear at first glance, but a decomposition of the distribution function with respect to the atom type of the closest saccharide atom clarifies the picture. The distribution functions in 3.6b show the normalized distributions of water molecules around polar (oxygen and nitrogen atoms) or non-polar (carbon atoms), just taking the volume available for water molecules at a given distance to the two saccharide atom types into account. From Figure 3.6b it is clear that the second peak in the full proximal distribution function can be attributed to water molecules around apolar parts of the saccharide. Comparing with the pure water oxygen-oxygen radial distribution function in Figure 3.3a, structural perturbations beyond the first peak also stem from water around hydrophobic parts of the saccharide. The distribution function of water around the saccharide carbons shows a

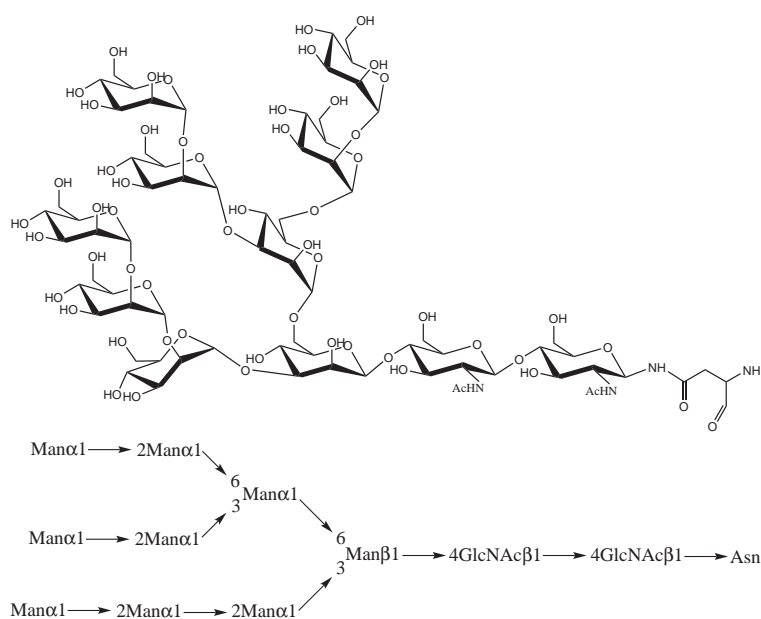


Figure 3.5: Pictorial description of the N-linked high mannose-type oligosaccharide used in the simulations.

hydrophobic hydration pattern, very similar to that of water around n -C₁₈ alkyl chains [62] and water on a graphite surface [134]. Thus, these distribution functions reveal a clear amphiphilic character of the saccharide. The majority of water molecules in the vicinity of the saccharide surface will experience the polar character of the oxygen and nitrogen atoms. A smaller subset of hydration water molecules will be around non-polar patches of the saccharide, giving rise to local pure hydrophobic hydration patterns. Using a Voronoi decomposition (see Section 2.2.4) of the simulation box, first hydration shell water molecules can be defined as water molecules sharing a common face of their Voronoi polyhedra with any polyhedron assigned to a saccharide atom. An average over a 4 ns time period of the trajectory gives 229 first hydration shell water molecules. Out of these are 184 water molecules nearest neighbors with polar atoms of the saccharide and 45 nearest neighbors with carbon atoms. Thus, approximately 20% of first hydration shell water experience a more or less hydrophobic environment. Many biologically important monosaccharides such as glucose, mannose and galactose have distinct top/down electrostatic character of the molecular surface of the sugar ring. The top side has a multipolar character whereas the bottom side of the ring is purely hydrophobic [99]. As mentioned in the introduction, unfavorable hydration of the hydrophobic parts of the saccharide has previously been hypothesized to make an

important contribution to lectin-carbohydrate recognition [99, 135].

The decomposition of the water oxygen distribution function in Figure 3.6b also reveals favorable hydrogen bonding of water to the saccharide. In comparison with the water oxygen-oxygen radial distribution function in Figure 3.3a, the saccharide hydroxyl oxygen-water oxygen distribution function shows a much higher first peak. This means that there is a higher probability of finding a water molecule in a given volume element around saccharide hydroxyl oxygens compared to the same volume around a given water molecule in bulk water. This is an indication of strengthening of water-saccharide hydrogen bonds. This subject will be discussed further later in Chapter 3.4.2.

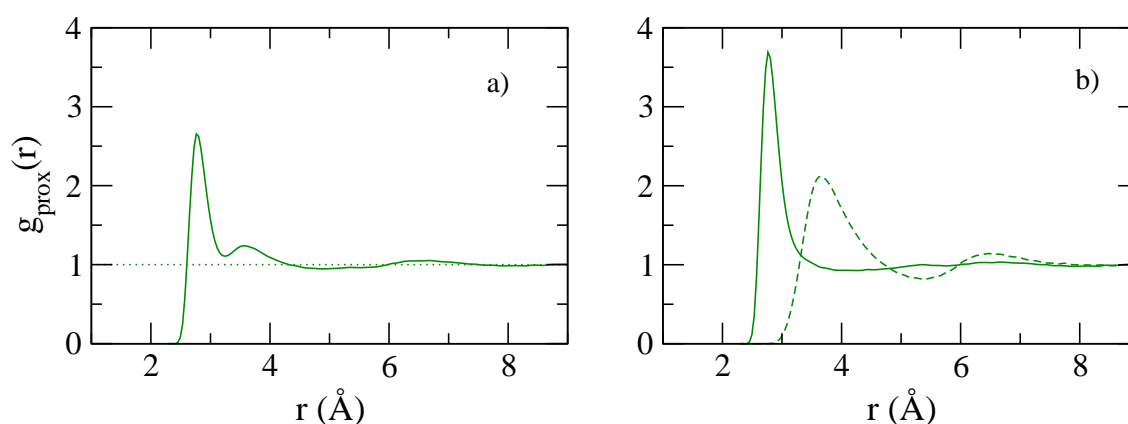


Figure 3.6: Proximal distribution functions of water oxygen atoms around the high mannose-type oligosaccharide. In Figure (a) the distribution of water oxygen atoms is calculated with respect to the closest non-hydrogen atom in the saccharide. Figure (b) shows the distribution function decomposed with respect to the closest atom type. The full line is the distribution around polar (O and N) saccharide atoms and dashed lines the distribution around apolar (C) saccharide atoms. The number of water oxygen atoms is collected in 0.1 Å bins and the average volume of each 0.1 Å shell around the saccharide is calculated as described in Section 2.2.1.

The effect of the saccharide on the geometrical arrangement of the hydrogen bond network of hydration water is measured by the tetrahedral order parameter, q , in Figure 3.7. For a perfect tetrahedral arrangement of a given water molecule and its four closest hydrogen bond donors/acceptors, which can be both water and saccharide atoms, q is unity. For a random arrangement of the five atoms, q takes a value of zero. For further details about the calculation of the order parameter see Section 2.2.5. Distributions of q -values for water molecules with an oxygen atom within a distance of 4

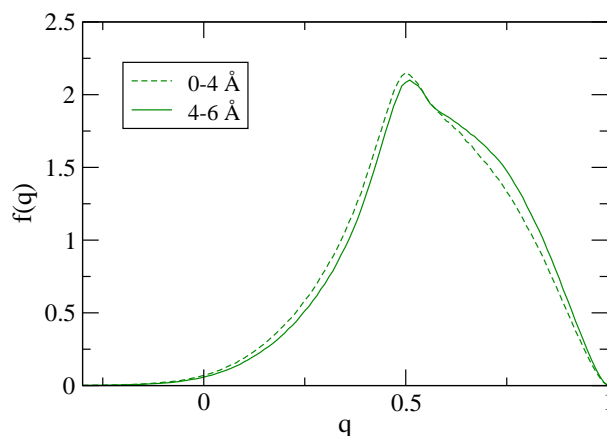


Figure 3.7: Distributions of the tetrahedral order parameter q of water molecules in different shells around the saccharide.

\AA , approximately accounting for water molecules in the first hydration shell, and for water located at a distance between 4 and 6 \AA are shown. The distribution of the order parameter for water molecules beyond a distance of 4 \AA is the same as for bulk water (see Figure 3.3b). The presence of the saccharide causes a slight perturbation of the geometrical arrangement of hydration water in direct contact with the saccharide, as seen in a slight shift towards lower q -values in Figure 3.7. Beyond the first hydration shell, the hydrogen bond network has geometrical properties the same as in bulk. Thus, the saccharide seems to have very modest effects on structural properties of the surrounding water. A shift towards lower q -values for water in direct contact with the saccharide is expected since, as previously shown, a substantial part of hydration water will be nearest neighbors to non-polar parts of the saccharide, which naturally will have lower q -values due to the lack of hydrogen bonding partners in the saccharide.

Further investigations of structural perturbations of hydration water were performed by directly analyzing hydrogen bond properties of hydration water. Hydrogen bonds can be calculated from computer simulations using the geometric criteria discussed in Section 2.2.2. The structure of the hydrogen bond network is measured here by the ratio of the number of hydrogen bonds formed by each water molecule, n_{hb} , and the coordination number, n_c , of the same water molecule calculated with a cutoff distance corresponding to that used in the hydrogen bond definition. A higher value compared to that of bulk water indicates a structuring of the water hydrogen bond network, such as induced by a temperature decrease, whereas a lower value shows a destructured

hydrogen bond network, which may result from either a reduced number of hydrogen bonds or an increased coordination number [136]. Figure 3.8a shows the average n_{hb}/n_c ratio, calculated for both water-water and water-saccharide hydrogen bonds, for water molecules at a given distance to the closest non-hydrogen saccharide atom. The most

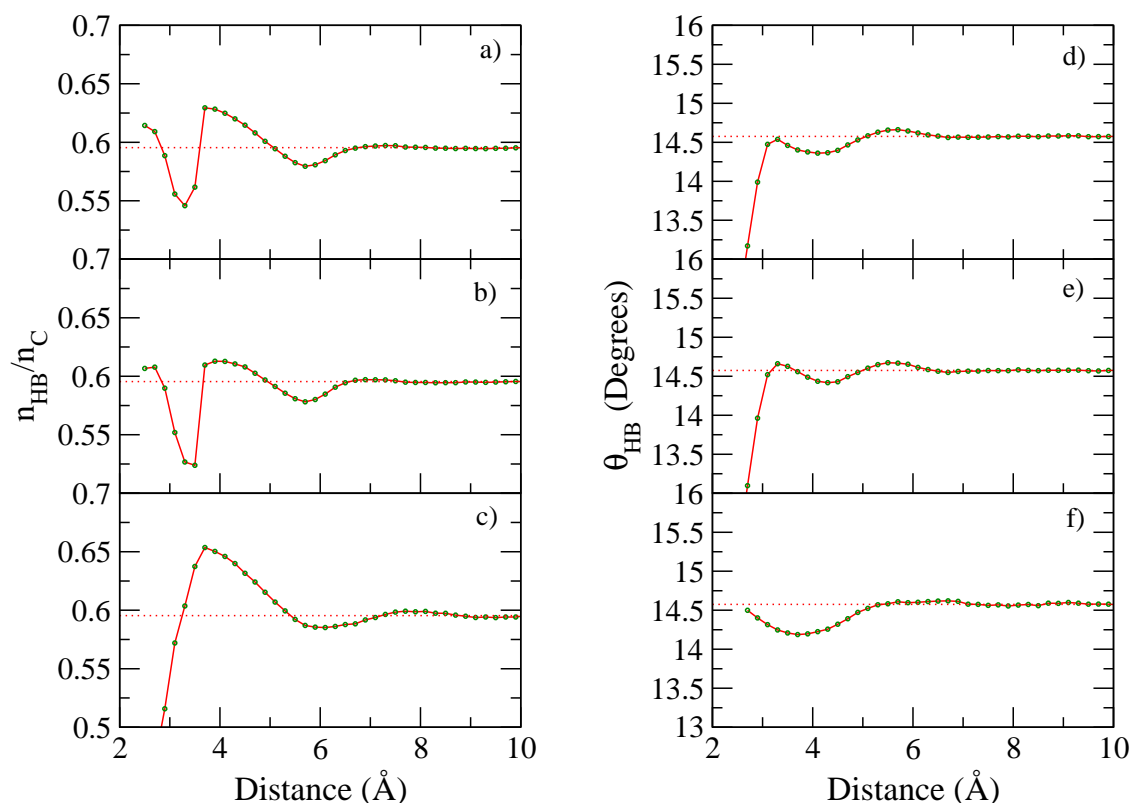


Figure 3.8: Structural properties of hydrogen bonds formed by water molecules, at a given distance r , to the closest non-hydrogen atom in the saccharide. Both water-water and water-saccharide hydrogen bonds are considered. Figures a)-c) show the average ratio of the number of hydrogen bonds formed by a water molecule to the coordination number of this water molecule. Figures d)-e) show the average hydrogen bond angle θ (see Figure 2.2) at distance r . In Figures a) and d) c), all non-hydrogen atoms in the saccharide are considered in the distance calculation. In Figures b) and e) and c) and f) properties of water hydrogen bonds are calculated for subsets of water molecules at distance r , being closest to saccharide hydroxyl oxygens and saccharide carbons respectively.

likely position of a water molecule, hydrogen bonded to the saccharide is, as seen in Figure 3.6, at a distance of approximately 2.9 Å to the closest non-hydrogen saccharide atom. Figure 3.8a shows that these water molecules form a relatively structured hydrogen bond network with hydrogen bonding partners in both the saccharide and other water molecules. A second peak of higher structure of the hydrogen bond network

can also be seen at a distance corresponding to the second peak in the experimental water-water oxygen radial distribution function. Hence, both first hydration and second hydration shell water molecules show an enhanced hydrogen bond structure. Water moving between the two first hydration shells, or from and to the second hydration shell, are less structured than bulk water. A detailed analysis (not shown here) showed that these water molecules, in between the hydration shells, on average both have higher coordination numbers and a lower number of hydrogen bonds. Figures 3.8b and 3.8c show the same hydrogen bond to coordination number ratio as in Figure 3.8a, but here partitioned based on the type of the closest saccharide atom (hydroxyl oxygens in Fig. 3.8c and carbons in Fig. 3.8b). Again, it is clear that the overall structure of hydration water is dominated by water around saccharide oxygens. However, hydration water molecules around apolar parts of the saccharide display the largest structural enhancements of the hydrogen bond network in terms of increased hydrogen bonding. A similar increase in the number of hydrogen bonds per water molecule has previously been observed for a water droplet on graphite [134].

Average water hydrogen bond angles, as a function of the distance to closest non-hydrogen saccharide atom, are shown in Figures 3.8d-f. Average hydrogen bond angles, at a given distance around the whole saccharide, are shown in Figure 3.8d. Figures 3.8e and 3.8f show average hydrogen bond angles decomposed into the closest saccharide atom type. The peaks in Figures 3.8a-c can be seen to correspond to minima in Figures 3.8d-f. These figures further highlight the increased structure of the hydrogen bond network, seen as increased linearity of hydrogen bond angles of water molecules in the first and second hydration shells. First and second hydration shell water molecules do also, on average, have shorter hydrogen bond lengths compared to bulk water (not shown here). This results in smaller volumes per water molecule and higher densities of both first and second hydration shell water (see Table 3.4.3). A similar increase in water density of hydration water around the protein lysozyme [129].

The average interaction energy per water molecule, as a function of the distance to closest non-hydrogen saccharide atom, is shown in Figure 3.9. Here the interaction energy is calculated from all atoms within a sphere with a radius of 12 Å. Electrostatic interactions beyond this cutoff radius are accounted for using a reaction field correction [137]. The structuring/destructuring hydration patterns shown in the previous two figures are clearly reflected in the potential energy profile of water around the saccharide. Water molecules hydrogen bonded to the saccharide, second hydration shell water molecules and water around the apolar parts of the saccharide all display

more favorable interaction energies compared to bulk water. Energetic perturbations of hydration water due to the presence of the saccharide can be seen up to a distance of approximately 8 Å.

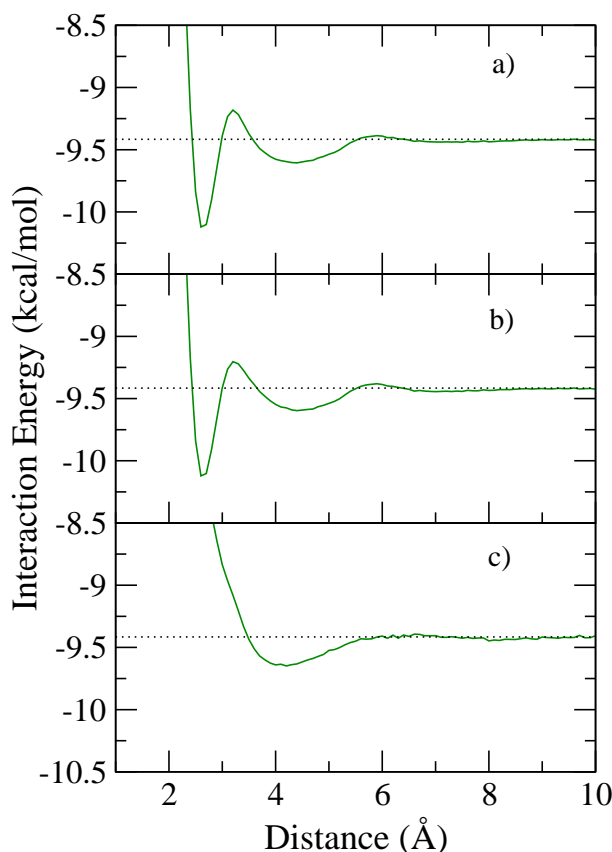


Figure 3.9: Figure a) shows the average interaction energy per water molecule at distance r to the closest non-hydrogen atom in the saccharide. In figures b) and c), the average interaction energy per water molecule is calculated for subsets of water molecules at distance r , being closest to saccharide hydroxyl oxygens and saccharide carbons respectively.

3.4.2 Dynamic properties of saccharide hydration water

The investigations in the previous section have shown that the majority of the first hydration shell water molecules are in contact with polar moieties of the saccharide, replacing water-water hydrogen bonds with saccharide-water hydrogen bonds. It was also shown that these water molecules are energetically more favorable compared to bulk water molecules. This is also reflected in the hydrogen bond lifetime autocorrelation

functions in Figure 3.10. Here, autocorrelation functions are shown for saccharide-water donor, saccharide-water acceptor and water-water bulk hydrogen bonds, calculated as described in Section 2.2.3. Figure 3.10a shows that lifetimes of saccharide-water hydrogen bonds, on average, are longer than corresponding bulk water-water hydrogen bonds. Figure 3.10b reveals a difference between the two different kinds of hydrogen bonds a water molecule can form with polar saccharide atoms. On average, saccharide donor hydrogen bonds are longer lived compared to saccharide acceptor hydrogen bonds. Integration of the hydrogen bond autocorrelation functions in Figure 3.10 gives mean continuous hydrogen bond lifetimes of 2.25 and 1.41 ps for saccharide donor-water and saccharide acceptor-water hydrogen bonds respectively. Calculated average bulk hydrogen bond lifetime of the TIP3P water model is, for comparison, 1.06 ps (see Table 3.2). The difference in strength between donor and acceptor hydrogen bonds is a consequence of the higher dipole moment of water compared to saccharide oxygens, reflected in the force fields as more negative charges of water oxygens compared to saccharide hydroxyl oxygens.

In order to investigate the dynamics of the complete first hydration shell, including hydration water around non-polar parts of the saccharide, we have calculated residence times of all water molecules within a distance of 3.6 \AA of the saccharide. The ARC/TAP approach of Henchmann and McCammon [138], originally developed to identify hydration sites in proteins, is here extended to carbohydrates. This method addresses the problem of calculating absolute positions of water relative to a flexible macromolecule, which is especially relevant for saccharides. The position of a water molecule in the vicinity of the saccharide is calculated in local coordinate systems assigned to each monosaccharide ring of the saccharide. Enumerating the saccharide ring carbon atoms, using index 1 for the anomeric carbon, the xy-plane of the local coordinate system is spanned by carbon atoms 1,2,4 and 5 in each saccharide ring, and the z axis is perpendicular to this plane using the right hand rule. An analysis of sugar ring conformations using the puckering parameter of Cremer and Pople [139] showed that the rings are stable and only showing small oscillations around a 4C_1 chair conformation over time. A local coordinate system, based on the sugar ring structure, should therefore provide sufficient accuracy for our purposes. The residence time of a water molecule is calculated as the time it stays within a distance of 2.8 \AA from its initial position. This distance approximately corresponds to the diameter of a water molecule. When a water molecule has moved beyond this distance, it is considered to have left a given hydration site and the residence time is binned. The position of a given water molecule

is averaged over positions in the local coordinate systems of saccharide rings with any atom within a distance of 3.6 \AA , weighted by the number of times the water is within this distance to a given residue from the time origin. In this manner, water in between saccharide residues are also accounted for. Figure 3.11 shows the distribution of residence times for water within 3.6 \AA from closest non-hydrogen saccharide atom, as well as the distribution of residence times for bulk water calculated in the global coordinate system using the same distance criterion. The dynamics of water in the proximity of

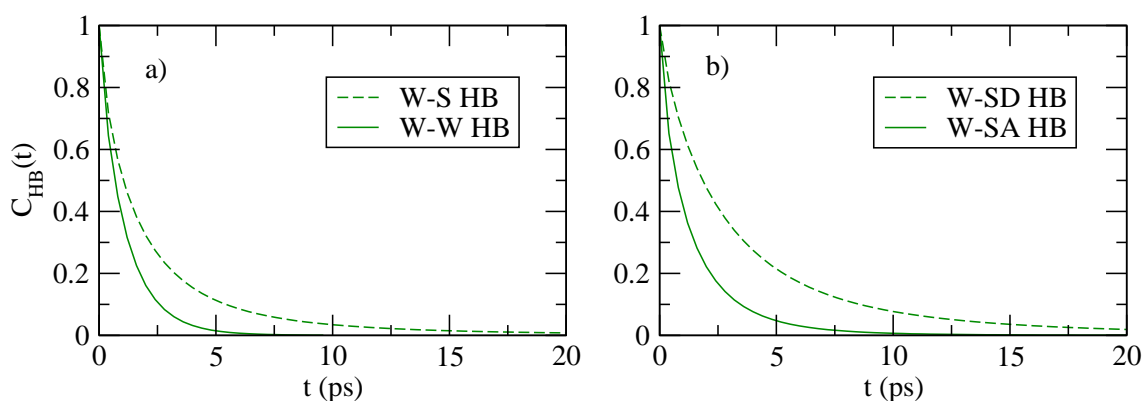


Figure 3.10: a) Continuous hydrogen bond correlation functions for water-saccharide (W-S) hydrogen bonds and for bulk water-water (W-W) hydrogen bonds. b) Continuous hydrogen bond correlation functions for water-saccharide donor hydrogen bonds (W-SD) and water saccharide acceptor (W-SA) hydrogen bonds.

the saccharide is similar to what has previously been found for both protein and DNA hydration water [140–142]. These studies showed that hydration water can be divided into two categories; water molecules tightly bound to the biomolecule, with greatly reduced rotational and translational dynamics, and the majority of hydration water with rotational and translational diffusion rates similar to bulk water. Figure 3.11 reveals a similar tendency of hydration water around the saccharide. The peak of bulk and hydration water residence times coincide, showing that the majority of first shell hydration water have bulk-like translational dynamics. There is however, a shift towards longer residence times for first hydration shell water. Several cases of residence times over 100 ps can be seen in this simulation, and the longest observed residence time of a water molecule in the hydration shell of the saccharide is 241 ps. A detailed analysis of the hydration shell showed that these long lived water molecules reside in the cavities of the molecular surface of the saccharide, being hydrogen bonded to the saccharide

and, at the same time, experiencing translational restraints from neighboring residues. The dynamics of water beyond the first hydration shell is further analyzed in Figure 3.12b, where local self-diffusion coefficients are plotted with respect to the distance to the closest non-hydrogen saccharide atom. The local diffusion coefficients are calculated from the slope of the mean square displacement of a water molecules with positions r at time t_0 , using two later time steps t_1 and t_2 , with fixed values of 1.2 and 2.4 ps respectively. Assuming that the diffusional regime has been reached after 1.2 ps, the diffusion coefficient can be calculated from the slope of the mean square displacement between times t_1 and t_2 [56, 143]. The distance in Figure 3.12b is that of time t_0 , which means that water molecules will have traversed some distance during the time the diffusion coefficient is measured. Thus, the local diffusion coefficients will be averaged over different parts of the hydration structure around the saccharide, and detailed information of water mobility at a given distance will be lost. The diffusivity of water molecules in the first hydration shell is greatly reduced compared to bulk water. Reduced translational mobility of water in second, third and even fourth hydration shells can also be seen in Figure 3.12b. Molecular dynamics studies have shown similar long range reduction of water diffusivity around DNA and various proteins [143]. The rotational

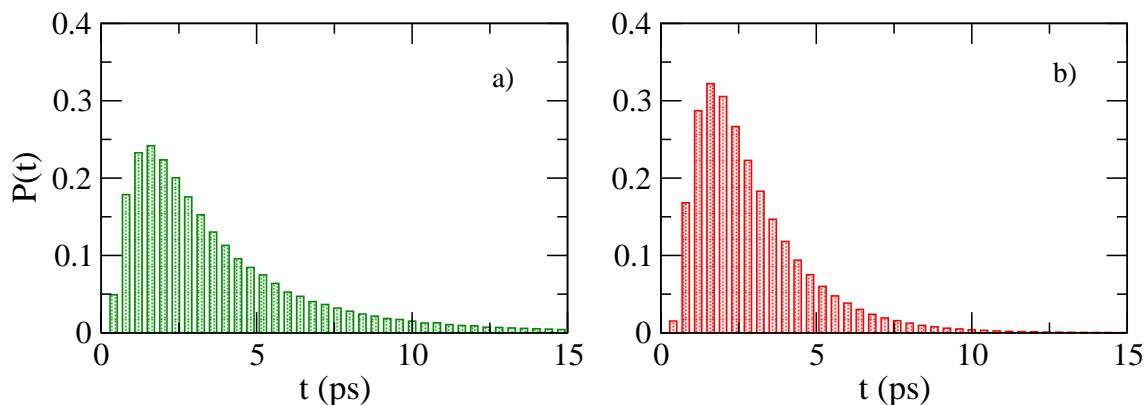


Figure 3.11: Probability distributions of residence times of water molecules calculated for hydration water in a local saccharide based coordinate system (a) and for bulk water in the global coordinate system (b).

dynamics of water around the saccharide is investigated by the orientational correlation function given by Equation 2.28. Figure 3.12a shows local second-order orientational relaxation times for water molecules in 1 Å thick shells around the saccharide. Only water molecules, continuously present in a given shell for 4 ps, are used in the calcula-

tion of the orientational correlation function. These correlation functions were fitted to stretched Kohlrausch-Williams-Watts exponentials [144, 145], and rotational relaxation times τ_R were obtained by numerically integrating the fitted functions from zero to infinity. It should be noted that the calculated orientational relaxation times of hydration water strongly depend on the time used to calculate the correlation function. Similar to the wide range of translational diffusion rates, shown in Figure 3.11, the rotational dynamics of hydration water molecules occurs on many different time scales. Longer sampling periods will increase the representation of saccharide bound water molecules and will therefore give longer rotational relaxation times. Using a short sampling period of 4 ps, contributions from water molecules with a wide range of different rotational rates are included in the calculated orientational relaxation times. Retardation of hydration water rotational dynamics is, as expected, most severe for

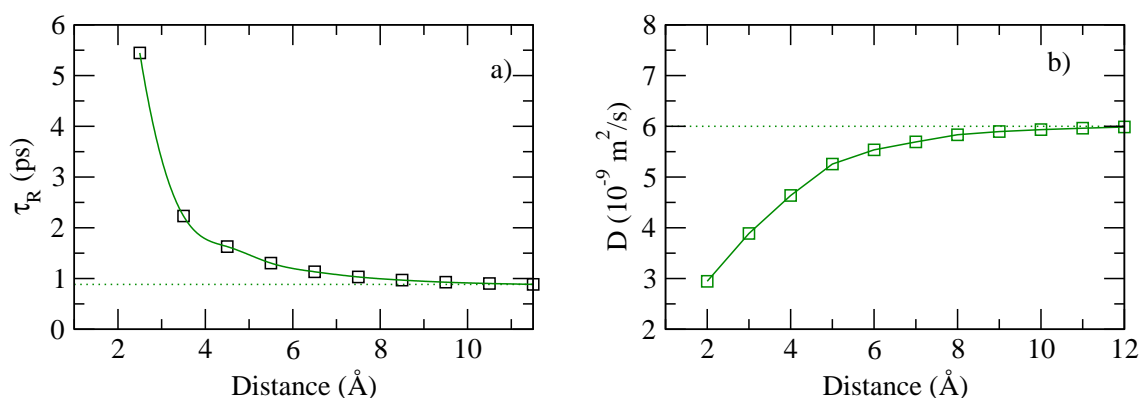


Figure 3.12: a) Second-order orientational relaxation times of water molecules as a function of the distance of the closest non-hydrogen saccharide atom. b) Local self-diffusion coefficients of water molecules in the saccharide hydration shell, plotted with respect to the distance to the closest non-hydrogen saccharide atom.

first hydration shell water, but is also observed in successive hydration shells. Comparing the structural modifications of the hydrogen bond network around the saccharide, in Figures 3.8-3.9, to Figures 3.12a and 3.12b, no clear correlation between structural and dynamic perturbations of hydration water can be seen. The immobilized saccharide acts as a molecular wall, reducing the dynamics of water, beyond distances where structural modifications of the water hydrogen bond network can be observed. This phenomenon can also be seen in the water-water hydrogen bond time correlation functions in Figure 3.13. These curves show the probability that a hydrogen bond, present

between two water molecules at a given time t_0 , has not been broken and is still intact at a later time t . An overall reduction in hydrogen bond dynamics can be seen up to a distance of 12 Å from the closest non-hydrogen saccharide atom. This distance extends far beyond the first and second hydration shells, where the structure and energetics of water differ from bulk water.

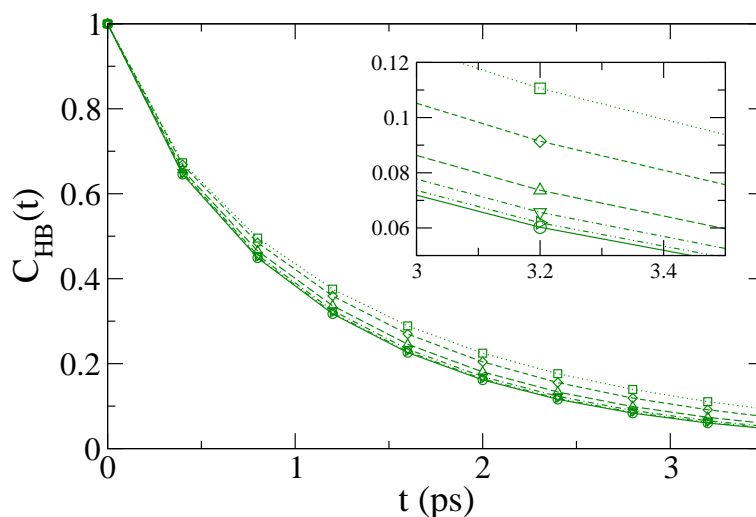


Figure 3.13: Continuous hydrogen bond lifetime autocorrelation functions for water in different regions around the N-linked high mannose-type oligosaccharide. Time correlation functions of water within 0-4 (squares), 4-6 (diamonds), 6-8 (up triangles), 8-10 (down triangles), 10-12 (right triangles) and 12-14 (circles) Å are shown. The correlation function for water in the 12-14 region is the same as for bulk water.

3.4.3 Thermodynamics of saccharide hydration water

In this section, we investigate how the previously discussed microscopical perturbations of the hydrogen bond network are reflected in thermodynamic properties of hydration water. This is interesting for understanding the driving forces behind lectin-carbohydrate or carbohydrate-carbohydrate association processes. Every non-covalent molecular association process, occurring in an aqueous environment, is accompanied by a reorganization of solvent molecules, where water molecules are transferred from the solute molecular surfaces to a bulk water environment. A pictorial description of this desolvation process is shown in Figure 3.14.

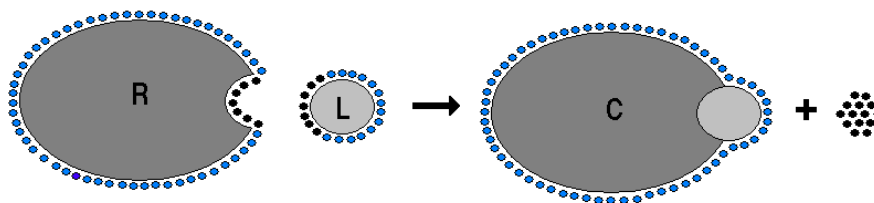


Figure 3.14: Pictorial description of a molecular association process, forming complex C from receptor R and ligand L. Solvent reorganization associated with the process involves a partial desolvation where water molecules (small sound circles) are transferred from the receptor and ligand molecular surfaces to bulk water.

It should be noted that this desolvation process, and the accompanying free energy change, is not the same as measurable solvation free energies, which refer to the free energy change for the process of transferring a solute from gas phase to water.

The results presented in the previous section showed that saccharide hydration water is both structurally and dynamically perturbed compared to bulk water. Structuring of the water hydrogen bond network could be seen for first and second hydration shells around the polar atoms of the saccharide, as well as, for water in the vicinity of apolar parts of the saccharide. These structural modifications of the hydrogen bond network result in, on average, more favorable interaction energies of hydration water molecules around the saccharide. Rotational and translational diffusion rates of hydration water were also shown to be much slower compared to bulk water. All together, these results indicate that removing hydration water from the high-mannose oligosaccharide is accompanied by a simultaneous increase in entropy and decrease in enthalpy of the system. This phenomenon is known as entropy-enthalpy compensation and is a common feature in many chemical reactions and processes in biological systems. It has been suggested that entropy-enthalpy compensation is a general feature of weak molecular associations [146] and would also apply to structural and dynamic differences between hydration water and bulk water. If the removal of hydration water results in a enthalpy loss that is fully compensated by a entropy gain, hydration water will be in thermodynamic equilibrium with bulk water and not contribute to the total free energy change. If either term is dominating, on the other hand, desolvation of the saccharide will either make a favorable or unfavorable free energy contribution to the binding process.

The 2PT method of Lin et al. [68] is used to calculate absolute entropies of water

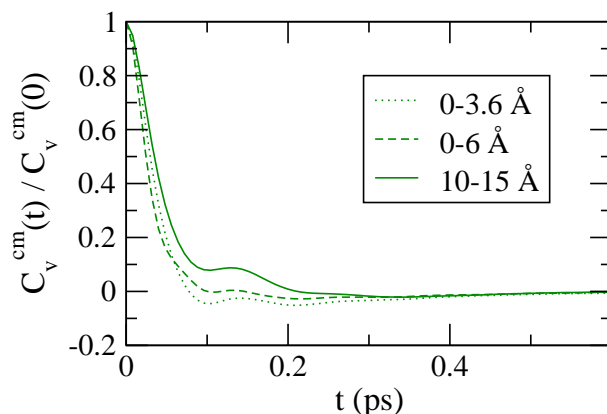


Figure 3.15: Normalized center of mass velocity autocorrelation functions for water molecules in three different domains around the N-linked high mannose-type oligosaccharide. Dotted, dashed and full lines show velocity autocorrelation functions averaged over water molecules within 3.6 Å, within 6 Å and beyond 10 Å from the closest non-hydrogen saccharide atom.

in different regions around the saccharide. The theory of this method is described in Section 2.2.8. In summary, the frequency distribution of translational motions of water molecules (density of states (DoS) function) is calculated from a Fourier transform of the water center of mass velocity autocorrelation function. These motions correspond to both high frequency vibrations of water in the hydrogen bond network as well as low frequency modes corresponding to slow diffusional motions of water molecules. Utilizing the fact that the DoS-function of a typical liquid takes the form of superposed solid-like and gas-like power spectra (see Figure 2.6), the intensity at a given frequency is decomposed into configurational (gas-like) and vibrational (solid-like) contributions. The contribution to the total translational entropy from vibrations within the hydrogen bond network is calculated using the analytical formula of the entropy of a quantum harmonic oscillator at a given frequency, weighted by the intensity of the decomposed solid-like density of states function. The configurational (gas-like) contribution is calculated in a similar way, using the analytical formula of the entropy of a hard-sphere fluid. The results presented here are calculated from a 300 ps trajectory in the NVT-ensemble (same simulation protocol as used before but without any pressure regulation). Velocity autocorrelation functions are calculated for a time period of approximately 8 ps, storing velocities every 8 fs, giving a resolution of the DoS-function of approximately 4 cm^{-1} . The results in the previous chapter showed that the properties of hydration

water around the saccharide is not uniform. The amphiphilic character of the saccharide results in different structural properties of hydration water in different parts of the saccharide surface. The dynamics of hydration water can also take many different forms, ranging from water with bulk diffusional rates to water molecules bound to the saccharide for hundreds of picoseconds. In these calculations, the velocity autocorrelation function of hydration water is averaged over water molecules in three different domains. Water within a distance of 3.6 Å, water within 6 Å and water within a distance 10 to 15 Å from the closest saccharide atom. The first domain corresponds to water molecules hydrogen bonded to the saccharide. The second domain accounts for the full first and second hydration shells, where the main structural and energetic perturbations of the hydrogen bond network can be seen (see Figure 3.8). Water beyond 10 Å is used to describe bulk water properties. The comparison of water properties in the two first domains can tell us how the non-polar parts of the saccharide influence the thermodynamic properties of hydration water. Figure 3.15 shows the calculated center of mass velocity autocorrelation functions for both hydration and bulk water, averaged over water molecules present in the three domains during the approximately 8 ps time interval for which the correlation function is calculated. The corresponding Fourier transforms of the center of mass velocity autocorrelations, i.e. density of states (DoS) functions, are shown in Figure 3.16. The area under each curve is here normalized to 3, giving the average frequency distribution from the three translational degrees of freedom per water molecule in each region. Properties of the bulk water DoS function has been characterized in previous studies [124, 147]. The intensity at $\nu=0$ is directly related to the self-diffusion coefficient of water molecules (see Eq. 2.77). The peak around 60 cm^{-1} has been attributed to the O-O-O bending mode between hydrogen bonded water molecules. Guillot *et al.* [124] showed that this peak is not specific to water and cannot be attributed to hydrogen bonded liquids. Rather, these low frequency modes are caused by rattling motions, due to caging effects present in all liquids. The shoulder around 200 cm^{-1} is generally accepted to correspond to the intermolecular O · · · H-O stretching motions.

The spectral densities of hydration water reveal a reduced intensity of low frequency modes, which, as previously discussed, can be attributed to diffusional motions in a liquid. An increase in intensity of modes above 100 cm^{-1} can also be seen. This shows that some of the loss of hydration water translational diffusivity is transferred into higher frequency vibrations due to interactions with the saccharide. The results of the two-phase decomposition of the DoS functions are presented in Table 3.3. The results

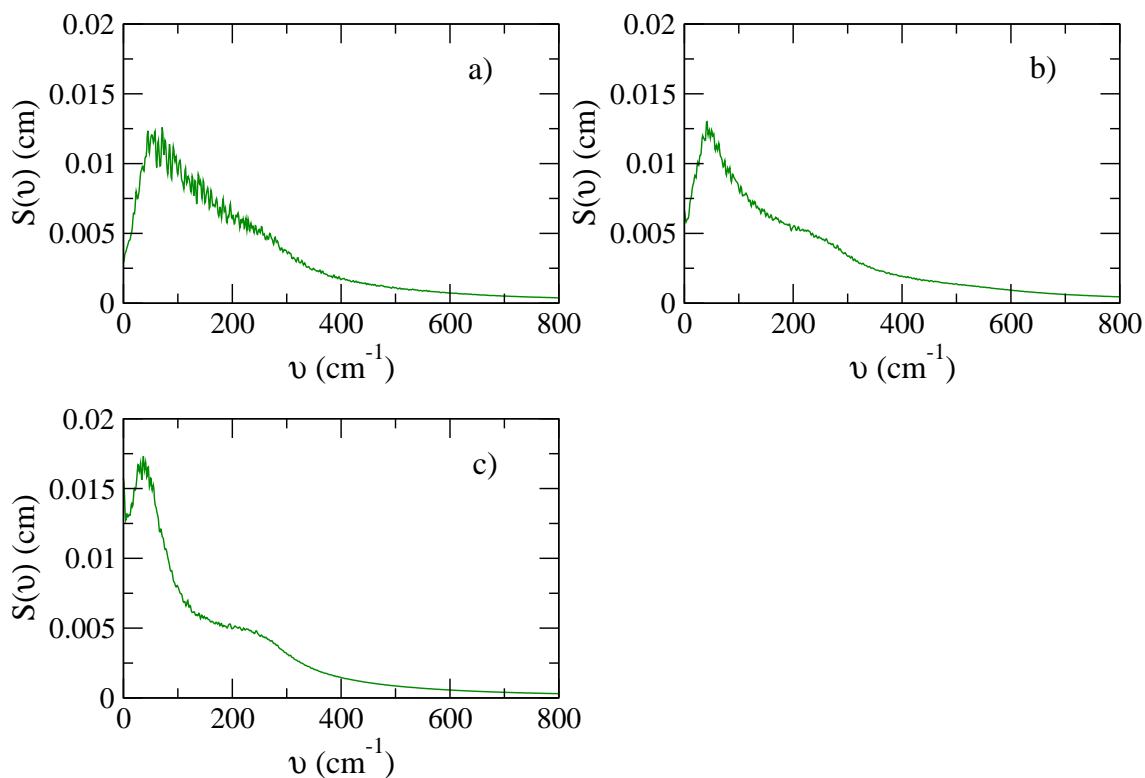


Figure 3.16: Translational density of states functions of TIP3P water in three different regions around the N-linked high-mannose oligosaccharide. a) within a distance of 3.6 Å, b) within a distance of 6 Å and c) between 10-15 Å from the closest non-hydrogen saccharide atom.

presented here are averaged over seven separate 40 ps segments of the NVT-trajectory. The perturbations in structure and dynamics of hydration water are reflected in the data from the 2PT-analysis. Diffusion coefficients, calculated from intensities of the DoS functions at $\nu=0$, are slightly lower for hydration water compared to the ones calculated from the average mean square displacements in Figure 3.12a. This can be explained by the fact that self-diffusion coefficients calculated from the velocity autocorrelation functions are averaged over water molecules constantly present in the domains over the sampling interval. The diffusion coefficients in Figure 3.12b also include dynamics of water molecules moving in and out of a given region. The fluidity factor determines the partition of the DoS function into solid-like and diffusive gas-like contributions. The fluidity factor depends on the diffusivity, temperature and density of the system and would take a value of zero for ice (low temperature, high density and low diffusivity) and approach unity for aqueous vapor (high temperature, low den-

Table 3.3: Calculated dynamic and thermodynamic data per TIP3P water molecule in different regions around the N-linked high mannose-type oligosaccharide.

2PT data	0-3.6 Å	0-6 Å	10-15 Å
ρ (g/cm ³) ^a	1.008	0.987	0.979
S(0) (10 ⁻² cm) ^b	0.300±0.02	0.642±0.02	1.541±0.05
D (10 ⁻⁹ m ² /s) ^c	1.16±0.1	2.47±0.1	5.93±0.2
Fluidicity factor (f)	0.1576±0.007	0.2331±0.005	0.3550±0.015
TS _{vib} (kcal/mol) ^d	1.97±0.03	1.75±0.02	1.48±0.02
TS _{conf} (kcal/mol) ^e	0.93±0.04	1.39±0.03	2.14±0.03
TS _{rot,ig} (kcal/mol) ^f	3.2	3.2	3.2
TS (kcal/mol) ^g	6.10 ±0.02	6.34±0.01	6.82±0.02
E (kcal/mol) ^h	-9.612±0.01	-9.543±0.005	-9.420±0.002
A (kcal/mol) ⁱ	-15.71±0.02	-15.88±0.01	-16.24±0.01
ΔA (kcal/mol) ^j	0.53±0.02	0.36±0.01	

^a Average water density calculated as the average number of water molecules in each domain divided by their average volume, determined by a Voronoi decomposition of the simulation box.

^b Intensity of the translational DoS-function at zero frequency.

^c Self-diffusion coefficient calculated from S(0).

^d Vibrational (solid-like) entropy contribution.

^e Configurational (gas-like) entropy contribution.

^f Entropy contribution from the three rotational degrees of freedom. Here approximated by the water ideal gas value.

^g Total absolute entropy per water molecule. $TS = TS_{vib} + TS_{conf} + TS_{rot,ig}$.

^h Average interaction energy per water molecule.

ⁱ Absolute Helmholtz free energy per water molecule, $A = E - TS$

^j $\Delta A = A_R - A_{bulk}$

sity and high diffusivity). The reduced fluidity agrees with the increased "ice-like" character of hydration water. The reduced translational mobility of hydration water is reflected in calculated absolute entropies for water in three regions. The average loss in entropy of a water molecule is, in these calculations, approximately 0.8 kcal/mol for water molecules hydrogen bonded to the saccharide and 0.5 kcal/mol per water molecule for the complete first and second hydration shells. These values are within range of the previously estimated entropy cost of 0-2.1 kcal/mol at 300 K for transferring a water molecule from bulk water to the hydration shell of a biomolecule [148], where the upper limit of the entropy cost in this estimation apply to tightly bound water molecules bound to polar or ionic moieties. Lu and Wong calculated absolute entropies of hydration water molecules around protein kinase A (PKA) to be in the

range of 1.2-4.8 kcal/mol, with a bulk water value of 4.95 kcal/mol at 300 K [149]. A similar loss of entropy of hydration water have been shown for hydration water around DNA [92] and polyamidoamine denrimers [93]. The loss of translational entropy is directly correlated to the low self-diffusion of hydration water, which is, as previously mentioned, observed around many macromolecules.

Average interaction energies per water in the three domains are calculated, as described before, with a cutoff of both electrostatic and van der Waals interactions at a distance of 12 Å. A reaction field correction is used beyond this distance to account for electrostatic interactions beyond the cutoff. The average energies presented in Table 3.3 are calculated for the exact same subset of water molecules used to calculate the center of mass autocorrelation functions. Using the interaction energy E and the total estimated entropy, the absolute Helmholtz free energy per water molecule can be calculated, as well as, free energy differences between water in the different regions around the saccharide. The desolvation free energy change, $\Delta A = A_{bulk} - A_{hydr}$, calculated from these simulations, is -0.36 kcal/mol per water molecule averaged over water molecules in the first and second hydration shells and -0.53 kcal/mol per water molecule for just water molecules hydrogen bonded to the saccharide. These results indicate that the main contribution for the unfavorable free energy of saccharide hydration water comes from water molecules hydrogen bonded to the saccharide and not from water around hydrophobic parts of the saccharide. The calculated entropies and free energies are averaged over water molecules with a wide range of different structural and dynamical properties. Removing a water molecule from the vicinity of the saccharide will however, on average, give a negative free energy change, which would give a favorable contribution to saccharide-saccharide or lectin-saccharide complexation. The cause of the unfavorable free energy of hydration water is from these calculations a result of greater entropy loss than the energetic gain due to favorable interactions with the saccharide. Although 2PT-entropies are calculated using a number of approximations, previous studies have shown that the method provides reliable entropies and free energies for Lennard-Jones systems [68] and liquid water at ambient conditions [93]. As seen in Table 3.3, the main reason for the entropy loss of hydration water can, according to the 2PT-analysis, be attributed to loss in configurational entropy. To check the implementation of the 2PT method and the consistency of the hard-sphere treatment of configurational degrees of freedom within this method, the logarithm of the translational diffusivity is plotted against the inverse TS_{conf} for the three domains in Figure 3.17. The Adam-Gibbs equation shows an exponential relationship between the self-

diffusional coefficient and configurational entropy of a given system of the form $D_T \propto \exp[-B(\rho)/TS_{conf}]$. This relationship connects the purely dynamic quantity D_T to the thermodynamic configurational entropy. Although we only have a small number of points, a clear linear relationship can be seen between calculated configurational hard-sphere entropies and the self-diffusion coefficients calculated from the DoS-functions. This shows that relative changes in translational diffusion rates, which we can calculate with good accuracy from velocity autocorrelation functions or mean square displacements, are well reflected in the calculated configurational entropies.

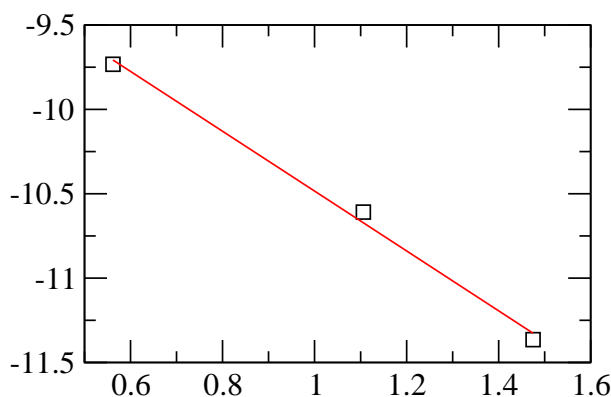


Figure 3.17: $\ln(D_T)$ plotted against $1/TS_{conf}$ for water in the three different regions around the N-linked high mannose-type oligosaccharide. The three points can be fitted to a straight line with a correlation coefficient of 0.998. Error bars are within the size of the data points.

The calculated absolute entropy of bulk water is larger than the standard entropy of 5.0 kcal/mol at 298 K [150]. One reason for the higher calculated entropy is likely to be the high diffusivity of the TIP3P water model (see Figure 3.2 and Table 3.2). Furthermore, the rotational entropy of water is, in these calculations, approximated by the ideal gas value of 3.2 kcal/mol. This is likely to be an overestimation of the rotational entropy of water in the condensed phase, due to the higher restriction of rotational motions in the hydrogen bond network of liquid water. Amzel [151] estimated the rotational entropy of liquid water at 298 K to be approximately 1.21 kcal/mol, less than half the ideal gas value. The use of this value gives an absolute entropy per bulk water molecule closer to the standard entropy, but does not change the conclusions drawn here since the same value is used for all three domains. Figure 3.12 shows that also the rotational dynamics of water is retarded in the vicinity of the saccharide, which would further increase the difference in entropy between hydration and bulk water.

3.5 Dependence of hydration properties on carbohydrate density

In this section, we investigate the effect of increased carbohydrate density on hydration water properties. This is interesting for understanding the role of water in cellular recognition processes involving extended parts of the saccharide units in the cell surface glycocalyx. It is reasonable to assume that hydration properties of monosaccharides, commonly used in studies of lectin-saccharide binding, are different to those of the high density saccharide forest of the glycocalyx. If that is the case, such differences should also be projected onto binding affinities of, for instance, lectins to respective carbohydrate system. Two further systems, presented in Figure 3.18, are modeled in

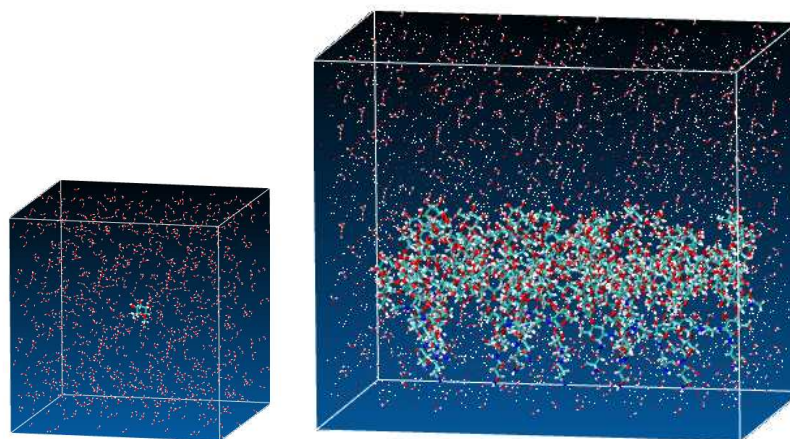


Figure 3.18: Simulation boxes of the O-methylated mannoside and the high mannose-type oligosaccharide array. The water density is reduced in both figures for a clearer representation.

this section. The first system consists of a single O-methylated mannoside in a large water box of initial dimensions $58.74 \times 57.74 \times 57.87 \text{ \AA}$. The second system models a very high density array of N-linked high mannose-type oligosaccharides. Here the simulation box consists of 12 high mannose-type oligosaccharide units separated by a distance of approximately 12 \AA in each dimension in a water box with 6293 water molecules giving a simulation box of dimensions $54.82 \times 71.20 \times 81.80$. Without speculating about how well this saccharide array reflects the actual density and arrangements of saccharides in the glycocalyx, this system is used to model the opposite extreme of

the "isolated" N-linked high mannose-type oligosaccharide discussed in the previous sections of this chapter. The same simulation parameters are used for all three systems (see Section 3.3), using GLYCAM04 and TIP3P force fields to model carbohydrates and water respectively. The analyses performed in the previous sections are repeated here for the two new systems. Comparing the three systems will show the cooperative effect of carbohydrates on hydration water properties. In order to compare the results to those in the previous chapter, all twelve high-mannose oligosaccharide units are in these calculations considered as a single large solute molecule. Water properties are, as before, calculated with respect to the distance to the closest non-hydrogen carbohydrate atom. In the following plots, Man denotes the O-methylated mannoside, S the N-linked high mannose-type oligosaccharide and SA the array of N-linked high mannose-type oligosaccharides.

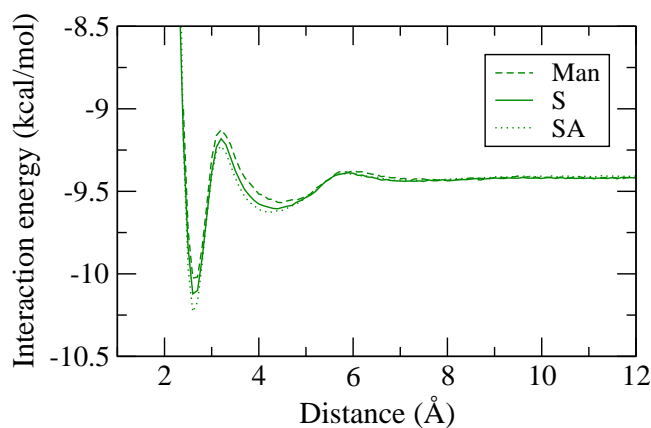


Figure 3.19: Average interaction energy per water molecule as a function of the distance to the closest non-hydrogen atom in the O-methylated mannoside (dashed lines), the high mannose-type oligosaccharide (full lines) and the array of high mannose-type oligosaccharides (dotted lines).

Structural modifications of the hydrogen bond network, as reflected in the average interaction energy per water molecule in Figure 3.19, are very similar for the three systems. A slight shift towards more negative interaction energies can be seen for water in the first and second hydration shells as the carbohydrate density increases. Beyond the second hydration shell, structural perturbations of the hydrogen bond network is basically the same for all three systems. The small effect of the increased saccharide density on structural properties of water is further evidenced by comparing the structural order parameter q of water within a distance of 4 Å (Figure 3.20a) and in the

interval 4 Å to 6 Å (Figure 3.20b) from the closest non-hydrogen carbohydrate atom in the three systems. A very slight increased destructuring effect can be seen for first hydration shell water as carbohydrate density increases. The tetrahedral order of water beyond 4 Å is identical to bulk water for all three systems (see Figure 3.3b). These

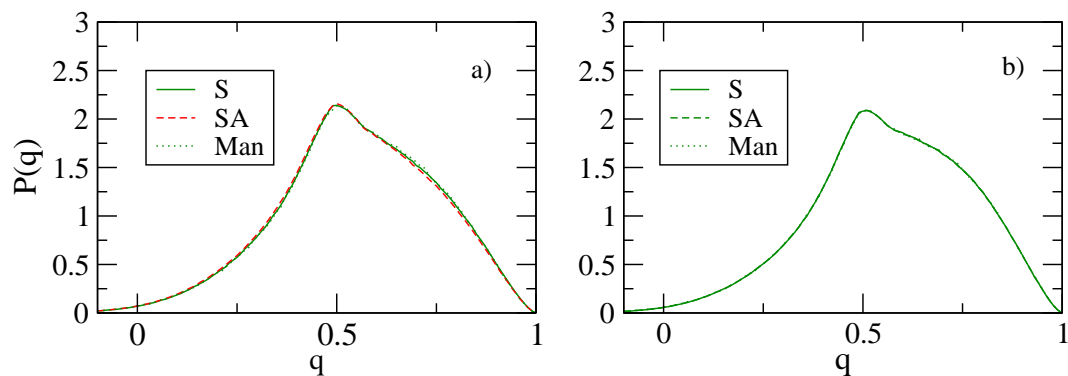


Figure 3.20: Probability distributions of the tetrahedral order parameter, q , of water molecules within a distance of 4 Å a) and in the region of 4-6 Å to the closest non-hydrogen atom in the three carbohydrate systems.

results are somewhat surprising given that a large amount of hydration water will be in between individual saccharide units in the saccharide array, experiencing a highly non-bulk environment. This phenomenon can be explained by a combination of a high flexibility of the water hydrogen bond network to accommodate to new surroundings and the ability of carbohydrates to blend in well with the water hydrogen bond network through interactions with their many hydrogen bond donors and acceptors.

Larger differences can be seen for the dynamics of hydration water. Diffusion coefficients and rotational second-order relaxation time, as functions of the distance of the closest non-hydrogen carbohydrate atom, are shown in Figures 3.21a and 3.21b. A clear correlation between carbohydrate density and dynamics of water can be seen. As the carbohydrate density increases, are both translational and rotational diffusion rates of hydration water decreased. Figures 3.21a and 3.21b show that the increase in saccharide density has a greater impact on the rotational dynamics of water than the diffusivity. A similar decoupling of water translational and rotational dynamics was observed around various mono- and disaccharides in the studies of Lee *et al.* [26]. In this study, it was shown that the rotational velocity of water decays faster than the self-diffusion, as water molecules enter the hydration shell of the saccharides. In these simulations, the decoupling is further increased with increasing saccharide density.

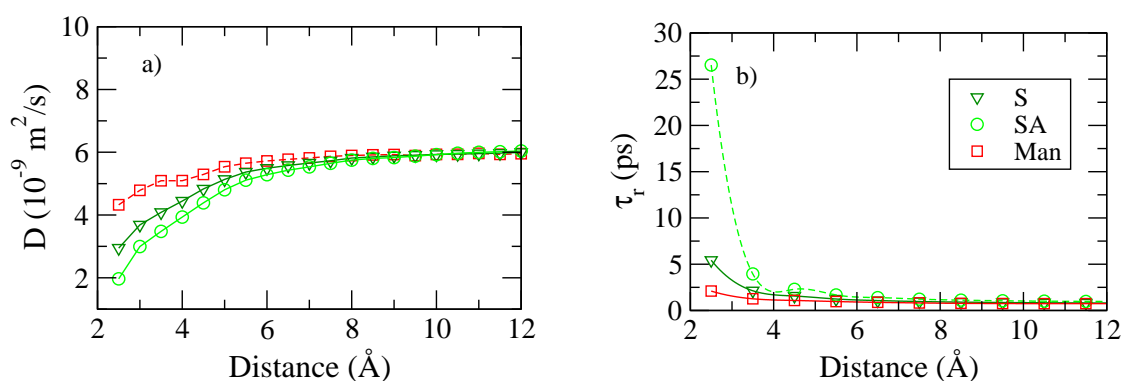


Figure 3.21: a) Local self-diffusion coefficients and b) local rotational correlation times of water molecules as a function of distance to the closest non-hydrogen atom in the O-methylated mannoside (Man), N-linked high mannose-type oligosaccharide (S) and the saccharide array (SA).

The retardation of water dynamics can be explained with a increased likelihood of a water molecule being trapped in the carbohydrate hydration shell with increasing carbohydrate density. Figure 3.22a shows the probability distribution of residence times of water in the first hydration shell around the saccharide array system. In agreement with the results for the N-linked high mannose-type oligosaccharide, presented in the previous sections, there is a wide spread in the distribution of residence times. A majority of first hydration shell water have residence times close to mean bulk values in Figure 3.22b. There is however a shift towards longer residence times of water around the saccharide array. This shift increases with the saccharide density. For water in the first hydration shell around the mannoside, is the longest residence time observed 28 ps. For the saccharide array system, several occurrences of residence times for hundreds of picoseconds are observed, and single occurrences of residence times on the nanosecond time scale. As seen in Figure 3.19, such changes in residence time are unlikely to be a result of the slight shift towards more favorable carbohydrate-water interactions. Rather, these simulations show that these long lived hydration shell water molecules are trapped in the saccharide array, due to steric hindrance, as a result of overlapping hydration shells of individual saccharide units. Comparing the maximum residence time of water molecules in the first hydration shell of the mannoside and the isolated saccharide of 28 ps and 241 ps respectively it is clear that such effects are also present in larger oligosaccharides.

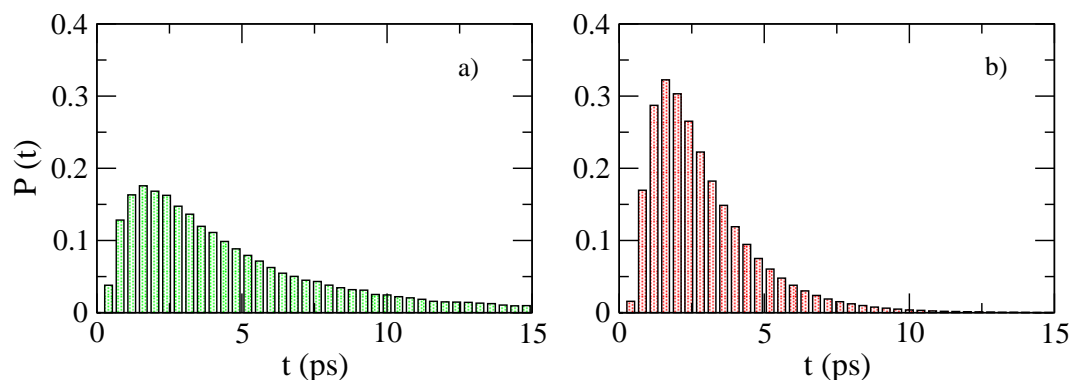


Figure 3.22: a) Probability distribution of water residence times in the first hydration shell around the saccharide calculated in local saccharide residual coordinate systems as described in Section 3.4.2. b) Probability distribution of water residence times in bulk water calculated in the global coordinate system.

The investigation of thermodynamic properties of hydration water is here repeated for the two new systems. Water is, as before, examined in three different regions around the saccharide. Water within 3.6 \AA , water within 6 \AA and water beyond a distance of 10 \AA to the closest non-hydrogen carbohydrate atom. Although these volumes are different for the three systems, using the same subset of water molecules in both energy and entropy calculations, a direct comparison of thermodynamic properties of hydration water can be made for the three systems. Due to the smaller number of water molecules in the regions around the mannoside, the center of mass velocity autocorrelation functions and corresponding DoS functions are calculated and averaged over four 160 ps trajectories. For the saccharide array, seven trajectories of 40 ps length were used. These trajectories were calculated, for both systems, in the NVT-ensemble at 300 K.

The results in Tables 3.3–3.5 show that thermodynamic properties of hydration water differ for the three systems. The relationship between rotational and translational dynamics of hydration water and carbohydrate density is reflected in the calculated entropies. The increased "ice-like" character of the dynamics of hydration water, discussed in the previous section, is here found to be further increased with increasing saccharide density. This can be seen as a constant decrease in intensity of the DoS functions at zero frequency ($S(0)$), as well as, the fluidity factor (f) of hydration shell water molecules, relative to the bulk water counterparts in each simulation. The fluidity factor of water molecules within a distance of $0\text{--}6 \text{ \AA}$ takes values of 0.31, 0.23

Table 3.4: Dynamic and thermodynamic data per TIP3P water molecule at 300 K in different regions around the O-methylated mannoside. See Table 3.3 for table captions.

2PT data	0-3.6 Å	0-6 Å	10-15 Å
ρ (g/cm ³)	0.991	0.983	0.976
S(0) (10 ⁻² cm)	0.553±0.1	1.150±0.07	1.447±0.04
D (10 ⁻⁹ m ² /s)	2.13±0.4	4.43±0.3	5.57±0.2
Fluidicity factor (f)	0.212±0.02	0.310±0.01	0.345±0.005
TS _{vib} (kcal/mol)	1.88±0.01	1.51±0.02	1.51±0.02
TS _{conf} (kcal/mol)	1.26±0.1	1.86±0.05	2.07±0.03
TS _{rot,ig} (kcal/mol)	3.2	3.2	3.2
TS (kcal/mol)	6.34±0.04	6.57±0.03	6.79±0.01
E (kcal/mol)	-9.50±0.02	-9.48±0.01	-9.42±0.002
A (kcal/mol)	-15.84±0.03	-16.05±0.02	-16.21±0.06
ΔA (kcal/mol)	0.37±0.03	0.16±0.03	

and 0.18 for the mannoside, the N-linked high mannose-type oligosaccharide and the saccharide array respectively. The calculated average entropy loss per water molecule in the same region is 0.22 kcal/mol for the mannoside, compared to 0.48 kcal/mol and 0.69 kcal/mol for the oligosaccharide and the saccharide array. As seen in Figure 3.19, an increase in carbohydrate density induces a shift towards more negative interaction energies. The relative loss in entropy is, however, larger than the energetic gain, resulting in increasing unfavorable absolute free energies of hydration water as carbohydrate density increases. This means that there is, on average, a higher desolvation free energy gain as the carbohydrate system gets larger. The calculated desolvation free energy change, per water molecule, is -0.16, -0.36 and -0.53 kcal/mol for water molecules within 0-6 Å of the mannoside, the oligosaccharide and the oligosaccharide array. This is a direct consequence of the increased retardation of water translational dynamics. Taking changes in rotational velocities into account would, as seen in Figure 3.21, most likely further increase the desolvation free energy differences between the systems.

Compensatory changes in enthalpy and entropy is often observed in microcalorimetric measurements of lectins binding with various mono- and oligosaccharides. A linear relationship between measured ΔH and $T\Delta S$ values can be seen for different saccharide ligands binding to the lectins, with a slope close to unity [152]. This shows that, although relatively large differences can be seen in the enthalpy and entropy terms,

Table 3.5: Dynamic and thermodynamic data per TIP3P water molecule at 300 K in different regions around the array of N-linked high mannose-type oligosaccharides. See Table 3.3 for table captions.

2PT data	0-3.6 Å	0-6 Å	10-15 Å
ρ (g/cm ³)	1.01	0.991	0.983
S(0) (10 ⁻² cm)	0.149±0.02	0.388±0.008	1.596±0.06
D _T (10 ⁻⁹ m ² /s)	0.58±0.02	1.49±0.03	6.14±0.2
Fluicidity factor (f)	0.1096±0.002	0.1808±0.002	0.3609±0.006
TS _{vib} (kcal/mol)	2.04±0.01	1.87±0.01	1.45±0.02
TS _{conf} (kcal/mol)	0.63±0.01	1.07±0.01	2.17±0.03
TS _{rot,ig} (kcal/mol)	3.2	3.2	3.2
TS (kcal/mol)	5.87±0.01	6.14±0.01	6.83±0.02
E (kcal/mol)	-9.659±0.01	-9.577±0.01	-9.415±0.005
A (kcal/mol)	-15.53±0.01	-15.71±0.01	-16.24±0.02
ΔA (kcal/mol)	0.71±0.02	0.53±0.02	

the total free energy change is typically small. From these simulations, we can directly investigate the contribution from solvent reorganization effects to this enthalpy-entropy compensation phenomenon. Figure 3.23 shows that changes in interaction energy and entropy of carbohydrate hydration water also display a linear relationship. Due to the relative larger change in entropy, the slope is approximately 4 from these calculations. To compare to experimental measurements, enthalpy and entropy changes of both receptor and ligand, as well as hydration water of the receptor, have to be taken into consideration. It is interesting that enthalpy-entropy changes associated with desolvation of the saccharide are opposite to those experienced by the solute molecules. Computer simulations have shown that the lectin-carbohydrate complexation is associated with an enthalpy gain from favorable receptor-ligand interactions and entropic loss, which is mainly due to restriction of translational and rotational degrees of freedom of the ligand, [31]. Just the loss in translational and rotational entropy has been estimated to be in the range of 12-15 kcal/mol for mono- and disaccharides [153]. The entropy gain from released carbohydrate hydration water could explain why only small changes in entropy are observed in experimental lectin-carbohydrate affinity measurements [152], even though saccharides are highly flexible molecules.

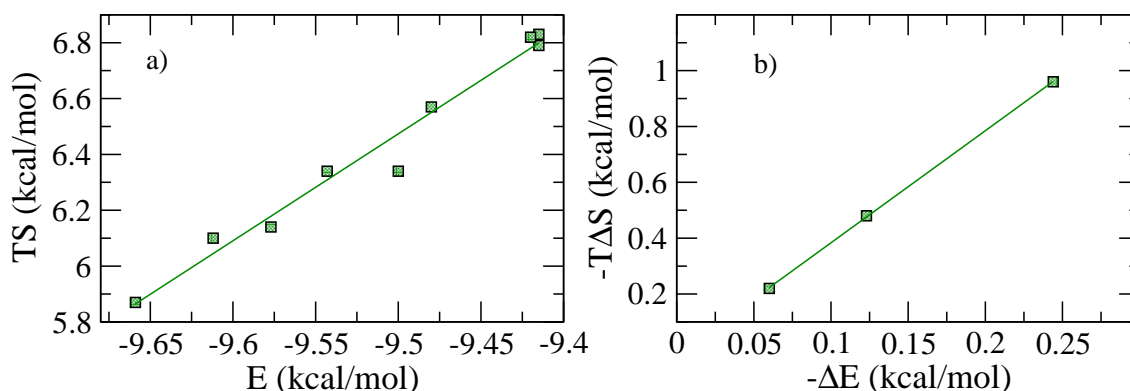


Figure 3.23: a) Calculated absolute entropies plotted with respect to calculated average interaction energies per water molecule in the three different domains for all three systems. The points can be fitted with correlation coefficient of 0.986 to a straight line with a slope of 3.88. b) Average desolvation entropy changes plotted with respect to the corresponding average desolvation interaction energy changes for hydration water in first and second hydration shells for the three different systems. The three points can be fitted with a correlation coefficient of 0.999 to a straight line with a slope of 4.01.

3.6 Saccharide hydration with SPC/E water

The force field comparison in Section 3.2 showed that the TIP3P force field models translational diffusional rates of water poorly. The diffusion coefficient of TIP3P water, calculated from the average mean square displacement from a pure water system, is $6.1 \times 10^{-9} \text{m}^2/\text{s}$, which is over two times as large as the experimental values of $2.3 \times 10^{-9} \text{m}^2/\text{s}$. Here the question arises if the relatively large entropy loss of hydration water, presented in the previous sections, is just a result of this inherent error of the TIP3P force field. The SPC/E model has previously been shown to reproduce experimental entropies of water at ambient conditions [154] and should therefore provide a good reference. Reference simulations of both the isolated high mannose-type oligosaccharide and the dense array of high-mannose oligosaccharides were performed, using the same simulation parameters described in Section 3.3, together with the SPC/E and GLYCAM04 force fields to model water and carbohydrates respectively. System sizes and systems setups are identical to the TIP3P simulations. Structure, dynamics and thermodynamic properties of hydration water, as modeled by the SPC/E force field, are compared to the previously presented results calculated with the TIP3P force field.

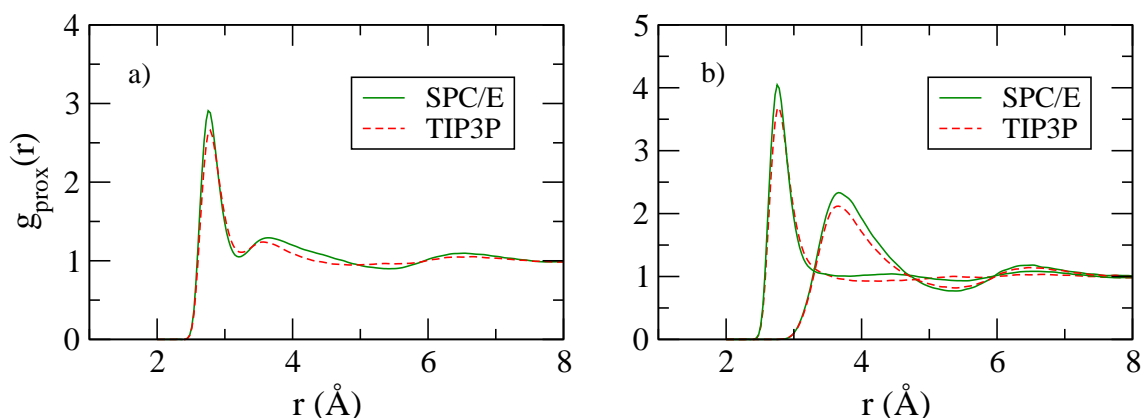


Figure 3.24: a) Proximal distribution functions of water calculated with respect to closest non-hydrogen atom in the N-linked high-mannose oligosaccharide for TIP3P (full line) and SPC/E (dashed line). b) Decomposed proximal distribution functions with respect to polar and non-polar atoms in the saccharide.

Although the SPC/E and TIP3P force fields model remarkably different structures of the water hydrogen bond network, the difference in structural properties of hydration water is, as seen in Figure 3.24, fairly similar. SPC/E water shows slightly larger first and second peaks in the proximal distribution function, corresponding to water around both polar and apolar saccharide atoms. The contribution of the non-polar hydration pattern can also be seen to be larger for SPC/E water, giving an increased structuring beyond first hydration shell water. The structuring/destructuring pattern of the TIP3P hydrogen bond network is very similar for the SPC/E force field. The interaction energy per water molecule, plotted as a function of the distance to the closest non-hydrogen atom in the saccharide for both force fields in Figure 3.25a, summarizes the perturbation of the hydrogen bond network of SPC/E water around the saccharide. Since the two water models have different average bulk interaction energies, bulk interaction energies are subtracted from each curve for a direct comparison.

SPC/E water molecules, hydrogen bonded to the saccharide, are slightly higher in energy compared to TIP3P water, relative to the average bulk water interaction energy of respective force field. The opposite trend can be seen for second hydration shell water. Figure 3.25b shows average interaction energies of SPC/E water from the single oligosaccharide (S) and for the saccharide array systems (SA). In agreement with the TIP3P force field, only minor structural differences can be seen for the two systems. The structural order of hydration water, calculated using the previously discussed tetra-

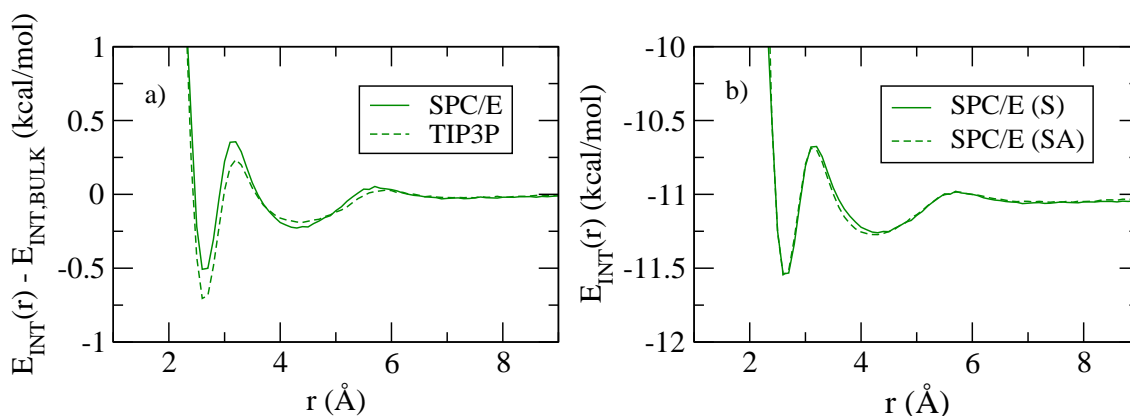


Figure 3.25: Average interaction energy per water molecule with respect to the distance to closest non-hydrogen atom in the saccharide for TIP3P (full line) and SPC/E (dashed line). The interaction energies are subtracted with the bulk values for each water model for a direct comparison.

hedral order parameter, is also in agreement with the TIP3P force field. Only a very small increase in structural perturbations of hydration water can be seen in the saccharide array simulation. The structure of the water hydrogen bond network is the same as bulk water beyond 6 Å for both force fields. These results show that the modest effect on water structure induced by the saccharide can not be attributed to the too weakly modeled hydrogen bond network of the TIP3P water force field. Figure 3.26 shows local self-diffusion coefficients and second-order orientational relaxation times for SPC/E water as functions of distance to the closest non-hydrogen saccharide atom in both the single saccharide and the saccharide array. The local diffusion coefficients and rotational correlation times are calculated as described in Section 3.4.2. The same trend, observed in the TIP3P simulations, can be seen here. Both translational and rotational diffusion rates are severely reduced in the vicinity of the saccharide systems. The relative difference in self-diffusion coefficients of first hydration shell and bulk water is approximately the same for the two water models. Calculated second-order rotational relaxation times are 19 ps and 5.4 ps for water molecules in direct contact with the N-linked high mannose-type oligosaccharide, compared to bulk values of 2 ps and 0.81 ps for SPC/E and TIP3P water respectively. Thus, the relative decrease in rotational velocity of hydration water is slightly larger for SPC/E water. Again, it is interesting to note that although the two water force fields model dynamic properties of water very differently, relative differences in translational and rotational

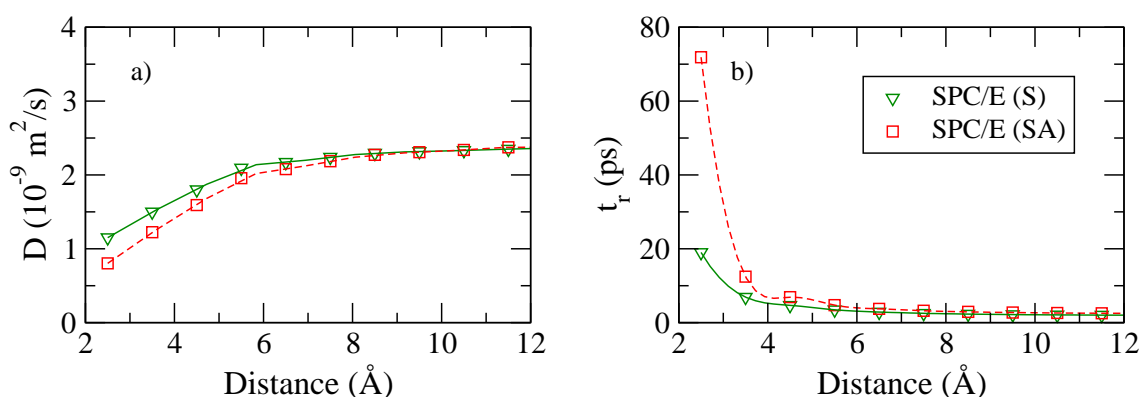


Figure 3.26: a) Local self-diffusion coefficients for SPC/E water around the N-linked high mannose-type oligosaccharide (full lines) and the oligosaccharide array (dashed lines) b) Second-order orientational relaxation times for the two systems. Distances are calculated with respect to the closest non-hydrogen saccharide atom.

diffusion rates of bulk and hydration water are in good agreement between the two force fields. The 2PT entropy and free energy analysis of SPC/E water is presented in Table 3.6 for hydration water around the N-linked high mannose-type oligosaccharide. For completeness, calculated SPC/E center of mass velocity autocorrelation functions and corresponding power spectra are shown in Figures 3.27 and 3.28.

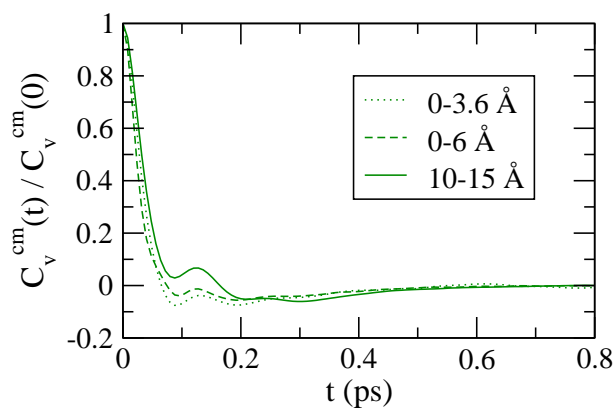


Figure 3.27: Center of mass water velocity autocorrelation functions for SPC/E water within 3.6 \AA (dotted line), within 6.0 \AA (dashed line) and beyond 10 \AA (full line) from the closest non-hydrogen atom in the N-linked high mannose-type oligosaccharide.

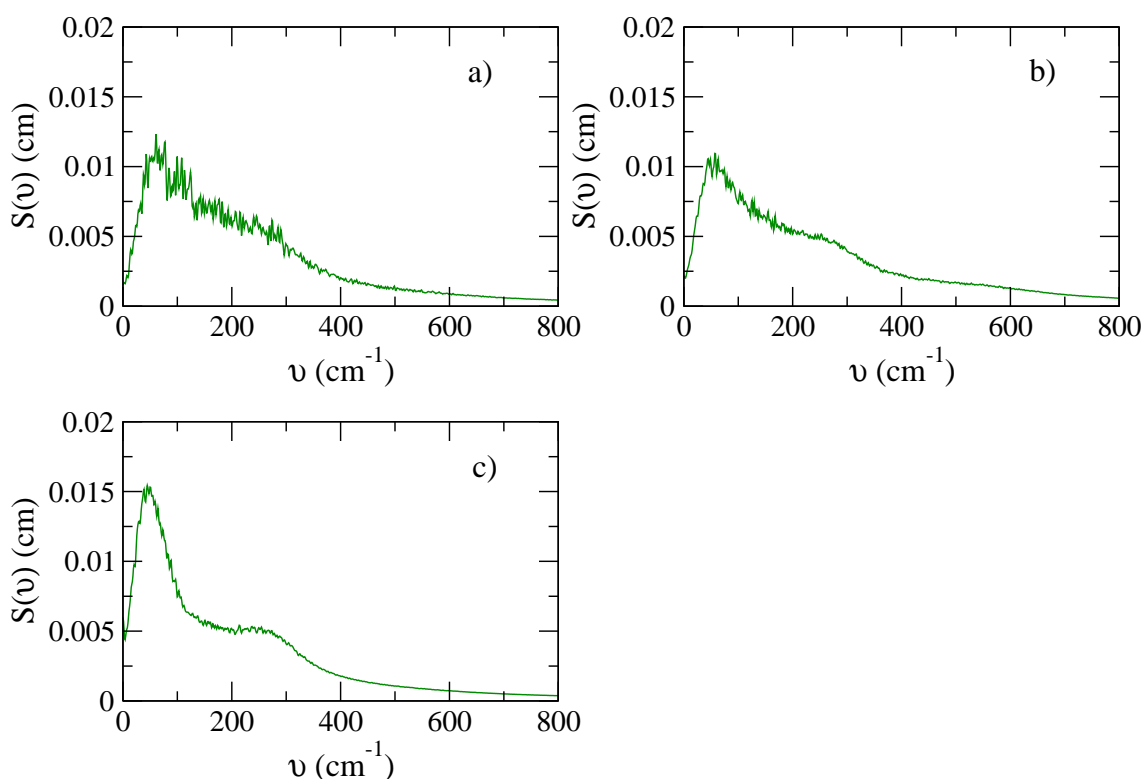


Figure 3.28: Translational density of states functions of SPC/E water in three different regions around the N-linked high-mannose oligosaccharide. a) within a distance of 3.6 Å, b) within a distance of 6 Å and c) between 10-15 Å from the closest non-hydrogen saccharide atom.

The data in Table 3.6 shows that SPC/E hydration water, similarly to TIP3P water, displays an increased ice-like character as seen in higher density, slower translational diffusion rates and lower fluidity factors. This results in an entropy loss of water in the vicinity of the saccharide. The average entropy loss per SPC/E water molecule in the first and second regions around the saccharide is 0.43 and 0.51 kcal/mol compared to 0.48 and 0.72 kcal/mol for TIP3P water. Together with changes in the average interaction energy per water molecule in each region, calculated desolvation free energy changes are -0.33 and -0.43 kcal/mol for SPC/E water, compared to the corresponding values of -0.36 and -0.53 kcal/mol for TIP3P water. Hence, both force fields agree on the observation that desolvation of the saccharide will give a favorable free energy contribution to the total free energy change of molecular association processes involving saccharides.

Table 3.6: Dynamic and thermodynamic data of SPC/E water at 300 K around different regions around the N-linked high mannose-type oligosaccharide. See Table 3.3 for table captions.

2PT data	0-3.6 Å	0-6 Å	10-15 Å
ρ (g/cm ³)	1.026	1.011	0.996
S(0) (10 ⁻² cm)	0.149±0.01	0.222±0.01	0.57±0.01
D (10 ⁻⁹ m ² /s)	0.57±0.5	0.85±0.3	2.2±0.05
Fluidity factor (f)	0.1089±0.005	0.1352±0.002	0.2211±0.002
TS _{vib} (kcal/mol)	1.97±0.02	1.90±0.01	1.81±0.01
TS _{conf} (kcal/mol)	0.63±0.03	0.79±0.02	1.32±0.02
TS _{rot,ig} (kcal/mol)	3.2	3.2	3.2
TS (kcal/mol)	5.81±0.01	5.89±0.01	6.32±0.01
E (kcal/mol)	-11.12±0.01	-11.14±0.01	-11.04±0.00
A (kcal/mol)	-16.93±0.01	-17.03±0.01	-17.36±0.01
ΔA (kcal/mol)	0.43±0.01	0.33±0.01	

The 2PT analysis of water around the saccharide array is not presented here, but as indicated by Figures 3.26 and 3.25, an increased entropy penalty is observed for the saccharide array, which only partly is compensated by a small decrease in energy. This results in increased unfavorable free energy of hydration water around the saccharide array. All together, these results show that all calculations presented in the previous sections can be qualitatively reproduced using SPC/E water. Relative differences between bulk and hydration are also in good quantitative agreement for the two force fields. This provides some support of the accuracy of the results presented in the previous sections and shows that TIP3P water provides sufficient accuracy to model biomolecular solvation properties.

3.7 Discussion

In this chapter, we have investigated structural, dynamic and thermodynamic properties of hydration water of systems with varying carbohydrate density. We have shown that structural modifications of the hydrogen bond network of water are rather insensitive to changes in carbohydrate density. Translational and rotational dynamics of hydration water do, on the other hand, display a greater dependence on the carbohydrate system. These studies show that carbohydrate hydration water molecules, on average, pay an entropic penalty which is greater than the gain in potential energy. These findings are in agreement with the hypothesis of Lemieux [104] stating that "Water behaves as a catalyst in the sense that it forms activated complexes that are necessary to the molecular association (of saccharide-lectin complexes)". In contrast to Lemieux's hypothesis, our studies show that the catalytic role of water is not a result of an enthalpy gain associated with the release of carbohydrate hydration water, but rather due to the entropic gain of water molecules removed from the molecular surface of the carbohydrate. The difference between the energy gain and the entropy loss is increased with increasing carbohydrate density, showing that randomly removing a water molecule from the hydration shell of an oligosaccharide would give a larger free energy gain than for a monosaccharide. For the dense saccharide array used to model a possible arrangement of the glycocalyx, the free energy gain is even further increased. All-in-all, these results show that favorable free energy changes due to solvent reorganization effects are likely to be present in all lectin-carbohydrate binding processes and that the free energy becomes more favorable with increasing size of the carbohydrate system. To get the full picture of the role of hydration water on the thermodynamics of lectin-carbohydrate binding desolvation of the lectin binding pockets needs to be taken into account. This adds more complexity since hydration properties of receptor binding pocket depend on local curvatures and structural properties of the receptor [155, 156]. Also, eventual structural water molecules immobilized at the binding interface and other solvation related effects, such as, pH, ionic strength and the presence of metal ions need to be considered. There are however, several experimental observations that can be explained from the results presented in this chapter. The findings in this chapter show a hydration pattern around carbohydrates, similar to the classical picture of hydrophobic hydration, i.e. increased structure and more "ice-like" character of water surrounding the solute. Williams et al. [157] showed that the thermodynamics of lectin-carbohydrate binding is similar to that of the binding of

small hydrophobic molecules to specific receptors. Since water is the main thing these systems have in common, it is reasonable to assume that thermodynamic changes due to solvent reorganization make an important contribution to the overall thermodynamic changes of the binding process. Experimentally measured free energy changes show that lectin-carbohydrate binding is enthalpy-driven, with relatively large negative changes in enthalpy compensated by smaller negative or positive contributions of the $T\Delta S$ term [103, 152, 158]. A much higher entropy loss would be expected taking changes losses in rotational, translational and configurational entropy contributions of the solute molecules into account. The entropy gain from released hydration water molecules from the carbohydrate molecular surface to bulk water environment could explain the total small entropy change observed in lectin-carbohydrate complexation processes. Another characteristics of lectin-saccharide complexation is a small negative change in specific heat capacity of the system [103, 152], which has generally been thought to reflect changes in the solvent structure. A general feature of hydrophobic hydration is a decrease in entropy and an increase in heat capacity [159]. Removal of the structured hydration water around the saccharide would, in agreement with these results, contribute to a negative heat capacity change. Thus, many important thermodynamic characteristics of lectin-carbohydrate interactions can be related to carbohydrate water-interactions. The results presented in this chapter might add a small piece to the puzzle of lectin-carbohydrate recognition and binding affinities.

4 Carbohydrate-ion interactions

Water is not the same everywhere in its liquid bulk phase. When the neighborhood of a water molecule changes from other water molecules to something else, one may expect that the local water structure, and hence also the local solvation properties of water, will change accordingly. This expectation has been confirmed in recent decades, by experimental and theoretical studies of water structure and ion-solvation propensities in various situations, including e.g. interfaces between the liquid water and air [160] as well as in confined situations [161] and clusters [162, 163]. Differences in hydration properties of ions are, in turn, projected onto interactions with other molecules, resulting in ion-specific phenomena frequently observed in biological systems. In 1888 Hofmeister [164] observed that the solubility of proteins depend on the type of the cations and anions in an ionic solution. This ion-specific phenomenon has been attributed to relative differences in solvation properties of the ions. Recent studies have shown that ion solvation properties alone cannot explain the phenomena observed by Hofmeister [165]. The interaction between proteins and ions in solution is rather a result of specific local interactions between ions and protein side chains, which in turn depend on both local solvation properties of the protein and ion solvation properties. In more recent studies, ion-specific phenomena have been observed for other biomolecules in the cell. The monovalent cations, Na^+ and K^+ , have received much interest in the literature due to their abundance in biological systems and importance for cellular functions. Studies of potassium channels have shown that the free energy barrier for replacing K^+ -coordinated water molecules by the carbonyl oxygen atoms of proteins in the interior of the channel is small compared to that of Na^+ and allows for a quick and selective transportation of K^+ [166]. Carillo-Tripp et al. [167] showed that the cost of constraining a hydrated potassium ion inside a narrow nanopore is lower than that of constraining a sodium ion, and that this is a result of the higher geometrical flexibility of potassium ions in their coordination with water. The simulation studies

of Chen et al. [168] found a higher binding preference of Na^+ to the phosphate groups of DNA, and a higher density of K^+ ions near the polar groups in the grooves of the DNA. A recent investigation of Na^+/K^+ affinities to different proteins by Vrbka et al. [169] observed a strong preference of Na^+ over K^+ close to the surface of different proteins, where the ion selectivity stems from cation specific interactions with side-chain carboxylate groups. A part of the clue for understanding these ion-specific effects lies in understanding differences in solvation properties of the cations. Previous studies have highlighted differences in how cations interact with solvation water. Using bulk water-water interaction energies as a reference, small monovalent ions are strongly hydrated and are classified as kosmotropes, whereas larger monovalent ions bind water weakly and are classified as chaotropes [170]. With this classification, Na^+ is a weak kosmotrope, or structure maker, in the sense that it has a relatively well defined and stable first hydration shell. The weak chaotrope K^+ is more flexible in its coordination with water, and has a less stable and more dynamic first hydration shell as water-water interaction dominates over ion-water interactions. These simple differences in the balance between the interactions water-water, Na^+ -water and K^+ -water have already been used successfully for a simple but effective explanation of the strikingly different structural propensities in cation-water clusters of Na^+ and K^+ [171]. The divalent metal cations are, in turn, stronger kosmotropes with increased strength and structure of cation-water interactions.

The aforementioned studies show that ion affinities and ion-specific phenomena in biological systems are results of a delicate balance between solvation properties of the ions, solvation properties of the macromolecule and direct macromolecule-ion interactions. The knowledge about ion-interactions with carbohydrates of the cell is, to this date, quite limited. Specific carbohydrate-ion interactions are of interest for understanding mechanisms behind cation dependent processes in the cell, such as, carbohydrate-carbohydrate interactions mediating cell adhesion processes and lectin-carbohydrate recognition of C-type lectins [172]. Furthermore, since all participating molecules of the glycocalyx environment of eukaryotic cell surfaces are water-exposed including the carbohydrate-recognition domains of lectins, molecular interactions of carbohydrates in aqueous, highly ionic systems should receive great attention in glycobiology. In this study, we have investigated the static and dynamic behavior of biologically relevant cations around the cell surface carbohydrates of the glycocalyx.

4.1 System and simulation protocol

The basic oligosaccharide unit $\text{Man}_6\text{GlcNAc}_2$ used in all our calculations is shown in Figure 4.1. Classical-mechanical molecular dynamics (MD) simulations were carried out on systems consisting either of one isolated $\text{Man}_6\text{GlcNAc}_2$ unit or of an array of nine of these units, in either 0.15 M NaCl or 0.15 M KCl aqueous solution. The single oligosaccharide/ion systems were all set up the same way. The saccharide was solvated in a large water box with unit cell dimensions of 51.88 Å x 72.31 Å x 48.93 Å. This setup models an isolated saccharide, separated from its periodic images by approximately 50 Å, on an imaginary surface spanned by the xz -plane of the simulation box. Na^+ , K^+ , and Cl^- ions were randomly added to the box by and Cl^- ions were randomly added to the box by removing overlapping water molecules. The resulting NaCl and KCl systems consisted of 6132 water molecules, 17 sodium or potassium ions and 17 chloride ions. Systems of higher carbohydrate density were set up in a similar way. Nine N-linked $\text{Man}_6\text{GlcNAc}_2$ oligosaccharides were placed in a regular arrangement in the simulation box, each saccharide separated by approximately 20 Å (see Figure 4.11). This setup models a dense saccharide array located on the imaginary surface spanned by the xz -plane of the simulation box. A large y -dimension of the box was chosen to model a bulk water environment above the saccharide array, giving box dimensions of 74 Å x 80 Å x 74 Å. For these systems, 34 sodium or potassium ions were added together with 34 chloride ions, giving a final number of 12718 water molecules. The divalent cation-carbohydrate systems were set up in the same manner, using either 15 Mg^{2+} or Ca^{2+} cations, together with 30 Cl^- counterions and 5428 water molecules, for the single oligosaccharide simulations. In the case of the saccharide array, 34 Mg^{2+} or Ca^{2+} cations, 68 Cl^- anions and 12684 water molecules were used. Starting conformations of the saccharides were taken from the SWEET database [173].

All ion force fields used in this study treat the ions as simple point-charged, non-polarizable Lennard-Jones particles. Na^+ , K^+ and Cl^- force fields are from Beglov and Roux [174]. Lennard-Jones parameters for Mg^{2+} and Ca^{2+} were calculated using the Åqvist parameters [175] together with TIP3P water, giving (r, ϵ) values of (1.5734 Å, 0.87504 kcal/mol) and (2.6522 Å, 0.44966 kcal/mol) for Mg^{2+} and Ca^{2+} respectively. Both the carbohydrate force field, GLYCAM04 [24], and ion parameters used in this study, were developed together with the TIP3P water model [44], giving a consistent set of parameters for the whole system.

All systems were simulated using NAMD version 2.6 [36], using the same simulation

protocol/parameters. Minimization, heating to 300 K at constant volume, and equilibration at a constant pressure of 1 bar was followed by production runs in the NVT-ensemble, using a 2 fs integration time step, for a total of 60 ns. A time step of 4fs was used for long range electrostatic interactions. The temperature was controlled by the Berendsen thermostat [51] with a time constant of 5 ps. Long range electrostatic interactions were accounted for using the PME method [132]. Lennard-Jones interaction was cut off at a distance of 12 Å, without using switching functions. SHAKE [55] was used to constrain all heavy-atom-hydrogen bonds. Positional restraints were applied to the main-chain asparagine atoms in the *N*-link (C_α , N and C(O)), allowing no diffusional motion of the saccharides and restraining each saccharide to the *xy*-plane of the simulation box. 1–4 bonded electrostatic and vdW scaling was set to unity in accordance with the GLYCAM04 force field [133]. Coordinates were saved every 0.4 ps of the trajectories, giving 150000 frames for data analysis.

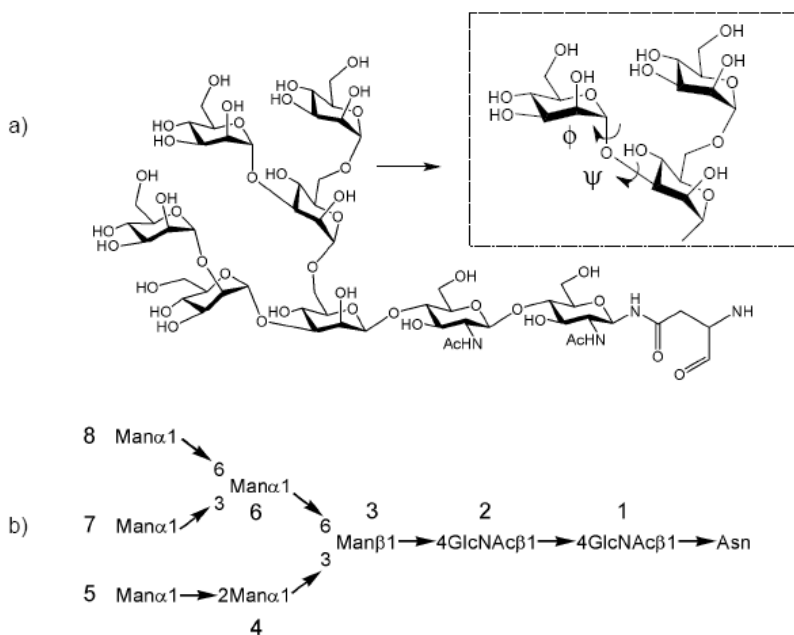


Figure 4.1: Pictorial description of the $\text{Man}_6\text{GlcNAc}_2\text{-Asn}$ oligosaccharide.

4.2 Ion hydration

The importance of local cation-water interactions for describing ion solvation properties is evidenced by recent studies showing that, rather than causing any long ranged perturbations of the hydrogen bond network of hydration water, structuring or destructuring effects are limited to water molecules in the vicinity of the ion [176–178]. Hence, we begin by presenting various structural and dynamic properties of the cation-first hydration shell complex for the ions used in this study, Na^+ , K^+ , Mg^{2+} and Ca^{2+} , calculated from the carbohydrate/ion/water trajectories in volumes of the simulation box where water has bulk properties. Given the obvious limitations of studying ion hydration with standard force fields, it is interesting to see how these force fields model ion solvation properties compared to experiment and to simulations at higher levels of theory.

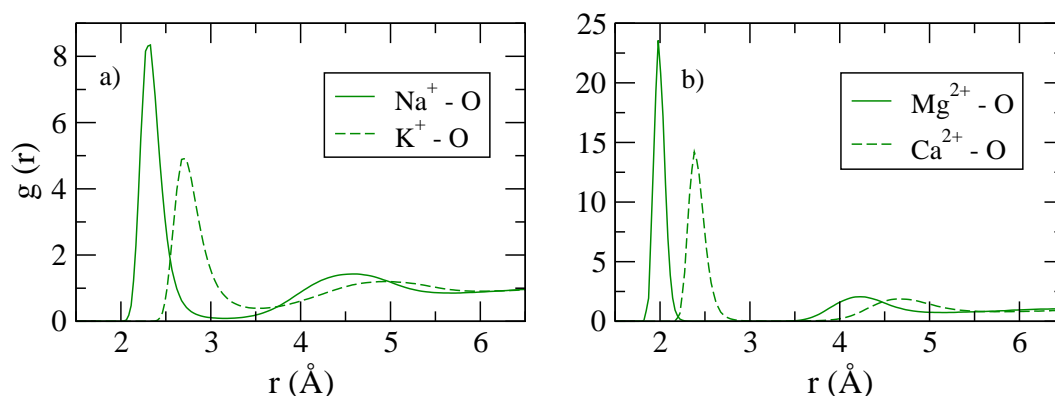


Figure 4.2: Na^+ -, K^+ -, Mg^{2+} - and Ca^{2+} -water oxygen radial distribution functions

The structure of the first hydration shell around the cations is more or less flexible and can take different conformations. In a first approximation, it can be quantified by the coordination number, i.e. the average number of water molecules in direct contact with the cations. The coordination numbers of the cations can be calculated by integrating the cation-water oxygen radial distribution functions (RDFs) from zero up to the first minimum, cf. Figure 4.2. Calculated coordination numbers of Na^+ , K^+ , Mg^{2+} and Ca^{2+} are here found to be 5.7, 6.7, 5.8 and 7.9. These coordination numbers are averages over different possible water coordination geometries. A more detailed picture of the

hydration structure around the cations is provided by the probability distributions of the instantaneous coordination numbers, shown in Figure 4.3. Comparing the two monovalent cations it is clear that the first hydration shell of K^+ is much more flexible than that of Na^+ , allowing for a wider range of possible conformations. Na^+ , on the other hand, can be seen to be rather restrictive in its coordination with water, favoring 5- or 6-fold coordinations. These results are in agreement with the simulation studies by Carillo-Tripp *et al.* [167] using polarizable potentials. The divalent cations show even lower coordination flexibilities. Ca^{2+} strongly favor a 8-fold coordination, whereas Mg^{2+} coordinate exclusively 5 or 6 water molecules.

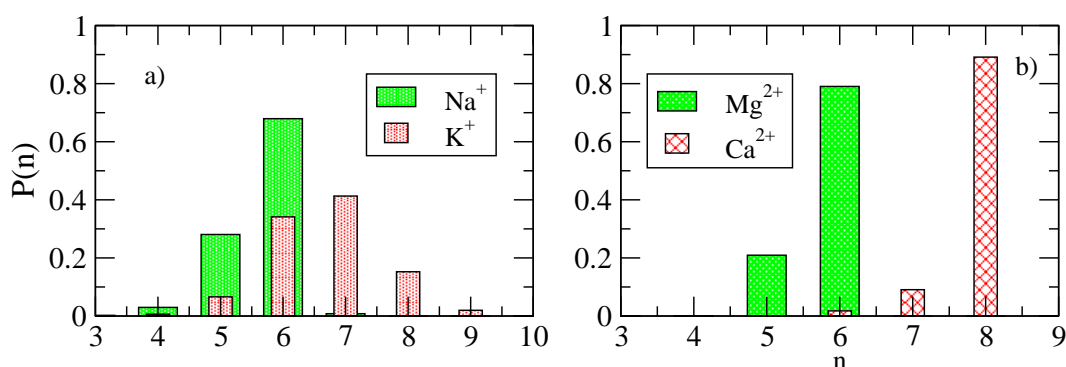


Figure 4.3: Probability distributions of Na^+ -, K^+ -, Mg^{2+} - and Ca^{2+} -water oxygen instantaneous coordination numbers.

The geometry of the cations and their coordinated water is investigated from properties of Voronoi polyhedra [66] constructed around the ions. The asphericity parameter [71] describes the shape of the ion Voronoi polyhedra (see Chapter 2.2.4). It is unity for a sphere and takes values of 1.21, 1.32, 1.65, 1.91 and 3.31 for an icosahedron, a dodecahedron, an octahedron, a cube and a tetrahedron, respectively. The distribution plots of the asphericity parameters for Na^+ and K^+ in Figure 4.4 show a narrow distribution around a close to cubic shape of the Na^+ Voronoi polyhedra, thus indicating an on average close-to-octahedral shape of Na^+ together with the, on average, close to six coordinated water molecules. Larger deviations can be seen for K^+ from any standard coordination polyhedron. The distributions of the asphericity parameter of Voronoi polyhedra calculated around the divalent cations show two narrow peaks around 1.92 and 1.62 for Mg^{2+} and Ca^{2+} respectively. These correspond to almost perfect cubic and octahedral geometries of the ion Voronoi polyhedra. Thus, similar coordination geometries for the monovalent cations and their divalent neighbors can

be seen. Stronger cation-water interactions do, however, lead to more rigid hydration structures of the divalent cations. It is interesting to note that the octahedral coordination geometry found in these force field simulations is the same as that found in DFT calculated gas-phase Mg^{2+} -water clusters calculated by Lightstone *et al.* [179].

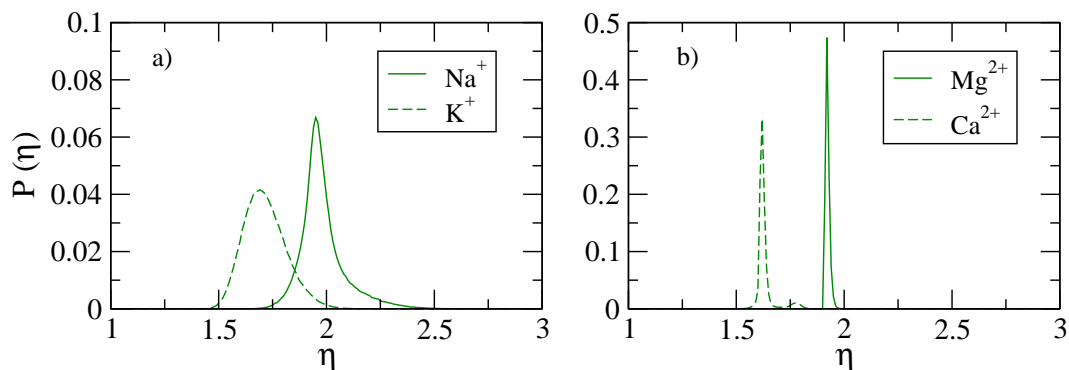


Figure 4.4: Probability distributions of asphericity parameters for a) Na^+ , K^+ and b) Mg^{2+} and Ca^{2+}

Coordination numbers provide an important link between experiment and simulations. However, difficulties involved in extracting radial distribution functions and coordination numbers from scattering data have resulted in measured coordination numbers in the range of 4–8 for both Na^+ and K^+ [180]. An alternative source for structural ion hydration data for force field validation is simulations using ab initio approaches [180–182]. Compared to the calculated coordination numbers from these studies, which are in the range 4.3–5.2 for Na^+ and 5.8–7.0 for K^+ , the force fields used in this study tend to overestimate the coordination numbers. The calculated coordination number of 5.8 for Mg^{2+} is in good agreement with the experimental value of 6.0 [183] which have also been calculated from first principle MD simulations [179] and simulations using polarizable force fields [184]. For Ca^{2+} , experimental coordination numbers are in the range of 6–10 [185], whereas recent computer simulations using ab-initio MD and standard MD with polarizable force fields agree on a coordination number of approximately 7.3 [178, 184]. The calculated coordination number of 7.9 for Ca^{2+} is well within the range of experimental data, but is overestimated according to the previously mentioned computational studies. It should be noted that although the coordination numbers of the divalent cations are in fairly good agreement with the simulations using higher levels of theory, larger differences can be seen in the cation-water radial distribution functions. For the divalent cations, polarization of the surrounding water

becomes more important and standard force fields tend to model a too structured first hydration shell, as seen in higher and narrower first peaks in the radial distribution functions, compared to simulations accounting for polarization of both the cations and the surrounding water. Thus, the rigidity of the first hydration shells of the divalent cations, as seen in Figure 4.4, may be overestimated by these force fields.

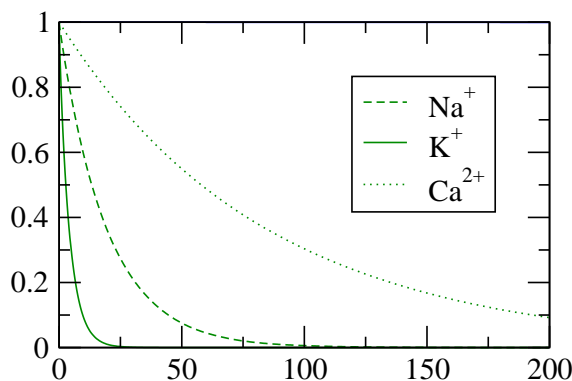


Figure 4.5: Survival time autocorrelation functions for water molecules in the first hydration shell around Na^+ , K^+ and Ca^{2+} cations.

Equally important for the study undertaken here is ion kinetics, in particular exchange times for water in the hydration shells of the cations. Calculated survival time correlation functions (see Chapter 2.2.3) for water in the first hydration shell of the cations are shown in Figure 4.5. Water molecules are considered to be in the first hydration shell of a given cation when the water oxygen atom is within a distance corresponding to the first minima of the radial distribution functions in Figure 4.2. Average residence times are calculated by integrating the time correlation functions from zero to infinity. Calculated first hydration shell water residence times and cation self-diffusion coefficients are presented in Table 4.1. The dynamics of the cations and the hydration water reflect the fundamental difference in hydration properties of Na^+ and K^+ . The chaotropic nature of K^+ is reflected in shorter residence times of hydration water and a higher overall mobility. The hydration shell of the weak kosmotrope Na^+ is by comparison relatively rigid, and lowers the mobility of Na^+ . Compared to the experimental self-diffusion coefficients at infinite dilution, 1.334 and $1.957 \times 10^{-5} \text{cm}^2 \text{s}^{-1}$, for Na^+ and K^+ respectively, [186], the mobilities of both cations are overestimated. However, this is likely due to the overly diffusive TIP3P water model [187] rather than an indication of problems with the ion force fields. The calculated residence times for first shell hy-

dration water around both cations are overestimated compared to experimental values [188], but are in reasonable agreement with other simulation studies using different force fields [177, 189].

Residence times of water in the first hydration shell of the divalent ions are much longer. The survival time-correlation function of hydration water of Mg^{2+} is not included in Figure 4.5 since it is close to unity for the time period presented in the figure. These simulations suggest average residence times of first hydration shell water of Mg^{2+} in the order of nanoseconds, which is much larger than calculated residence times of hundreds of picoseconds found for Mg^{2+} hydration water in the simulations of Jiao *et al.* [184] using polarizable force fields. The average calculated residence time of hydration water of Ca^{2+} is 84.6 ps from these simulations. Thus, a clear correlation between increased water-cation interaction strengths and stability of the ion-first hydration shell complex can be seen, in agreement with the increasing kosmotropic character in the series K^+ , Na^+ , Ca^{2+} and Mg^{2+} . Calculated self-diffusion coefficients from average mean square displacements for the two divalent cations are presented in Table 4.1. Here the largest deviations from experiment can be seen. Experimental values are 0.71 and $0.79 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ compared to calculated values of 1.96 and $2.15 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ for Mg^{2+} and Ca^{2+} respectively [190]. The cause of this discrepancy is not clear from these simulations, but might again be related to the highly diffusive TIP3P water.

Table 4.1: Ion-water solvation data from our MD simulations.

	$r(\text{\AA})$ ^a	ϵ (kcal/mol) ^a	n_{coord} ^b	D ($10^{-5} \text{ cm}^2 \text{ s}^{-1}$) ^c	τ_{res} (ps) ^d
Na^+	2.7275	0.0469	5.7	1.43	19.5
K^+	3.5375	0.0870	6.7	2.28	4.1
Mg^{2+}	1.5734	0.87504	5.8	1.96	-
Ca^{2+}	2.6522	0.44966	7.9	2.15	84.6

^a Cation ionic van der Waals parameters.

^b Average cation-water coordination numbers, calculated from the radial distribution functions in Figure 4.2.

^c Self-diffusion coefficients, calculated from the average mean square displacement of respective cation.

^d Average residence times for water molecules in the first hydration shell of the cations, defined by a spherical volume with a radius corresponding to the distance of the first minima of the cation-water RDFs in Figure 4.2.

These results show that, although using simple models, many important cation hydration properties can be reproduced with these force fields. The neglect of polarization

effects seems to be a reasonable approximation for, at least, the alkali cations. For the alkaline earth metals, the force fields tend to over-structure the cation-first hydration shell complex.

When studying systems at higher ion concentrations, which is needed for example for a statistical analysis of ion distributions around biomolecules, proper descriptions of cation-anion interaction are of equal importance as ion-water interactions. Since ion force fields are typically derived to reproduce hydration energetics of an ion-species at infinite dilution, ion-ion interaction is modeled very differently in different force fields. An example of this can be seen in Figure 4.6, which shows Na^+ - and K^+ - Cl^- radial distribution functions calculated with ion force fields of Beglov and Roux and from reference simulations with the AMBER PARM99 force field [191]. There is a striking difference in how these two force fields model cation-anion interactions. The AMBER force fields show a much larger tendency to form cation-anion pairs, or even clusters, in solution. Further, the AMBER force fields show a difference of almost one order of magnitude in the height of the first peaks of the Na^+ - and K^+ - Cl^- radial distribution functions in Figure 4.6. The Beglov and Roux force fields show first peak heights in much better agreement with the first principle simulations of Cavallari *et al.* [182]. Large discrepancies in ion-ion interactions have previously been observed in the force field comparison of Patra *et al.* [192] and for comparative analysis of ion affinities, large differences in cation-anion pairing tendencies will be projected onto the results, making many ion force fields unsuitable for the study undertaken here.

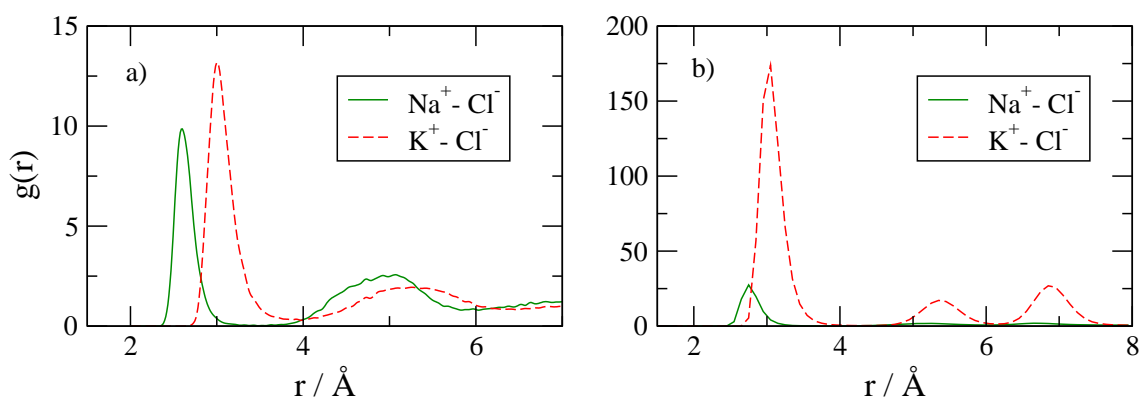


Figure 4.6: a) Na^+ - and K^+ - Cl^- radial distribution functions calculated with Roux ion force fields. b) with AMBER PARM99 force field

4.3 Saccharide ion affinities

4.3.1 Man₆GlcNAc₂-Asn-cation interaction

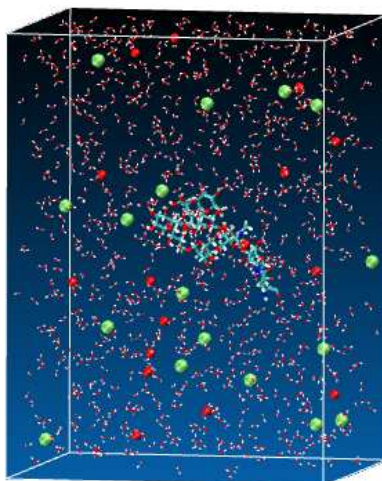


Figure 4.7: Snapshot of the Man₆GlcNAc₂-Asn/water/NaCl simulation box. Water density is reduced for clarity. Sodium and Chloride ions are represented as red and green spheres

In this section, we investigate affinities of Na⁺, K⁺, Mg²⁺ and Ca²⁺ cations to a single Man₆GlcNAc₂-Asn oligosaccharide. A pictorial description of the simulation box is shown in Figure 4.7. For all four systems an ion concentration of 0.15 M is used. Distribution functions of Na⁺ and K⁺ around the oligosaccharide are shown in Fig. 4.8a. The distribution functions are calculated with respect to the closest non-hydrogen atom of the saccharide, and normalized with respect to the bulk density of the ions, as described in Chapter 2.2.1. Both distribution functions show a distinct first peak located at a distance equivalent to the first peak of the cation-water distributions (see Figure 4.2a). Here, this peak represents the case where one or more ion-coordinated water molecules have been replaced by contacts with the polar oxygens (or the 2-nitrogen atom of the *N*-acetyl-D-glucosamine residues) of the saccharide. A second broader peak is also visible for both cations which corresponds to water mediated saccharide interactions, i.e. the cation and the saccharide are connected via a water molecule of the cation hydration shell, which in turn is hydrogen bonded to the saccharide. The different radial positions of these two peaks for Na⁺ and K⁺ simply reflect the different sizes of these

ions. However, the differing relative heights of the first to the second peak indicate a different behavior of Na^+ and K^+ as they approach the saccharide: Apparently, K^+ has a greater probability of forming direct contacts with the saccharide, while Na^+ has a higher probability of forming water-separated saccharide contacts.

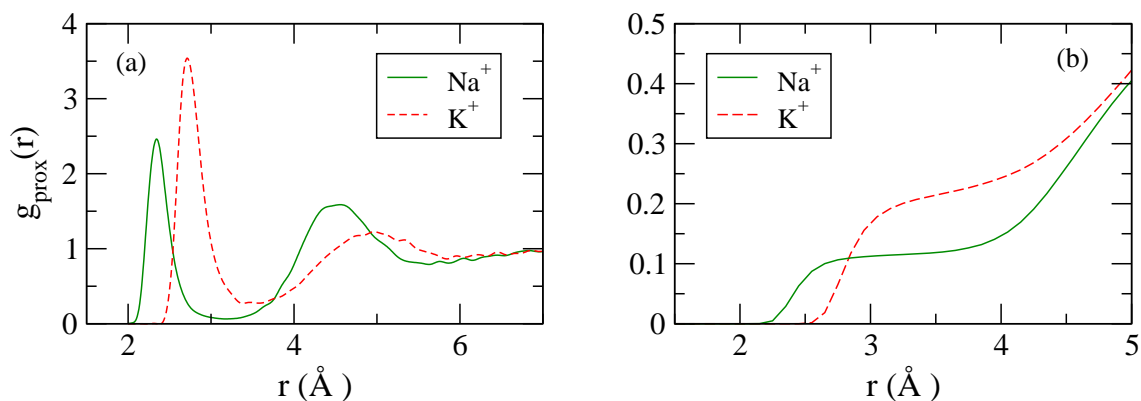


Figure 4.8: (a) Normalized $\text{Man}_6\text{GlcNAc}_2$ -cation proximal distribution functions, calculated with respect to the closest non-hydrogen atom in the saccharide. (b) The corresponding cumulative sums (integrated distribution functions).

The result of this difference in the ion distributions can be further interpreted with the cumulative sums of Figure 4.8b. The cumulative sums show the average numbers of ions per frame, within a given distance of the saccharide. Again, the different radial “onset” of the two curves merely reflects the different ion sizes. More interesting is the difference in the height of the first step. Since the different volumes of the two ions have been taken into account (see above), the data for the two ions are directly comparable. Hence, it is clear that on average there are approximately twice as many potassium ions in direct contact with the saccharide.

The differences observed here can partly be related to the differences in hydration properties of the cations shown in Section 4.2. There will obviously be a higher energetic barrier in the case of Na^+ to remove a coordinated water molecule and replace it with a saccharide contact. The observed differences in saccharide-cation affinities do, however, also depend on the energetic cost of removing hydration water around the saccharide, as well as, on the energetics of direct saccharide-ion interactions. As mentioned in the introduction, ion affinities vary between different bio-ligands (proteins, lipids, carbohydrates) and also vary with e.g. the electrostatic character of the ion binding sites and their steric propensities. Collins [185] recently proposed the so called law of matching

water affinities, which states that the degree to which oppositely charged ions form ion-pairs in solution is determined by the difference in hydration energies between the cation and anion. Forming inner sphere contacts between two ions is only energetically favorable when the cost of removing coordinated water molecules is compensated by the newly formed cation-anion and water-water interactions. Within this theory, the formation of kosmotrope cation and anion pairs is driven by the energetic gain from cation-anion interactions, whereas chaotropic cation and anion pairing is driven by the energetic gain from water-water interactions formed by released water molecules [170]. For the relatively strongly hydrated biologically common phosphate and carboxylate anions, the law of matching water affinities predicts a greater affinity for Na^+ over K^+ , which is supported by the studies of Vrbka *et al.* [169] and Cheng *et al.* [168]. Similar arguments can be applied to the polar oxygen and nitrogen atoms of uncharged carbohydrate moieties, which can be considered to be weakly hydrated anions of low charge density, and would have a greater tendency of pairing with K^+ . This trend can also be seen in ion affinities to the two different types of oxygen atoms of the saccharide. Figure 4.9 shows the proximal distributions and cumulative sums around the hydroxyl oxygens and the glycosidic and endocyclic oxygens of the saccharide. The relative difference in ion affinities is indeed larger around the lower charge density glycosidic and endocyclic oxygens.

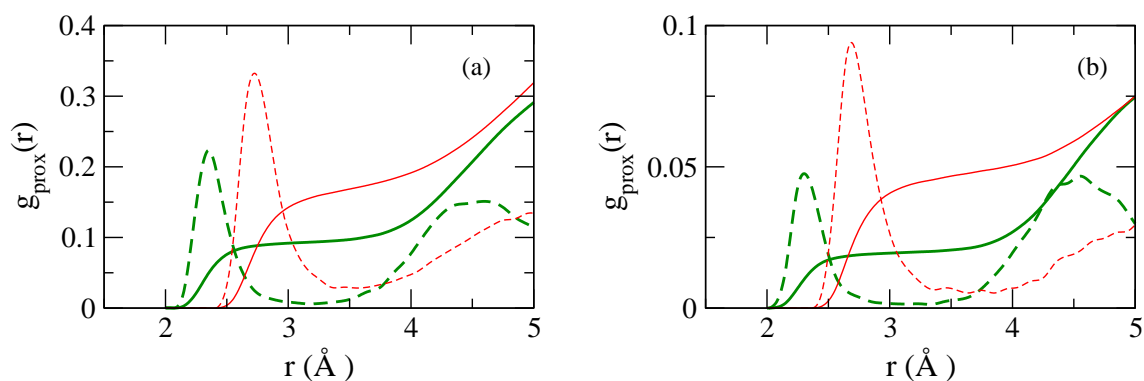


Figure 4.9: Proximal distribution functions (dashed lines) and cumulative sums (solid lines) for Na^+ (thick lines) and K^+ around $\text{Man}_6\text{GlcNAc}_2$ hydroxyl oxygen atoms (a) and endocyclic and glycosidic oxygen atoms (b). The distribution functions are multiplied by a constant factor for a clearer representation.

The corresponding normalized proximal distribution functions for the divalent cations Ca^{2+} and Mg^{2+} are shown in Figure 4.10. The average cation-water oxygen distance is,

as seen in Figure 4.2b, approximately 2 Å and 2.4 Å for Mg^{2+} and Ca^{2+} respectively. As seen in Figure 4.10, not a single inner sphere contact between the saccharide and Mg^{2+} can be seen in the whole 60 ns simulation, and only a very small number of direct saccharide- Ca^{2+} contacts. However, the distributions of both cations show distinct peaks at distances corresponding to water separated contacts. This shows a clear correlation between the charge density and the behavior of the cations as they approach the saccharide surface. The tendency of the cations to form direct, inner sphere contacts with the saccharide decreases with increasing charge density and stronger hydration in the series $\text{K}^+ > \text{Na}^+ > \text{Ca}^{2+} > \text{Mg}^{2+}$. For the divalent cations, these results indicate that interactions with polar non-charged saccharides almost exclusively takes place through water-mediated contacts. Again, the observed ion affinities can be seen to change with the electrostatic character of the interacting moieties. Ca^{2+} readily binds to proteins, mainly through interaction with side-chains containing deprotonated carboxyl groups, but also through carbonyl and hydroxyl groups with decreasing affinity [193]. The strongly hydrated Mg^{2+} has been found to form a mixture of inner and outer sphere contacts with both carboxylate and phosphate anions [193–195]. Thus, the interaction of divalent cations with carbohydrates containing charged groups may differ from what is seen here with neutral groups. Carbohydrate-carbohydrate interactions in cellular adhesion processes are, for instance, believed to be stabilized by Ca^{2+} bridging charged carbohydrate residues containing deprotonated carboxylate groups [196].

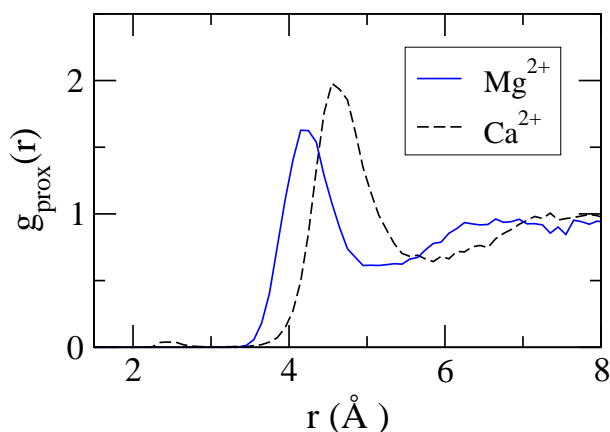


Figure 4.10: Normalized proximal distribution functions for Mg^{2+} (solid line) and Ca^{2+} (dashed line), calculated with respect to the closest non-hydrogen atom in $\text{Man}_6\text{GlcNAc}_2$.

4.3.2 Dependence of ion affinities on saccharide density

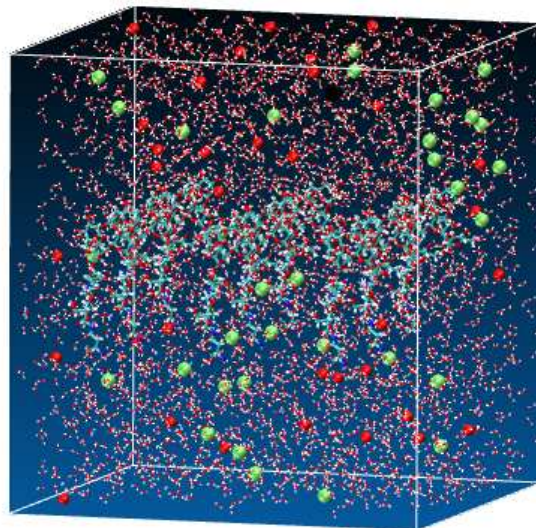


Figure 4.11: Initial configuration of the $\text{Man}_6\text{GlcNAc}_2$ oligosaccharide array in 0.15 M KCl solution. Water density is reduced for a clearer representation. Potassium and Chloride ions are shown as red and green spheres.

We have investigated structural and dynamical details of alkali cation-saccharide interactions, and also the effect of carbohydrate density on these properties. For this purpose, additional simulations were performed on systems mimicking a dense saccharide surface at the same ion concentrations (0.15 M NaCl or KCl). Nine $\text{Man}_6\text{GlcNAc}_2$ -Asn units were placed in a regular arrangement, each oligosaccharide equally spaced at a distance of approximately 20 Å, with main-chain atoms of the asparagine residues restrained to an imaginary surface, as depicted in Figure 4.11. At this distance, the saccharides are close enough to interact with one another, forming a dense carbohydrate layer. This is meant as a zero-order cartoon of a possible situation in the glycocalyx, acknowledging the facts that neither the detailed composition of the oligosaccharides in the glycocalyx nor even their distances are known experimentally.

Distribution functions of the the alkali cations and the alkaline earth metal cations around the saccharide array are shown in Figures 4.12 and 4.13 respectively. Comparing the cumulative sums of the Na^+ and K^+ distribution functions in Figure 4.12b with those in Figure 4.8b, a smaller density difference between the cations can be seen

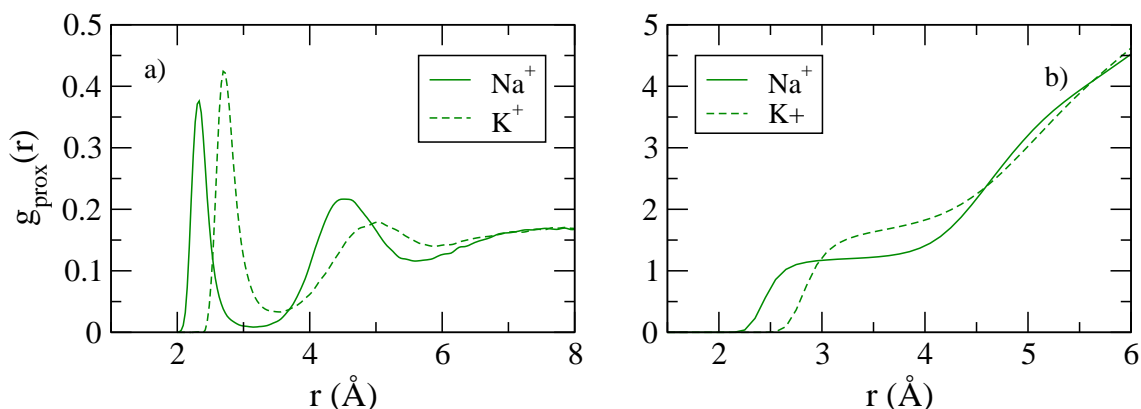


Figure 4.12: a) Unnormalized proximal distribution functions for Na^+ (full lines) and K^+ (dashed lines) calculated with respect to the closest non-hydrogen atom in the $\text{Man}_6\text{GlcNAc}_2$ saccharide array. b) The corresponding cumulative sums.

around the saccharide array, compared to the single oligosaccharide. Although there are on average more potassium ions in the vicinity of the saccharide surface, the relative difference between Na^+ and K^+ is smaller in the saccharide array. The distribution functions of the divalent cations are similar to those of the single saccharide simulations. Also here, not a single Mg^{2+} -saccharide contact can be seen for the whole 60 ns trajectory. Calcium cations show an slight increased tendency to pair with polar saccharide atoms in the saccharide array.

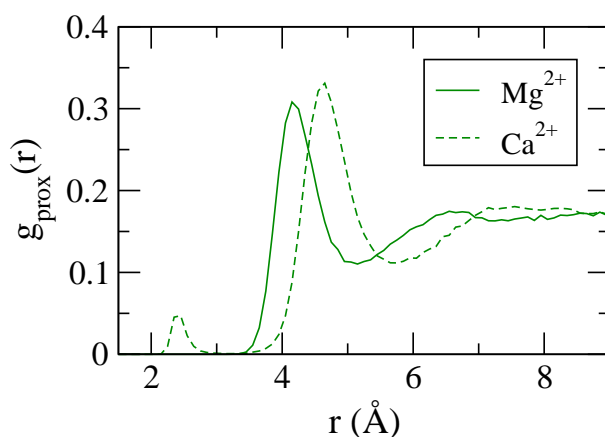


Figure 4.13: Unnormalized proximal distribution functions for Mg^{2+} (full lines) and Ca^{2+} (dashed lines) calculated with respect to the closest non-hydrogen atom in the $\text{Man}_6\text{GlcNAc}_2$ saccharide array.

Thus, we observe, at least for the alkali cations, what seems to be a dependence on carbohydrate density on ion affinities. To check that these results are not a consequence of poor statistics, due to the smaller number of saccharide-cation contacts in the single oligosaccharide simulations, these simulations were extended another 30 ns. Since the results presented in the previous section did not change significantly, this is likely to be a relevant observation. In an attempt to understand the chemistry behind this phenomenon, we have taken a closer look at details about saccharide-cation interactions for the two alkali cations. The probability distributions of ion-saccharide coordination numbers for saccharide bound sodium and potassium ions, calculated for the single $\text{Man}_6\text{GlcNAc}_2$ oligosaccharide (S) and the saccharide array (SA) are shown in Figure 4.14. Ion-saccharide coordination numbers are calculated as the number of

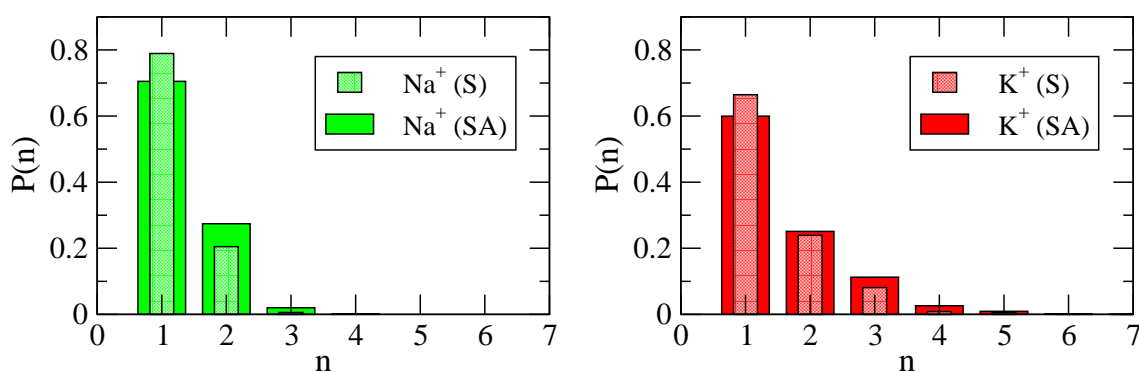


Figure 4.14: Probability distributions of cation-saccharide coordination numbers for Na^+ (left) and K^+ (right) bound to $\text{Man}_6\text{GlcNAc}_2$ (S) or to the array of $\text{Man}_6\text{GlcNAc}_2$ saccharides (SA).

saccharide oxygens and nitrogen within a distance corresponding to the first minimum of the cation-saccharide distribution functions (Figure 4.2a). The maximum numbers of ion-saccharide contacts observed in these simulations are 4 and 7 for Na^+ and K^+ , respectively. Although the interaction mainly takes place through a few saccharide contacts for both cations, the larger size and a weaker and more flexible hydration shell of K^+ also allows for higher saccharide coordination numbers, where K^+ is almost fully solvated by the saccharides. A snapshot of the trajectory showing a seven-fold coordinated potassium ion, buried inside an oligosaccharide, can be seen in Fig. 4.15. The geometrical constraints imposed by Na^+ on the coordinated atoms seem to be hard to meet for the saccharides. Hence, regardless of the saccharide density, no more than four-fold coordinated sodium ions are observed. For both cations there is a shift

towards higher coordination numbers with increasing saccharide density, this shift is, however, larger for potassium cations.

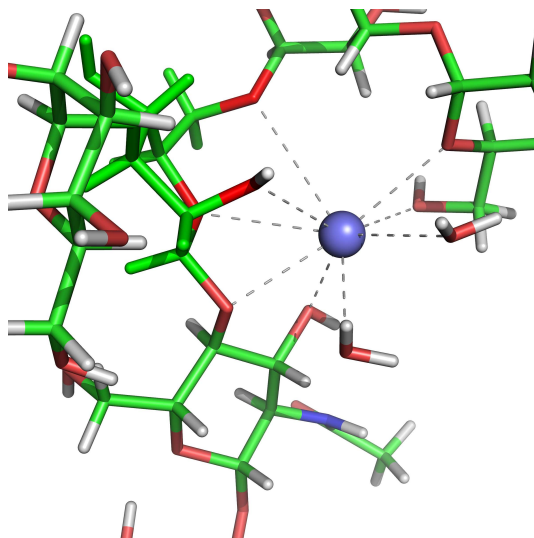


Figure 4.15: Snapshot from the trajectory showing a potassium ion, seven-fold coordinated to the $\text{Man}_6\text{GlcNAc}_2$ saccharide. Two additional coordinations of the potassium ion to water molecules are also shown. For clarity, the cation is drawn smaller than its usual contact radius. Also, only part of the oligosaccharide is included in this graph, the remainder is cut off. The figure was generated using PyMOL[197].

Figure 4.16 shows the "layer survival-time-correlation functions" $C_R(t)$ (see Chapter 2.2.3). $C_R(t)$ gives the probability that an ion remains within a distance to the closest non-hydrogen saccharide atom, corresponding to the distance of the first minima of the proximal distribution functions in Figure 4.2a, i.e., ions bound to the saccharide, for a time period t . The correlation function can be fitted to a double exponential of the form

$$f(x) = A * e^{(-x/\tau_{short})} + B * e^{(-x/\tau_{long})} \quad (4.1)$$

from which the average residence time can be calculated from the long and short decay constants and the corresponding amplitudes. The decay constants and the calculated average residence times of Na^+ and K^+ are shown in Table 4.2. The difference in hydration strengths of the two cations is clearly projected onto their interaction with the saccharide. On average, sodium ions can be seen to bind to the saccharide for longer periods of time, in spite of an on average lower saccharide coordination number. The steeper short time decay of the K^+ survival-time-correlation plots in Figure 4.16 show that K^+ -saccharide contacts are weaker and have a greater tendency to be broken at shorter time scales, compared to their Na^+ analogues. While the probability for higher

coordination numbers increases together with the carbohydrate density, the increase in the residence times for sodium ions is larger. This is clearly the reason of the smaller differences between the two cations seen in the distribution functions and the cumulative sums around the saccharide array. Thus, ion affinities to carbohydrates are therefore expected to depend on the size, shape, and composition of the carbohydrate system, and the resulting coordination possibilities.

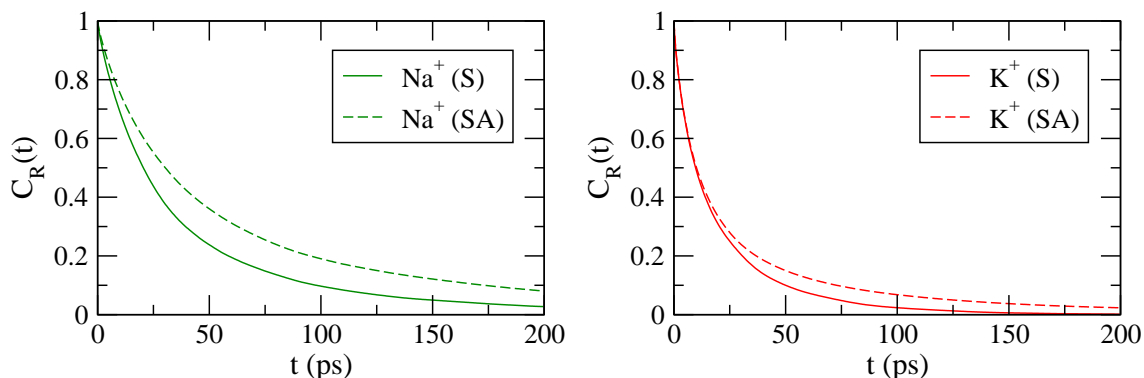


Figure 4.16: a) Survival autocorrelation functions for Na^+ (left) and K^+ (right) coordinated to one or more oxygen or nitrogen atoms of $\text{Man}_6\text{GlcNAc}_2$ (S) (solid lines), or the array of $\text{Man}_6\text{GlcNAc}_2$ saccharides (SA) (dashed lines).

Table 4.2: Temporal characteristics of Na^+ and K^+ interaction with a single $\text{Man}_6\text{GlcNAc}_2$ saccharide (S), or with an array of $\text{Man}_6\text{GlcNAc}_2$ saccharides (SA)

	A^a	τ_{short}^a	B^a	τ_{long}^a	τ_{avg}^b	n_{scoord}^b
Na^+ (S)	0.62	20.3	0.36	75.6	39.3	1.25
Na^+ (SA)	0.55	26.2	0.42	119.3	64.4	1.46
K^+ (S)	0.45	7.65	0.50	31.6	19.4	1.32
K^+ (SA)	0.64	10.8	0.30	69.8	27.7	1.60

^a Short and long decay constants τ_{short} , τ_{long} and the corresponding amplitudes A and B of a double exponential (Eq. 4.1) fitted to the survival-time-correlation-functions in Figure 4.16.

^b τ_{avg} is the average residence time and n_{scoord} is the average cation-saccharide coordination number for saccharide bound ions.

4.4 Ion specific effects on saccharide conformations

Many oligosaccharides display a high degree of conformational flexibility and may not have a well-defined three-dimensional shape, but rather, exist as an ensemble of conformers in solution at room temperature [198]. The overall conformational flexibility is determined by the flexibility of the individual glycosidic linkages, joining neighboring sugar residues in the saccharide. The potential energy surface for rotations around the glycosidic linkages can exhibit multiple minima, sometimes separated only by low barriers [199], allowing for a high conformational flexibility of the whole saccharide on longer time scales, but also conformational rigidity on shorter time scales [200]. In vacuo, transitions between different local minima occur on a time scale of 10-100 ps [201]. In water, transitions occur on much larger time scales, and long simulation times are needed to fully sample the conformational space of an oligosaccharide [33].

To show the effects ions can impose on saccharide conformation and flexibility, we have taken a closer look at the Man(α 1-3)Man glycosidic linkage of the saccharide, connecting residues six and seven in Figure 4.1. The average dihedral angles $\langle\phi\rangle$ and $\langle\psi\rangle$ describing this linkage are -40° and -15° from NMR experiments [199], but computational studies have shown a considerable flexibility of this linkage due to multiple minima on the conformational energy map [199, 202]. Starting with SWEET-derived dihedral angles of -48° and -7.5° for ϕ and ψ , the man(α 1-3)Man the configurational space sampled in our trajectories is restricted to the region around the three lowest, closely located energy minima on the adiabatic Ramachandran energy maps of Naidoo *et al.* [199]. Figure 4.17 shows different properties of the two dihedral angles, ϕ and ψ , for a 95 ps time segment where a sodium ion, dually coordinated to oxygen O6 of residue seven and oxygen O3 of residue six, was observed. As reference, average values from the whole NaCl trajectory, and from a trajectory without any ions, are shown. Figures 4.17c and 4.17d show the dihedral autocorrelation functions calculated as [203],

$$C_A(t) = \langle \cos(\theta(\tau) - \theta(\tau + t)) \rangle. \quad (4.2)$$

Here, $C_A(t)$ shows to which degree the dihedral angle θ at time τ correlates with that at the later time $\tau + t$. From these figures, it can be seen that the sodium ion is able to reduce the flexibility of this particular glycosidic linkage, and stabilize the saccharide in an intermediate conformation for a short period of time. The dihedral distribution plots in Figures 4.17a and 4.17b show that the presence of the cation may stabilize

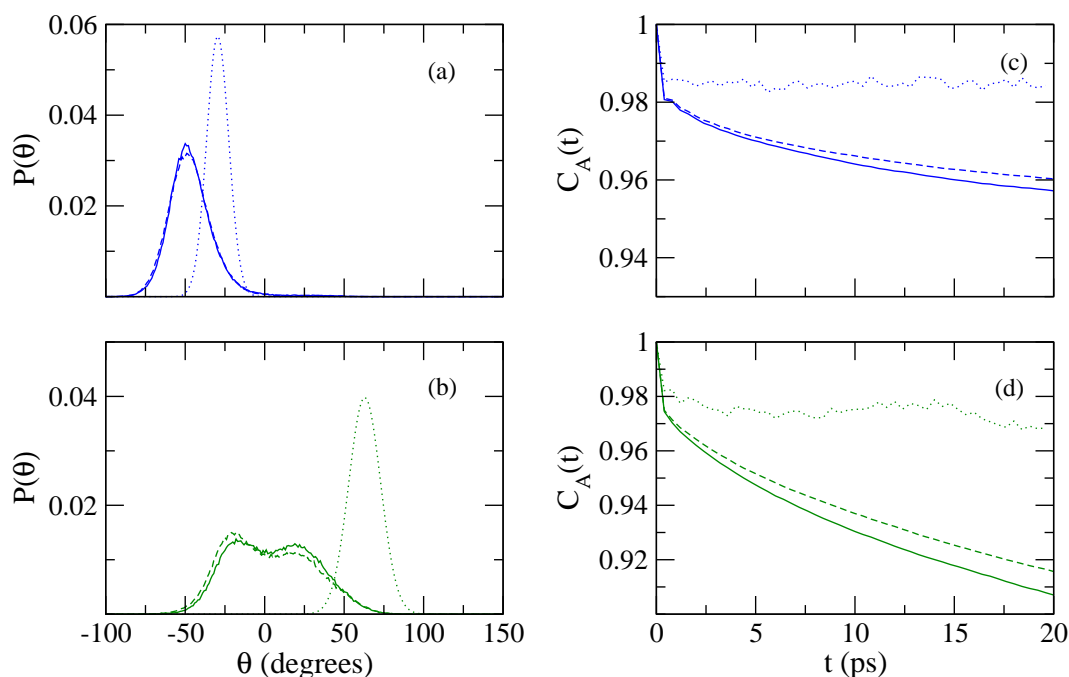


Figure 4.17: Properties of the two dihedral angles ϕ and ψ describing the Man(α 1-3)Man glycosidic linkage connecting residues six and seven of Man₆GlcNAc₂. Results from three different trajectories are shown. A full (60 ns) trajectory in water (full lines), the full 0.15 M NaCl trajectory (dashed lines) and a 95 ps long segment of the NaCl trajectory (dotted lines), where a sodium ion is dually bound to oxygens O3 and O6 of residues six and seven. (a) and (b) show the dihedral angle distributions of ϕ and ψ respectively. For the shorter time segment, the distributions of ϕ and ψ are shown as fitted Gaussian curves. Autocorrelation functions of ϕ and ψ for the different trajectories are shown in (c) and (d).

the angles in conformations that have low probability to occur when compared to the whole trajectory. For example, for this time segment, the ψ angle is shifted to an average value of 62° , compared to 2° for the whole trajectory. Figure 4.18 shows the time evolution of the two dihedral angles before, during and after the interaction with the ion. This figure shows that the conformation of the saccharide clearly is influenced by the cation. During the time the ion is dually bound to the saccharide, in the time interval between 41790 and 41885 ps, the ψ angle is shifted towards higher values, allowing the ion to coordinate with oxygens of both carbohydrate residues, and relaxes as the ion leaves. Similar to the ϕ and ψ distributions of the Man(α 1-3)Man linkage, the other dihedral angles, determining the total conformation of the saccharide, only show minor differences in water and in the ionic solutions. The only exceptions to this trend are the two 1–6 linkages, which can be seen to populate the three possible staggered

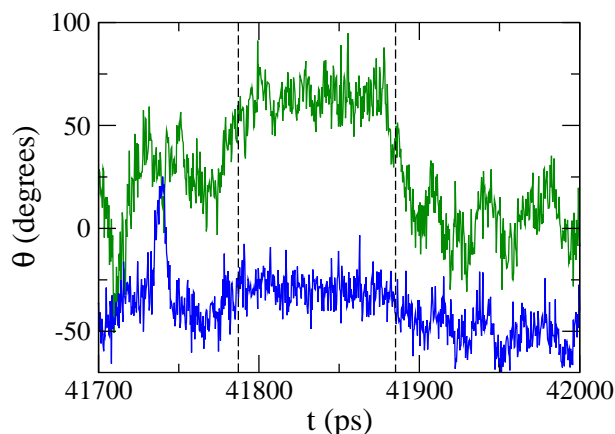


Figure 4.18: Time histories of the dihedral angles ϕ (lower curve) and ψ (upper curve) describing the Man(α 1-3)Man glycosidic linkage connecting residues six and seven of Man₆GlcNAc₂ in 0.15 NaCl solution. A time segment where a sodium ion is dually bound to oxygen atoms of residues six and seven is marked with vertical lines.

conformers, gg, gt and tg [25], differently in each simulation. Using a starting structure closest to the gt conformer, the relative populations of the different conformers, for the Man(α 1-6)Man glycosidic linkage between residues three and six in Figure 4.1, are (100:0:0), (34:63:3) and (81:18:1) in the water, NaCl and KCl simulations, respectively. These differences cannot be attributed to the presence of the ions, but rather show the flexible nature of this particular glycosidic linkage and the long simulation times needed to fully sample the relative populations of the conformers of oligosaccharides. The dihedral angle correlation plots in Figures 4.17c and 4.17d show an on average higher short-time correlation for the angles from the NaCl simulation than in pure water. Similar trends can be seen for several of the other dihedral angles, with the general trend that the correlation is higher for the ion-trajectories compared to the pure water simulations. Thus, the overall effect of the alkali cations on saccharide conformation is small, but the presence of the ions can induce short time rigidity whenever multiple saccharide coordination occurs.

4.5 Discussion

We have studied solvation properties of simple cations of biological relevance (focusing on Na^+ and K^+ , with additional data on Mg^{2+} and Ca^{2+}), comparing the situation close to typical glyocalyx saccharides with that in the bulk, taking a N-linked $\text{Man}_6\text{GlcNAc}_2$ oligosaccharide as a representative glyocalyx unit. We have tested various force field choices and settings, arriving at a combination that is internally consistent and also appears to yield validation results in acceptable agreement with previous theoretical work and with experiment. Hence, we are confident that our results are at least qualitatively correct, and possibly extendable to similar glyocalyx oligosaccharides. Our simulation data for oligosaccharide-cation interactions reveal that K^+ interacts with oligosaccharide oxygen atoms preferentially directly, whereas Na^+ shows an enhanced indirect interaction via an intermediate water molecule (the latter mode is almost exclusively preferred by Mg^{2+} and Ca^{2+}). This agrees with the propensity of Na^+ for holding on more tightly to its first solvation shell water molecules. Accordingly, discounting the indirect water-mediated contacts, there are about twice as many K^+ cations in *direct* contact with the oligosaccharide, compared to Na^+ . The higher flexibility of the first solvation shell of K^+ also allows for up to 7 direct cation-oligosaccharide contacts, while Na^+ can have at most 4, reflecting the restricted ability of the oligosaccharide to accommodate the more stringent steric preferences around Na^+ . Despite the above tendencies favoring oligosaccharide- K^+ contacts, the fewer contacts between the oligosaccharide and Na^+ are longer lived than their K^+ counterparts. Again, this simply reflects the stronger interaction between Na^+ and its first solvation shell: It is harder to induce ligand replacements there, but once it happens the changed situation is again more stable. In our particular oligosaccharide array setup, these opposed tendencies lead to a decreased difference between Na^+ and K^+ residence times close to the oligosaccharide, when compared to the single oligosaccharide solvation situation. In general and in detail, these differential solvation effects of saccharides on Na^+ and K^+ differ from what has been found for other biomolecules (proteins, DNA) (with the reservation that charged groups may change this finding). Clearly, this might be important for glyocalyx functions. Finally, the cations also influence the oligosaccharide conformations. Angular correlations reveal a small but significant overall tendency of Na^+ (and even more so of K^+) to hinder oligosaccharide flexibility. On close inspection of the MD trajectories, the oligosaccharides execute strong but short-lived excursions away from their pure-water preference conformations to temporarily participate in ion solvation.

5 FimH-carbohydrate interactions

5.1 Introduction

Lectins are the class of proteins found in a wide range of organisms, ranging from viruses to animals, with the common ability to recognize and bind carbohydrates with high specificity. In this manner, lectins mediate cell-cell interactions by binding to receptors in the cell surface glycocalyx and are involved in a myriad of biological processes [204]. It has been shown that microorganisms, such as bacteria and viruses, utilize lectin-cell surface carbohydrate interactions to mediate adhesion, as a first step in host colonization upon infection. An important discovery was made in 1979 by Aronson *et al.* [205], showing that urinary tract infection by the bacterium *E. coli* in mice could be inhibited by methyl α -D-mannoside, presumably by blocking the carbohydrate recognition domain (CRD) in the bacterial surface structure and thereby inhibiting adherence to the cell surface. Since then, a lot of research has been dedicated to understanding the mechanisms behind adhesion of *E. coli* to host cell receptors. It is now known that Bacterial adhesion is mediated by the hair-like filaments extracting from the bacterial cell surface, called pili or fimbriae. The type 1 pili, found on the cell surface of *E. coli*, is made up by protein subunits of mainly FimA type, with smaller contributions of FimF, FimG and FimH protein subunits [206]. Abraham *et al.* [207] showed that the FimH subunit, located at the tip of type 1 pili of *E. coli*, contains the carbohydrate recognition domain (CRD), which is responsible for the adhesive function of the bacterium. The structure of the FimH lectin was determined in 1999 by Choudhury *et al.* [9] from the X-ray crystal structure of FimH-FimC complex. FimH is an all-beta protein consisting of two domains connected by a short linker. The CRD is a deep pocket located in the amino-terminal domain of FimH. The structural basis for FimH-mannose recognition was further elucidated by the crystal structure of the FimH-FimC chaperone-adhesin complex with a bound D-mannopyranoside in the CRD of FimH, determined by Hung

et al. [208]. This study showed that the FimH binding pocket is perfectly designed to accommodate and form tight interactions with D-mannose through extensive hydrogen bonding and additional hydrophobic interactions. Further X-ray experiments and affinity measurements have revealed carbohydrate based ligands with increased binding affinity over a single mannose monosaccharide [209, 210]. These studies have increased the understanding of interactions between FimH and its natural cell surface glycoprotein ligands and also provided information about preferred structures for potential anti-adhesion drugs.

Computer simulations can provide additional information of FimH-carbohydrate interactions over the static picture given by the X-ray structures from experiment. Structural and energetic details of the complex can be calculated from an ensemble of structures collected from, for instance, a MD-simulation. The statistical thermodynamics framework can be used to connect molecular mechanics energetic changes, associated with the binding process, to macroscopical free energy changes measured in experimental binding affinity measurements. Computer simulations can, in this manner, be used to probe the structural basis of binding at a microscopical scale and, at the same time, how changes in the structure of the ligand are reflected in macroscopical binding free energy changes.

The following sections present results from a series of explicit water molecular dynamics simulations of the lectin FimH and different of mannose-based ligands. The structure of each complex is investigated in detail and connected to binding free energy changes of respective ligand using the MM-PB(GB)SA methodology. These studies aim at giving a deeper understanding of FimH-carbohydrate recognition and aid the design of new carbohydrate-based anti-adhesion drugs, as well as, artificial carbohydrate receptors.

5.2 System and simulation protocol

The starting structure of FimH was taken from the FimH-FimC chaperone complex with D-mannose from the Protein Data Bank (1KLF). FimH chain B together with crystallographic water within 6.0 Å of the protein was used as the initial structure of the protein. In order to maintain a reasonable system size and to lower the statistical errors of the calculated results, only the amino-terminal FimH domain (residues 1-158) were used to represent FimH in these simulations. Reference simulations with the full

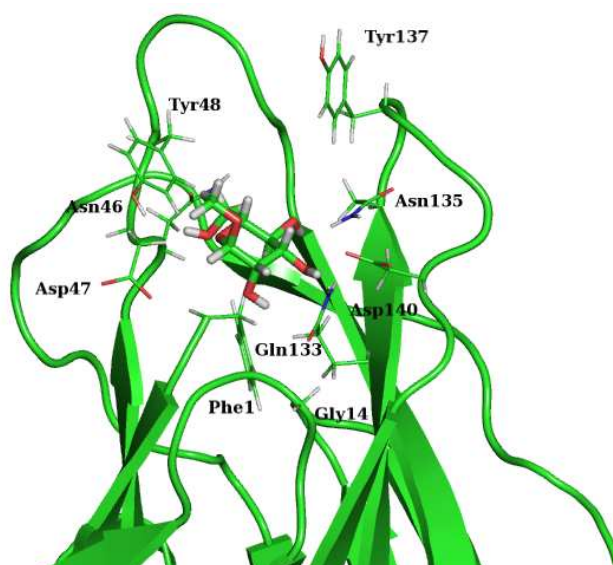


Figure 5.1: Pictorial description of the binding pocket of FimH. The structure is taken from the FimH-FimC chaperone complex with D-mannose from the Protein Data Bank (1KLF)

FimH protein were also performed to validate that this approximation did not induce any structural alterations of the FimH binding pocket. Cysteine residues 3 and 44 were assumed to be connected by a disulphide bond and the nitrogen of residue His45 was assumed to have both nitrogen atoms protonated [211]. Carbohydrate ligands were added to the FimH binding pocket by superposing terminal α -D-mannosyl residues onto the mannose in the PDB-file 1KLF. Starting conformations of the ligands were taken from the SWEET database [173]. To avoid steric clashes with FimH side chains, slight alterations of glycosidic torsional angles were done whenever necessary. Standard classical molecular dynamics simulations were performed for FimH, the ligands and the FimH-ligand complex systems using the NAMD2.6 [36] molecular dynamics program. Force field parameters for FimH were taken from the AMBER PARM99 force field [43]. Carbohydrates were modeled by the GLYCAM04 force field [24] and water by the TIP3P water model [44]. An inherent problem with this combination of force fields is the different 1-4 non-bonded van der Waals and electrostatic scaling used in the PARM99 and GLYCAM04 force fields. In order to correctly model the structure of the FimH binding site, we used the PARM99 scaling factors for 1-4 non-bonded van der Waals and electrostatic interaction for all simulations in this study. The PME [49] method was used to account for long-ranged electrostatic interactions in the periodic

system. A cutoff of 12 Å was used for van der Waals interactions, without using any switching functions. All simulations were performed at a constant temperature of 300 K, using a Berendsen thermostat [51] with a weak coupling coefficient of 5 ps. A constant pressure of 1.01325 bar was applied using a Berendsen pressure bath coupling with a rescaling frequency of 8 fs, a barostat relaxation time of 100 fs and a compressibility of $4.57 \times 10^{-5} \text{ bar}^{-1}$. All heavy atom-hydrogen bond lengths were restrained by the SHAKE algorithm [55]. A multiple time step approach was used in the time-propagation of the system. A time step of 2 fs was used for forces from immediate and intermediate neighboring atoms. Long range forces were updated every 4 fs. Simulation lengths are in all cases in the range of 23-25 ns with snapshots stored every 1 ps, giving 23000-25000 conformations over which the presented results are averaged. In total, nine systems were simulated. Four of these systems are the four ligands in Figure 5.2, each in an explicit water box. The size of the water boxes were chosen so that the closest distance in each dimension to the closest periodic image is at least 30 Å from the starting conformation. The number of atoms in these simulations are in the range of 11-19000 atoms. Five additional systems, consisting of each ligand in complex with FimH and FimH alone, were also simulated. In these cases, the water box was chosen to be large enough for the solute to rotate freely in each dimension without interacting with its periodic images. The number of atoms in these systems are approximately 75000 atoms.

5.3 Binding free energy calculations

Calculating the free energy change of binding for a given complex is a non-trivial task. The accuracy of the calculated results depend on both the molecular mechanics potential energy function and the sampling of the phase space of the system. On a microscopical scale, lectin-carbohydrate binding affinities are, to a great extent, determined by intermolecular non-covalent interactions between the two molecules in the complex. Hydrogen bonding and van der Waals interactions are recognized as being responsible for the main attractive forces in receptor-ligand interactions and need to be modeled properly by the force fields. Furthermore, energetic changes in the surrounding water need to be considered when water molecules are transferred from the vicinity of the solute binding interface to bulk water. Several different approaches have been

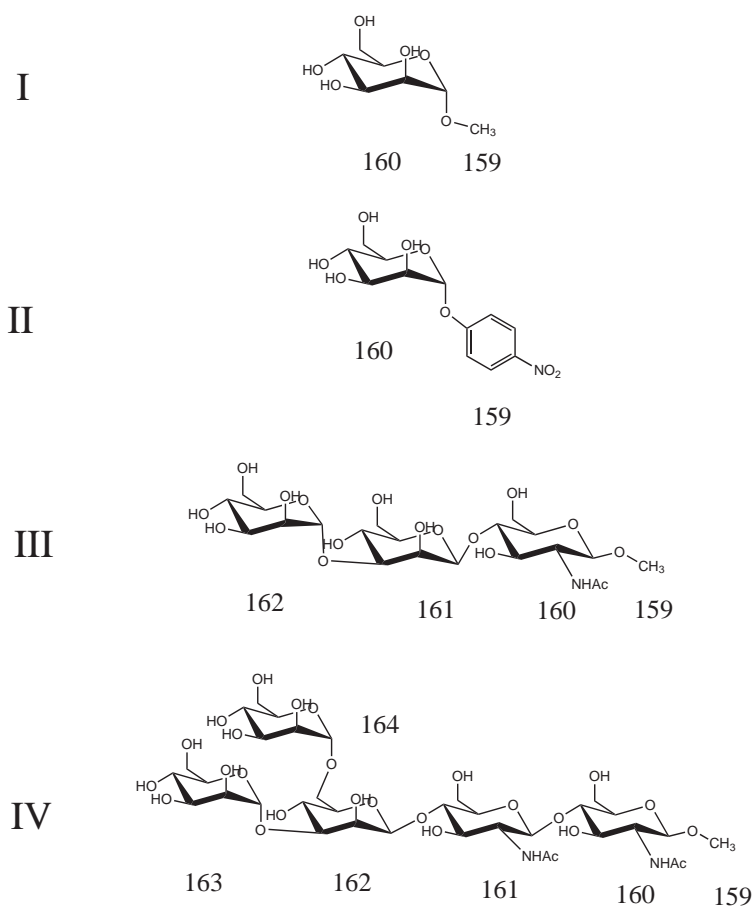


Figure 5.2: Structures of the four ligands investigated in this study. I - $\text{Man}\alpha\text{Me}$. II - $\text{Man}\alpha\text{-p-Nitrophenyl}$. III - $\text{Man}\alpha1,3\text{Man}\beta1,4\text{GlcNac}\beta\text{Me}$. IV - $\text{Man}(\alpha1,6)[\text{Man}(\alpha1,3)]\text{Man}(\beta1,4)\text{GlcNac}(\beta1,4)\text{GlcNac}\beta\text{Me}$.

developed for calculating free energy changes of binding from the microscopic energetic changes calculated from MD-simulations [6]. Free energy perturbation (or thermodynamic integration) has successfully been applied to calculate $\Delta\Delta G_{bind}^0$ for different, but chemically similar, ligands. Given that simulation times are sufficient, this method can reproduce free energy changes within experimental accuracy. For this study, we investigate structurally very different ligands and the free energy perturbation approach would require very long simulation times. Instead, standard MD-simulations are performed on the complexes and FimH and the ligands in the free state, and the MM-PB(GB)SA approach [212], as implemented in the AMBER molecular dynamics suite, was used to post-process the calculated trajectories to estimate binding affinities of the complexes. The theory behind the MM-GBSA method is described in Section 2.2.6. In practice,

snapshots of the complex, the ligand and the receptor are collected from explicit water simulations. The free energy change of binding is calculated, using a thermodynamic cycle, as the sum of differences in solute molecular mechanics potential energies plus solvation free energy contributions, averaged over all stored snapshots. The free energy change related to solvent reorganization is treated, either by the Poisson-Boltzmann (PB), or the Generalized Born (GB) implicit solvation model. The Generalized Born model of Onufriev et al. [83] is used in this study, which has been designed to better describe interior regions of large macromolecules. The radius of hydrogen atoms bonded to nitrogen atoms is, for these calculations, set to 1.3 Å as suggested by the AMBER8 manual. General Born solvation free energies are calculated using a dielectric constant of unity for the interior of the solute molecules and a constant value of 80 in the exterior regions. The boundary between the solute and the solvent was determined by a probe with a radius of 1.4 Å. The non-polar contribution to the solvation free energy for each solute species is calculated using parameters $\gamma=0.005$ kcal/mol and $b=0.0$ kcal/mol [213]. The solvent accessible surface area is calculated using either the LCPO [214] method or the ICOSA method. The latter method is only used for the free energy decomposition calculations. Solvation free energies calculated with the PB method are using the same dielectric constants, but different parameters for the non-polar solvation free energy contribution. Here, the parameters $\gamma=0.00542$ kcal/mol and $b=0.92$ kcal/mol are used [213].

Two different approaches can be taken when performing the MM-PB(GB)SA free energy analysis. Either structures of the complex, receptor and ligand are taken from three separate simulations, or the solute structures are extracted from the complex trajectory. Using complex structures from a single trajectory has the advantage that statistical uncertainties, due to insufficient configurational sampling of the molecules, are minimized. The drawback of this approach is that the calculated free energy change of binding processes involving conformational changes in the receptor or ligand structures can be overestimated, since internal energetic penalties are neglected. Both approaches are used in this study and are discussed further in later sections of this chapter.

The MM-PB(GB)SA method gives the free energy change of binding, without taking contributions from entropy changes in the receptor and ligand into account. A number of different methods, discussed in Section 2.2.7, are used in the following sections to estimate entropy changes of the complex relative to the free receptor and ligand. Changes in translational and rotational solute entropies, estimated from the ideal gas one-molecule partition functions, are denoted $TS_{trans,ig}$ and $TS_{rot,ig}$ respectively. The

entropy contribution from rotational rigid body degrees of freedom depends on the conformation of a given solute molecule and is averaged over the stored snapshots of the trajectory. The translational rigid body entropy term only depends on the mass of the molecule in the standard state and can be calculated from a single snapshot. Absolute vibrational entropies, $TS_{vib,nm}$, are calculated from normal mode analysis using the Nmode program of the AMBER8 suite. Starting coordinates for the normal mode analysis are taken from a series of uncorrelated snapshots collected from the trajectory. Prior to the calculation of the vibrational frequencies, each solute structure is minimized by a conjugate gradient method until the root-mean-square of the Cartesian elements in the gradient is less than the default value of 1.0^{-4} kcal mol $^{-1}$ Å $^{-1}$. Vibrational normal mode frequencies were calculated at a temperature of 300K using all atoms in the solute molecules in the normal mode analysis. The translational contribution to the rigid body free energy term of Swanson *et al.* [86], ΔG_{rbt} , was calculated from ligand center of mass motions in the complex trajectory. The covariance matrix of ligand center of mass motions was evaluated, after overall translational and rotational motions of the complex were removed, by superposing FimH backbone atoms onto a reference frame (typically the first frame of the production run). The covariance matrix of Euler angle fluctuations of the ligand in the complex was calculated from the rotation matrix needed to superpose all, or a subset of ligand atoms, onto a reference structure, after global rotational and translational motions of the complex were removed. The covariance matrices were, in both cases, evaluated for all snapshots of the stored trajectory. The absolute configurational entropy of a given solute molecule, using the method of Schlitter [90], TS_{config} , was in all systems calculated from all atom covariance matrices using all snapshots of the trajectory. Translational and rotational motions of the solute molecules are also in these calculations removed by superposing respective molecule onto a given reference frame. The entropy of all the abovementioned methods was evaluated at the simulation temperature of 300 K.

5.4 Structure of the FimH CRD

The key for understanding the structural basis for lectin-carbohydrate recognition and corresponding binding affinities lies in the structure of the receptor binding pocket. In the case of FimH, the binding pocket is perfectly designed to fit a D-mannose monosac-

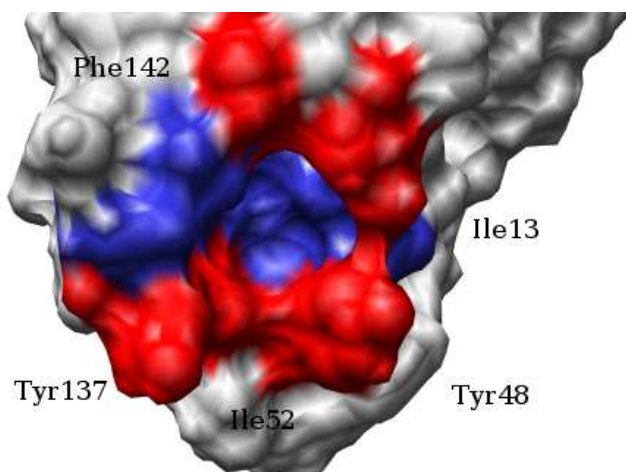


Figure 5.3: Surface representation of the FimH binding pocket with residues in the extended binding region containing non-polar patches at the molecular surface highlighted in red. FimH residues with side chains of predominately polar character are shown in blue. The structure is taken from the FimH-FimC chaperone complex with D-mannose from the Protein Data Bank (1KLF).

carbohydrate through extensive hydrogen bonding and non-polar interactions with a subset of important residues in the binding pocket. Hung *et al.* [208] identified residues Phe1, Asn46, Asp47, Asp54, Gln133, Asn135, Asp140 and Phe142, located in the amino-terminal domain (residues 1-158) of FimH, as essential residues for binding mannose. FimH residues Phe1, Asn46, Asp47, Asp54, Gln133, Asn135 and Asp140 form hydrogen bonds with the hydroxyl oxygens and the endocyclic oxygen of the mannose. Phe142 further contributes to the binding affinity, stabilizing the complex with hydrogen bonds and non-polar interactions. This study showed that all of the residues involved in D-mannose binding is an invariant feature in 200 FimH structures in *E. coli* from different pathogens. Using site-directed mutagenesis, this study also showed that mutations of several of these residues decrease, or completely abolish, D-mannose binding affinities. This subset of FimH residues form a deep binding pocket in the the FimH amino-terminal domain, which fully envelopes the mannose (see Figures 5.1 and 5.3), resulting in a relative high lectin-carbohydrate binding affinity. FimH binds mannose in the micromolar range, compared the millimolar range often observed for other lectin-monosaccharide complexes [209].

Binding affinities in the nanomolar range have been measured for larger oligosaccharides. Sharon [215] showed that the trisaccharide $\text{Man}\alpha 1\text{-3Man}\beta 1\text{-4GlcNAc}$ shows a 21

fold higher inhibitory efficiency compared to Man α Me of agglutination of yeast and red blood cells. Larger mannose-terminal saccharides, found in the cell surface glycocalyx, were in this study found to have even larger inhibitory efficiency, indicating an extended binding region beyond the binding pocket. The stronger interactions between FimH and oligosaccharides have been attributed to stacking of non-polar surface patches of carbohydrate rings outside the binding pocket, onto non-polar patches of FimH [208]. Increased affinity, resulting from hydrophobic interactions, has been verified in several studies using synthetic mannosides with hydrophobic aglycons [209, 216, 217]. Figure 5.3 shows a surface representation of the FimH binding pocket. Hydrophobic residues Ile13, Tyr48, Ile52 and Phe142, surrounding the FimH binding pocket, are highlighted in the figure. These residues form a hydrophobic ridge around the binding pocket, allowing for favorable interactions with hydrophobic patches of additional carbohydrate residues or synthetical hydrophobic substituents.

5.5 FimH-Man α Me complex

Man α Me is the simplest carbohydrate ligand investigated in this study. From experiment, this is also the weakest anti-adhesive ligand, with binding affinities in the micromolar range (see above). Structural, dynamic and thermodynamic properties of the the FimH-Man α Me complex are presented in this section, as a reference for the larger ligands investigated in later sections of this chapter. The results presented in this section are averaged over 25000 snapshots from a 25 ns molecular dynamics trajectory.

5.5.1 Hydrogen bond analysis

Intermolecular hydrogen bonds between lectins and carbohydrates in the CRD are important contributors to the binding affinity for lectin carbohydrate complexes. Just considering the enthalpy changes of the binding process, it is clear that lectin-carbohydrate hydrogen bonding must compensate the energetic cost of removing water molecules hydrogen bonded to, both the lectin and the carbohydrate, for the complex to be thermodynamically stable. In this section, we present results from hydrogen bond analysis of the FimH-Man α Me complex. The geometric criteria used for identifying a hydrogen

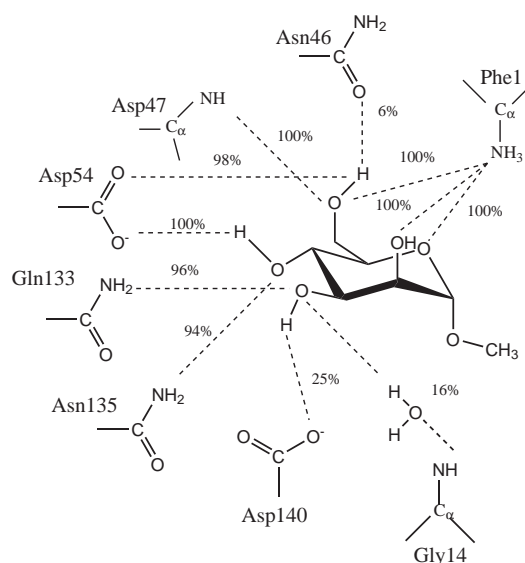


Figure 5.4: Important hydrogen bonds formed between $\text{Man}\alpha\text{Me}$ and FimH residues in the CRD. Calculated hydrogen bond occupancies are given as percentages in the figure.

bond between a given hydrogen bond donor and acceptor are that the distance between the two atoms is $\leq 3.6 \text{ \AA}$ and that the hydrogen bond angle (see Section 2.2.2) is $\leq 60^\circ$. These simulations showed that the intermolecular hydrogen bond network of the complex is not well described by a static picture. Reorientations of both hydroxyl groups in the mannoside and side chains in the binding pocket result in many different hydrogen bond topologies. All possible hydrogen bonds formed between the mannoside and FimH are difficult to represent graphically. Therefore, only hydrogen bonds with high probabilities to occur are shown in Figure 5.4. All oxygen atoms in the mannoside, with the exception of oxygen O1, are engaged in hydrogen bonding with polar atoms in side chains or the backbone of FimH. The N-terminal FimH residue Phe1 forms simultaneous hydrogen bonds with mannose oxygens O2, O5 and O6, which all have occupancies of approximately 100% in the whole 25 ns trajectory. Other important hydrogen partners in FimH are Asp47, Asp54, Gln133 and Asn135, which are all present around 100% of the stored snapshots of the trajectory. Lower hydrogen bond occupancies can be seen for interactions with residues Asp140 and Asn46. The hydrogen bond network shown in Figure 5.4 is in good agreement with those derived from crystal structures [10, 208].

The strength of individual hydrogen bonds can further be analyzed by the geometric arrangement of the donor and acceptor atoms. Average hydrogen bond lengths and

hydrogen bond angles of a selected subset of Fimh-Man α Me intramolecular hydrogen bonds are presented in Table 5.1. The interaction energy of a given hydrogen bond becomes more negative with decreasing hydrogen bond lengths and angles. Average hydrogen bond distances and angles of water-water hydrogen bonds, calculated with the TIP3P water model, using the same hydrogen bond criteria, of 3.0 Å and 27.5° can be used as a direct reference. Hydrogen bonds formed between Asp54 and hydroxyl oxygens O6 and O4 of the mannoside stand out as strong hydrogen bonds with much shorter hydrogen bond distances and more linear hydrogen bond angles compared to water-water hydrogen bonds.

Table 5.1: Average geometries of selected hydrogen bonds formed in the FimH-Man α Me complex.

Acceptor-DonorH,Donor ^a	r (Å) ^b	θ (degrees) ^c	Occ. (%) ^d
Man02-Phe1:1H1,N	2.867	21.95	26
Man02-Phe1:1H2,N	2.856	21.12	44
Man02-Phe1:1H3,N	2.854	21.45	30
Man03-Gln133:HE21,NE2	3.044	13.12	96
Man04-Asn135:HD22,ND2	2.951	16.90	94
Man05-Phe1:H1,N	3.018	46.88	44
Man05-Phe1:H2,N	3.014	47.05	31
Man05-Phe1:H3,N	3.030	46.72	26
Man06-Asp47:H,N	3.024	12.48	100
Man06-Phe1:H1,N	2.860	15.99	44
Man06-Phe1:H2,N	2.865	16.26	29
Man06-Phe1:H3,N	2.866	16.32	25
Asp140OD1-Man:H30,O3	2.713	12.21	25
Asp54OD1-Man:H40,O4	2.583	8.61	100
Asp54OD2-Man:H60,O6	2.596	8.12	98
Asp46OD1-Man:H60,O6	3.178	54.14	5

^a Atom names from the GLYCAM04 and PARM99 force fields.

^b Average hydrogen bond distance (see Chapter 2.2.2).

^c Average hydrogen bond angle (see Chapter 2.2.2).

^d Occurrence probability of the hydrogen bond.

5.5.2 Binding free energy analysis

The binding free energy change associated with complex formation of FimH and Man α Me is, in this section, estimated with the MM-PB(GB)SA approach. Free energy analysis was performed using both a single and a three trajectory approach. Detailed analysis showed that, even though the trajectories are much longer compared to previous similar studies [31, 213], the 25 ns trajectory clearly is too short for completely sampling the conformational space of FimH. Using a three-trajectory approach for calculating free energies of binding resulted in large statistical uncertainties. Relatively large contributions to the calculated free energy of binding could also be seen for residues far away from the CRD, as a consequence of sampling different parts of the configurational space of FimH in the complex and receptor simulations. The MM-PB(GB)SA results, presented in this and the following sections, are therefore calculated using structures of the ligand and receptor from the trajectory of the complex.

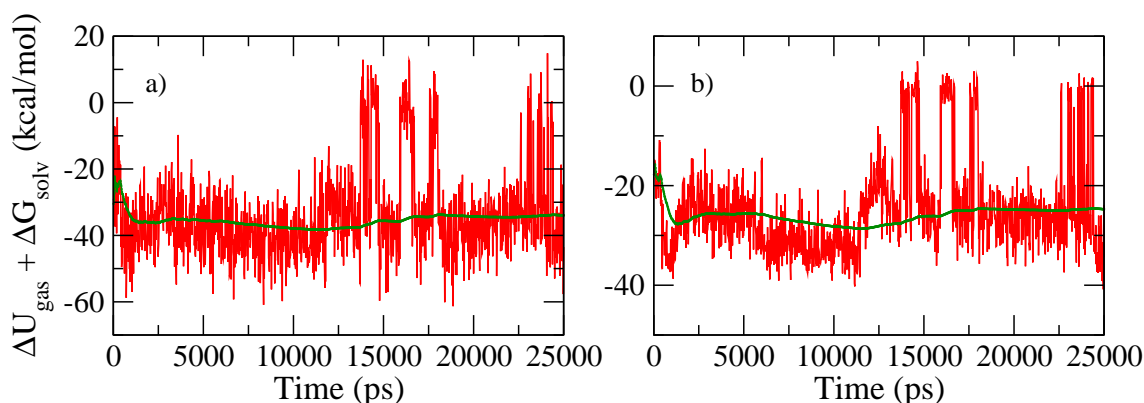


Figure 5.5: Calculated instantaneous $\Delta U_{vac} + \Delta G_{solv}$ values and the running average thereof, for the MM-GBSA method (a) and the MM-PBSA method (b)

Both the Poisson-Boltzmann and Generalized Born methods are used to estimate solvation free energies of the solute molecules. Figure 5.5 shows instantaneous values, and the running average, of changes in vacuum binding potential energy plus the solvation free energy change of the complex, relative to the free FimH and Man α Me molecules, estimated with the GBSA and PBSA methods. Calculations of center of mass positional fluctuations of the mannoside in the complex showed that only small displacements from the initial crystallographic structure occur during the 25 ns trajectory. The MM-PB(GB)SA results in Figure 5.5 show that small conformational

changes of the complex can result in large fluctuations in the free energy estimates. The running averages are, however, fairly stable and do not drift significantly with time. A rather large difference can be seen between the two implicit solvation models. Although the time-evolution of the binding free energy estimates are similar for the two methods, a shift towards more negative $\Delta U_{vac} + \Delta G_{solv}$ values can be seen for the GBSA results. From these calculations it is difficult to determine which method gives a better estimate of the solvation free energy of the solute molecules. Although the GB method is an approximation of the numerical solution of the Poisson-Boltzmann equation, it has also been shown that PB-method needs reparameterized atomic radii and charges to reproduce experimental solvation free energies of small organic molecules [79].

The results from the MM-PB(GB)SA free energy analysis of the FimH-Man α Me complex are shown in Table 5.2. These results are averaged over 1250 snapshots, separated by 20 ps, from the complex trajectory. The free energy change, calculated as the sum of changes in vacuum potential energy, plus the solvation free energy contributions, adds up to -37.5 ± 0.4 kcal/mol using the GBSA implicit solvation model. With the PBSA method, the free energy change is much smaller and adds up to -24.8 ± 0.3 kcal/mol. Table 5.2 shows that the main contribution to the negative free energy change upon complexation comes from favorable Coulomb interactions between FimH and the mannoside. Favorable van der Waals interactions give a contribution of approximately -14 kcal/mol to the vacuum interaction energy. The total solvation free energy contribution in the thermodynamic cycle ($\Delta G_{PB/GB,solv}$), adds an unfavorable free energy contribution to binding. The ΔU_{vac} contribution is, however, dominating, showing that binding of Man α Me to FimH is enthalpy driven, mainly due to strong hydrogen bonding between the two solutes in the bound state. The free energy change, given by the MM-PB(GB)SA analysis, is not complete, since entropy changes of the solute molecules are not taken into account. Limitations in the available space for rigid body motions and internal motions of both the mannoside and FimH can add an additional entropy penalty to the binding free energy change. Entropy changes are here calculated separately from contributions from translational, rotational and configurational degrees of freedom, and finally added up to give the total entropy change from all 3N degrees of freedom of the solute molecules. The first contribution to the solute entropy change investigated, is from the available rotational and translational degrees of freedom of FimH and Man α Me in the complex, relative to the free state. The standard method for estimating rigid body entropy changes, within the MM-PB(GB)SA

Table 5.2: MM-PB(GB)SA results from the FimH-Man α Me complex.

	Complex	σ_M	Receptor	σ_M	Ligand	σ_M	Delta	σ_M
U_{ele} ^a	-4647.9	1.5	-4665.6	1.2	90.0	0.1	-72.2	0.7
U_{vdw} ^b	-598.1	0.5	-587.6	0.5	3.4	0.0	-13.9	0.2
U_{int} ^c	3140.8	1.0	3117.2	1.0	23.6	0.1	0.0	0.0
U_{vac} ^d	-2105.2	1.6	-2136.0	1.0	116.9	0.6	-86.1	0.8
$G_{GB,ele}$ ^e	1329.6	1.2	-1351.4	1.1	-28.7	0.0	50.5	0.5
$G_{PB,ele}$ ^f	-1275.2	1.6	-1297.2	1.1	-42.3	0.1	64.3	0.6
$G_{GB,np}$ ^g	37.8	0.0	38.4	0.0	1.28	0.0	-2.0	0.0
$G_{PB,np}$ ^h	43.8	0.0	44.0	0.0	3.0	0.0	-3.1	0.0
$G_{GB,solv}$ ⁱ	-1291.8	1.2	-1313.0	1.1	-27.4	0.0	48.5	0.5
$G_{PB,solv}$ ^j	-1231.3	1.3	-1253.2	1.1	-39.33	0.0	61.23	0.6
$U_{vac} + G_{GB,sol}$ ^k	-3397.0	1.0	-3449.0	0.9	89.5	0.1	-37.5	0.4
$U_{vac} + G_{PB,sol}$ ^l	-3336.5	1.1	-3389.2	1.0	77.6	0.1	-24.9	0.3

^a Electrostatic molecular mechanics energy in vacuum.

^b Van der Waals molecular mechanics energy in vacuum.

^c Internal energy from bond, angle and torsional terms in the force field.

^d Total vacuum potential energy (sum of electrostatic, Van der Waals and internal energies).

^e Electrostatic contribution to the solvation free energy from the Poisson-Boltzmann approach.

^f Electrostatic contribution to the solvation free energy from the Generalized Born approach.

^g Non-polar contribution to the solvation free energy calculated with the SA (GB) approach.

^h Non-polar contribution to the solvation free energy calculated with the SA (PB) approach.

ⁱ Sum of non-polar and (GB) polar solvation free energies.

^j Sum of non-polar and (PB) polar solvation free energies.

^k Sum of gas phase energy and the solvation (GB) free energy contribution.

^l Sum of gas phase energy and the solvation (PB) free energy contribution.

approach, is to use ideal gas one-molecule partition functions for translations and rotations separately, and calculate absolute entropies from statistical thermodynamic formulas (see Section 2.2.7). The calculated absolute ideal gas translational and rotational entropies, $TS_{trans,ig}^0$ and $TS_{rot,ig}$, are presented in Table 5.3. The translational entropy depends on the concentration and is here calculated at a standard state of 1 M (1 solute molecule/1660 Å). The total rigid body entropy penalty for complex formation of FimH and Man α Me is, using the ideal gas treatment, approximately 22 kcal/mol and indicates a large entropy barrier for complex formation. As seen in Table 5.3, this entropy penalty corresponds to the total absolute rigid body entropy of

Table 5.3: Calculated entropies of the FimH-Man α Me complex.

	Complex	Receptor	Ligand	Delta
$TS_{trans,ig}^{\circ}$ ^a	16.5 \pm 0.0	16.5 \pm 0.0	12.5 \pm 0.0	-12.5 \pm 0.0
$TS_{rot,ig}$ ^b	16.8 \pm 0.0	16.7 \pm 0.0	9.40 \pm 0.0	-9.3 \pm 0.0
G_{rbt}° ^c				10.6
$TS_{vib,nm}$ ^d	1689.4 \pm 0.7	1671.6 \pm 0.6	13.7 \pm 0.0	4.1 \pm 0.8
TS_{config} ^e	2670.0	2721.80	27.6	-79.4
$TS_{config,rec}$ ^f	23.0		23.5	-0.5
$TS_{config,lig}$ ^g	2645.1	2721.8		-76.7

^a Standard translational entropy at 1 M calculated from ideal gas one molecule partition functions.

^b Translational entropy calculated from ideal gas one molecule partition functions.

^c Free energy contribution from the loss in rigid body degrees of freedom of the ligand calculated with the method of Swanson *et al.*

^d Vibrational entropy calculated from Normal Mode Analysis.

^e Configurational entropy calculated using Schlitter's method.

^f Configurational entropy of just the receptor atoms calculated using Schlitter's method.

^g Configurational entropy of just the ligand atoms calculated using Schlitter's method.

the mannoside in the free state. In reality, all rotational and translational degrees of freedom of the ligand in the complex are not completely lost, but rather transformed into internal motions of the complex.

The method of Swanson *et al.* (see Chapter 2.2.7) provides an alternative to the ideal gas treatment for estimating the entropy loss of rigid body motions of the mannoside in the complex. Within this method, entropy changes are calculated, taking translational and rotational degrees of freedom available to the mannoside in the complex into account. The superposition of the FimH backbone atoms, to remove overall rotational and translational motions of the complex, was done with root mean square deviations within 1.5 Å with respect to the first snapshot of the trajectory. Fluctuations around average distances and angles of the six translational and rotational modes of the mannoside in the complex could be seen to take the form of Gaussian distributions around a single minimum in all cases, showing that the quasi-harmonic treatment is a good approximation in this case. Maximum deviations from the average value are smaller than 1 Å and 10° for the positional and angular fluctuations respectively. The free energy contribution from restrictions of ligand rigid body motions in the complex, estimated

by the method of Swanson *et al.*, ΔG_{rbt} in Table 5.3, is 10.6 kcal/mol, less than half the value of the corresponding entropy loss estimated with the ideal gas approach.

Complex formation may further introduce entropy penalties, due to restrictions of the internal degrees of freedom of the receptor and the ligand in the complex. Within the harmonic approximation, these internal motions are described by $3N-6$ normal modes of vibrations of each solute molecule. Normal mode analysis (NMA) of the complex, and both FimH and the mannoside in the unbound state, was performed on a series of snapshots from the FimH-Man α Me trajectory. Due to the computational cost of the method, the normal mode analysis was limited to a total of 260 uncorrelated snapshots of the trajectory. Using the analytical formula of a quantum harmonic oscillator, absolute entropies of the complex, receptor and ligand were calculated by summing over contributions from all vibrational modes and averaged over all 260 structures. The results, presented in Table 5.3, show that the vibrational entropy of the complex is larger than that of the sum of vibrational entropies of the receptor and the ligand in the free state, giving a positive entropy contribution for the complex. It is important to note that this analysis includes the six additional internal degrees of freedom of the complex, i.e. rigid body motions of the ligand transformed into internal degrees of freedom of the complex. Thus, the vibrational entropy can be added directly to the ideal gas rigid body entropy to give an estimate of the solute entropy change, but is not compatible with the rigid body free energy contribution calculated with the method of Swanson *et al.*, since contributions from the six additional degrees of freedom of the complex would be double counted. The entropy analysis indicates that only small changes in internal degrees of freedom of FimH and the mannoside occur on complex formation. In the NMA analysis, entropy penalties due to limitations in the number of conformational isomers available for, for instance, protein side chains in the binding pocket are not taken into account. For the FimH-Man α Me complex, where the mannoside is buried in a deep binding pocket, further entropy penalties from limitations of both FimH side chain and carbohydrate conformational isomers are expected.

The method of Schlitter provides an upper limit of the total configurational (vibrational and conformational) entropy of a given macromolecule. This method requires three separate trajectories of the complex, receptor and ligand. Calculated absolute configurational entropies, TS_{config} , of the the solute molecules in the three 25 ns trajectories are also presented in Table 5.3. The calculated configurational entropy difference between the complex and the free receptor and ligand is with this method approximately -79 kcal/mol. This is clearly is an overestimation. Combined with the

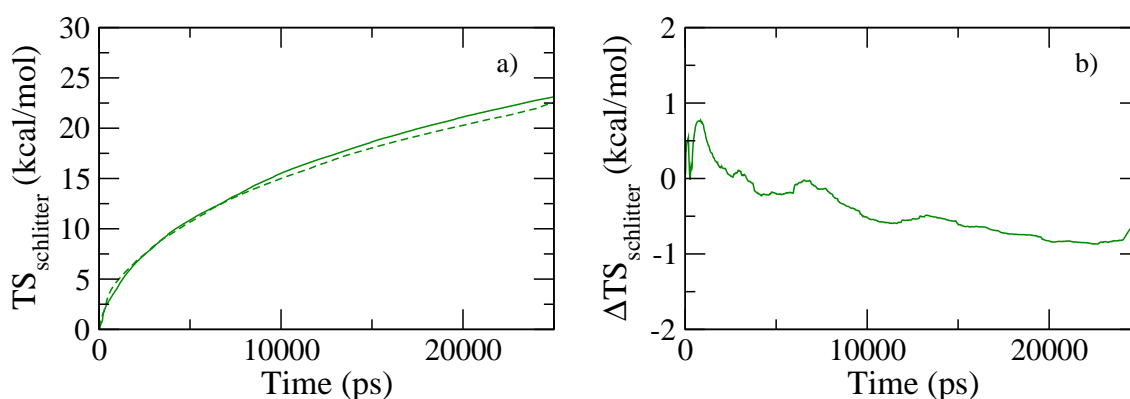


Figure 5.6: a) Calculated TS_{config} values of the Man α Me ligand as a function of the simulation time. The absolute configurational entropy of the ligand calculated from the complex trajectory is shown as dashed lines and the corresponding entropy of the ligand in the free state is shown with solid lines. b) Time evolution of the calculated configurational entropy change of Man α Me in the complex relative to the free state.

MM-PBSA or MM-GBSA results, this estimate of the entropy change gives a positive free energy change of binding. Although Schlitter's method has successfully been verified for small systems where the quasi-harmonic approximation is valid [91], its application to biomolecules has not been so successful. There are several reasons why Schlitter's method overestimates the configurational entropy difference of larger molecular complexes. The first issue is insufficient sampling of the configurational space, due to too short simulation times. In the light of the observation that proteins display conformational motions occurring on a time scale of a few milliseconds [218], the configurational space of FimH is likely to be poorly sampled in these simulations. Another potential problem can be related to anharmonicity of the correlated motions of the solute molecules, which may overestimate the calculated absolute entropies and entropy differences [219]. Additional problems have been related to spurious correlations in the covariance matrix, as a result of calculating it from finite simulation data [220]. Table 5.3 also shows calculated configurational entropy changes of FimH and the mannoside in the complex relative to the free state separately. For the case of the mannoside, the six ring atoms of the initial structure were used as a reference for the superposition used to remove rigid body motions. Figure 5.6a shows calculated absolute configurational entropies of the mannoside, in the complex and in the free state, as a function of the simulation time. Even after 25 ns, both curves are increasing as the mannoside explores new parts of the configurational space. Figure 5.6b shows the difference between the

calculated absolute configurational entropies in the bound and free states. The entropy difference has not converged either, and shows fluctuations around 0.5 kcal/mol during the last 5 ns of the trajectories. After 25 ns the configurational entropy loss of the mannoside in the complex is -0.5 kcal/mol. Analysis of atomic fluctuations show that this is mainly due to restrictions of the 6 arm of the mannose, but also of the hydroxyl hydrogens (H2O,H3O and H4O in the GLYCAM04 force field) of the mannose in the complex. From $\Delta TS_{config,rec}$ in Table 5.3, it is clear that the major contribution to the calculated large configurational entropy difference comes from FimH atomic fluctuations. Since the goal of this study is to compare $\Delta\Delta G_{bind}^0$ values for different ligands to a common receptor, one can make the approximation that absolute configurational entropies of FimH in the different complexes is the same, and just compare differences in the configurational space of the ligand. Together with the rigid body free energy term, ΔG_{rbt}^0 , the configurational entropy of the ligand can be used as an alternative to the ideal gas + vibrational entropy combination to investigate relative binding affinities of the different ligands. The combination of $\Delta U_{vac} + \Delta G_{PB,solv} - (T\Delta S_{trans,ig}^0 + T\Delta S_{rot,ig} + T\Delta S_{vib,nm})$ gives a ΔG_{bind}^0 value of -7.1 ± 0.8 kcal/mol, which is very close to the experimentally determined solution affinity of mannose for FimH with $K_d = 2.3 \mu\text{M}$ ($= -7.6$ kcal/mol). The good agreement between the calculated and experimental results is encouraging, but based on the approximations on which these calculations are done this might just be a coincidence. Calculations of ΔG_{bind}^0 of the other investigated ligands and calculations of $\Delta\Delta G_{bind}^0$ values of the different ligands will further elucidate the accuracy and limitations of the methods. The MM-GBSA approach is strongly overestimating the binding affinity with a ΔG_{bind}^0 of -19.8 kcal/mol. Similar results were found in the study of study of Gohlke and Case [213] which indicate that the Generalized Born approach underestimates the ΔG_{solv} contribution to the free energy change. The combination of $\Delta U_{vac} + \Delta G_{PB,solv} + \Delta G_{rbt}^0 - T\Delta S_{con,fig,lig}$ contributions gives a ΔG_{bind}^0 value of -13.8 kcal/mol, which overestimates the binding affinity, due to neglect of changes in configurational entropy of FimH in the complex.

5.5.3 The role of structural water molecules

In order to predict receptor-ligand affinities it is necessary to, not only, consider free energy contributions from desolvation of the solute molecular surfaces, but also the energetic contributions from structural water molecules at the binding interface. These

water molecules are immobilized at the binding interface for long time intervals and can be considered to be part of the complex by bridging polar receptor and ligand atoms through hydrogen bonding. Due to the small size, water molecules can enter cavities in the receptor-ligand binding interface and contribute favorably to binding affinity by both forming additional hydrogen bonds, as well as, shielding unfavorable receptor-ligand Coulomb interactions [221]. If the (favorable) enthalpy change of the system, due to the presence of a structural water molecule, is larger than the corresponding entropy loss, structural water molecules can make a significant contribution on the overall binding affinity. Ladbury [222] estimated the upper limit of the free energy contribution for interfacial water molecules to be -1.67 kcal/mol. Given that experimental free energy changes measured for lectin-carbohydrate complex formation, typically are in the range of 5-11 kcal/mol, it is clear that the presence of structural water molecules at the binding interface have to be considered when trying to understand lectin-carbohydrate binding affinities.

A detailed analysis of water-mediated hydrogen bonding between FimH and the mannoside, revealed one, or occasionally two water molecules, present in the binding pocket. A single water molecule can also be involved in cooperative hydrogen bonding between different FimH residues and the mannoside, leading to many different bridging possibilities. The most frequently observed water mediated hydrogen bond is indicated in Figure 5.4, bridging Gly14 and oxygen O3 in the mannoside. Several other indirect FimH-Man α Me interactions are also observed. Time series of frequently occurring water-mediated hydrogen bonds are shown in Figure 5.7. Although individual water molecules can reside in the binding pocket for very long times, exchanges of bridging water molecules are frequent. Taking only water molecules with residence times over 100 ps in the FimH binding pocket, 35 different water molecules are mediating interactions between FimH and the mannose during the 24 ns trajectory.

The presence of structural water molecules should, in principle, be captured by the implicit solvation treatment of solvation free energy changes since the structures are taken from an explicit water simulation. As a test, we included the water molecule bridging Man160-O3 and Gly14, present in the time interval of 15 to 20 ns of the complex trajectory, in the MM-GBSA analysis. In these calculations, the water molecule was considered a part of the ligand, both in bound and free states. Solvation effects beyond this water molecule were treated by the GBSA methodology. The $\Delta\Delta H_{gas} + \Delta\Delta G_{solv,GB}$ value for the 5 ns block for the structures with and without the structural water molecule is -3.4 kcal/mol. Taking the value of 2.1 kcal/mol as the upper

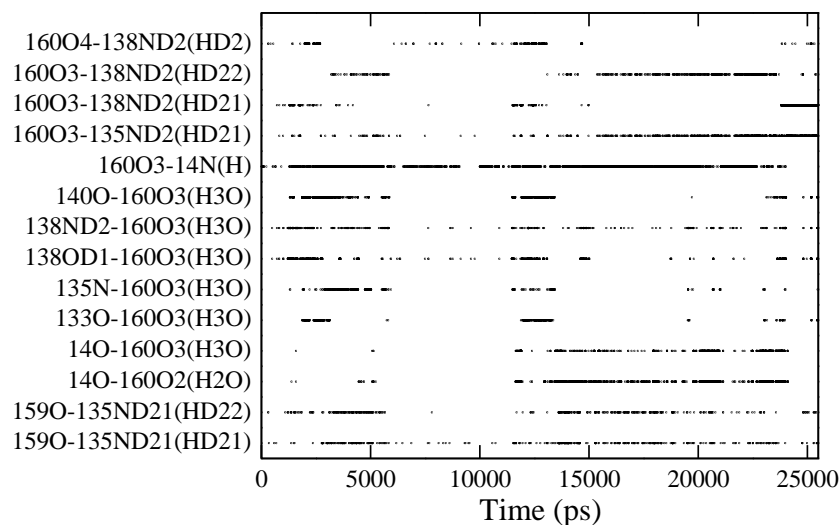


Figure 5.7: Time series of a selected number of water mediated FimH-Man α Me hydrogen bond interactions. Atom and residue names are those of the PARM99/GLYCAM04 force fields.

value for the entropy loss of structural water molecules from Dunitz [146], the free energy difference is still lowered with 1.3 kcal/mol by including the structural water molecule. A recent study of Yu *et al.* [223] showed that the inclusion of a structural water molecule was needed to reproduce explicit water calculated solvation free energies for the acetate-guanidinium ion pair, using both generalized Born and Poisson-Boltzmann implicit solvation models. Thus, structural water molecules can make important contributions to the binding affinity, which may not be well treated by implicit solvation models.

5.5.4 Per-residue free energy decomposition

In this section, we investigate contributions of the MM-GBSA results on a residue basis. Using the GB implicit solvation model, the calculated free energy change of binding can be decomposed into contributions from a given subset of atoms. Figure 5.8 shows contributions to the binding free energy change from FimH residues within a distance of 6 Å from any atom in the mannoside in the initial crystal structure. The calculated values are averaged over 1250 snapshots, separated by 20 ps, of the complex trajectory. Standard errors of the mean are displayed as error bars in the figures.

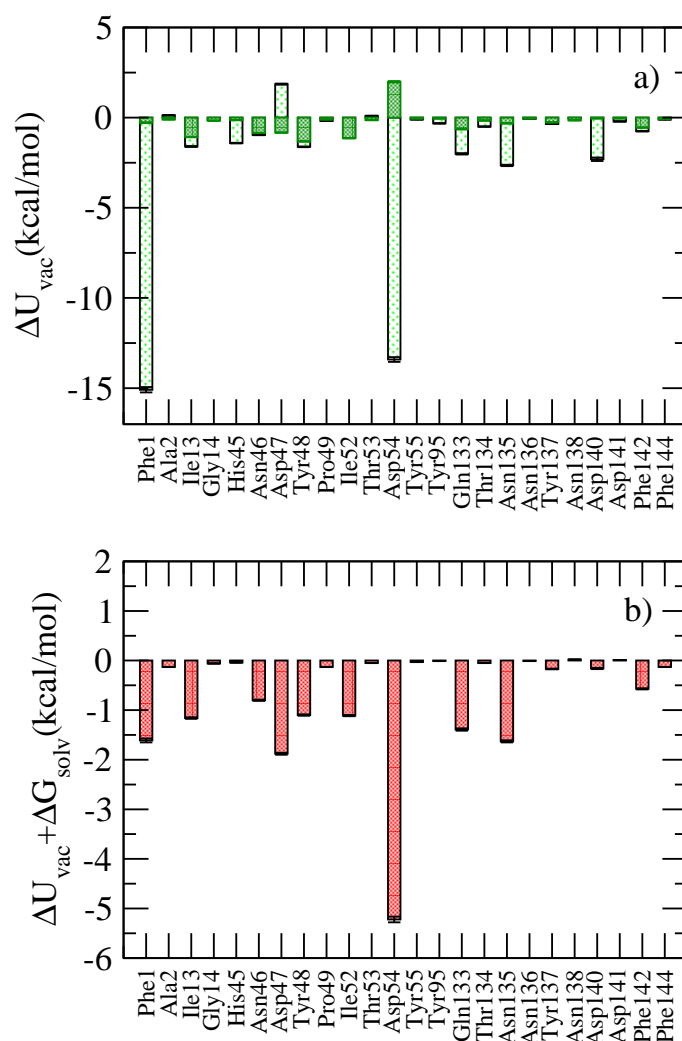


Figure 5.8: a) Decomposition of ΔU_{vac} values of the FimH-Man α Me complex on a FimH residue basis. Van der Waals contribution to the potential energy is shown in a) as shaded bars. b) Decomposed $\Delta U_{vac} + \Delta G_{solv}$ values calculated with the MM-GBSA approach.

In Figure 5.8a, the difference in vacuum molecular mechanics energy for FimH residues in the complex and in the free state is displayed. The total ΔU contribution from each residue is shown, together with just the contribution from van der Waals interactions (shaded bars). Large negative ΔU_{ele} values can be seen for residues Phe1 and Asp54. Smaller contributions can be seen from the other residues involved in hydrogen bonding with the mannoside. Although the FimH residue Asp47 forms hydrogen bonds with mannoside (see Figure 5.4), it still gives a positive ΔU_{ele} contribution. This can be caused by conformational changes of the binding pocket in the complex, affecting the position the side chain of Asp47, and also possibly disrupting internal hydrogen bonds

with other residues in FimH. A negative ΔU_{ele} contribution also comes from His45, although not directly involved in hydrogen bonding with the mannoside. FimH residues Ile13, Tyr48, Ile52 and Phe142, containing non-polar and aromatic side chains, all show additional favorable ΔU_{vdW} contributions in the complex. Visual inspection of the complex geometry showed that contributions from Tyr48 and Ile52 is mainly due to interactions with the methyl group of the mannoside, whereas Tyr48 and Phe142 interact with non-polar patches of the mannose residue.

Figure 5.8b shows the free energy contribution to the FimH-Man α Me complex for the corresponding FimH residues, implicitly taking solvent reorganization effects into account using the GBSA method. The large negative vacuum ΔU_{ele} values for Phe1 and Asp54 are compensated by the $\Delta G_{GB,solv}$ contribution. Together with the solvation free energy change of the complex, all terms in the thermodynamic cycle, with the exception of solute entropy changes, are included. Enthalpy and entropy changes of water of the binding process are implicitly taken into account in the $\Delta G_{GB,solv}$ term. Including solvent reorganization effects, FimH residues Phe1, Asp47, Asp54, Gln133, and Asn135 stand out as important contributors to the overall binding affinity of the mannose. Also Asp47 makes a favorable contribution to the binding free energy change when including the $\Delta G_{GB,solv}$ contribution. This observation can be explained by the position of the residue in the binding pocket. The part of Asp47 that is in the binding pocket of FimH is mainly non-polar in nature, with the charged side chain pointing out towards the solvent. The site-directed mutagenesis study of Hung *et al.* [208] showed that mutation of residues Asp54 resulted in a complete loss in binding affinity of D-mannose. From these calculations, this is also the FimH residue making the largest contribution to the binding affinity. This study shows that additional contributions to the binding affinity of Man α Me to FimH come from residues Ile13, Asp47, Tyr48, Ile52 and Phe142. In order to obtain the total binding free energy contribution per residue, also residual entropy changes need to be incorporated. Still, the results presented here can be used as an indication of the residues in FimH responsible for recognition and binding of the the Man α Me ligand.

5.6 Fimh-Man α PNP complex

Synthetic mannosides with aliphatic or aromatic aglycons have been found to increase the affinity of FimH binding, from dissociation constants in the micromolar range to the nanomolar range [209, 216, 217]. The enhanced affinity for aliphatic and aromatic mannosides has been attributed to interactions with the aromatic residues in the tyrosine gate of FimH (see Figure 5.3). In this section, we investigate the structural basis and thermodynamics of FimH interactions with the ligand Man α -p-Nitrophenyl (ligand II in Figure 5.2), which from hereon is abbreviated as Man α PNP.

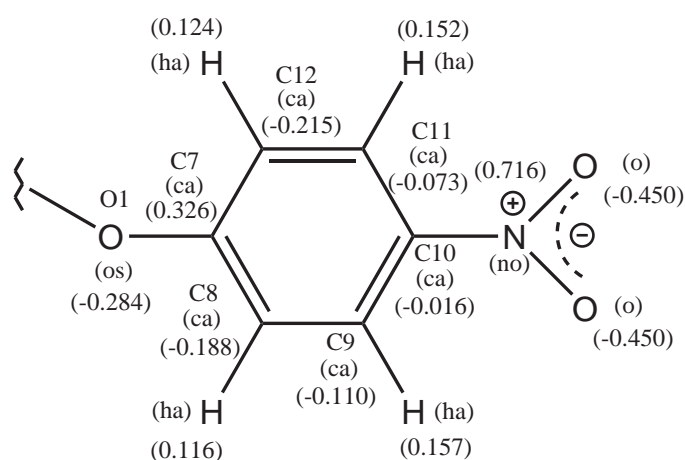


Figure 5.9: Force field parameters of the p-Nitrophenyl group. GAFF atom types, partial charges and atoms names are shown in the figure.

5.6.1 Force field development

Force field parameters for the nitrophenyl group are missing in the AMBER and GLYCAM04 force fields, and are derived specially for this study. RESP charges for the nitrophenyl group were calculated at the HF/6-31G* level of theory, in a similar way used to calculate the charges in the GLYCAM04 force field. In the charge calculations, the linking oxygen in Figure 5.9 was capped with a methyl group. In order to keep the charges on the mannose, given by the GLYCAM04 force field, RESP charges were derived by applying intramolecular charge restraints using the R.E.D program of Pigache *et al.* [224].

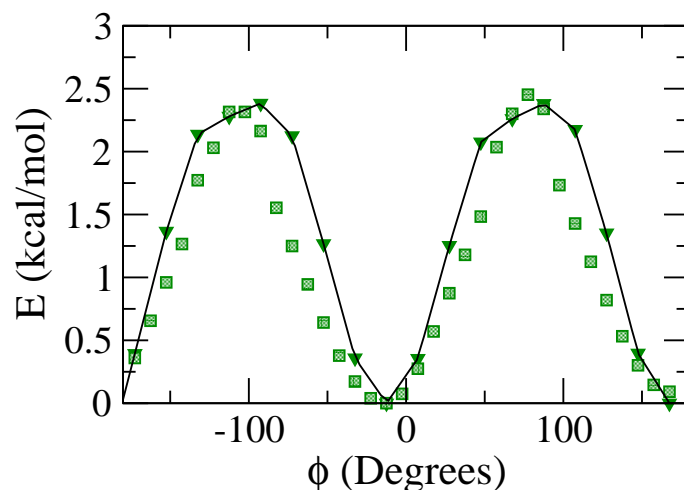


Figure 5.10: Relaxed potential energy scans of the C1-O1-C7-C12 dihedral angle of the Man α PNP ligand. The HF/6-31G* energy profile calculated is shown as full lines. AMBER results using the derived force field parameters are shown as rectangular points. The AMBER energy profile is shifted for a direct comparison.

A charge restraint of +0.1940 e was applied to the methyl group, which is equal to that of α -D-mannose 0MA (mannose without the glycosidic oxygen) in the GLYCAM04 force field. In this manner, the total molecule is neutral and no reparameterization of the mannose-part is necessary. Atomic partial charges of the nitrophenyl group is presented in Figure 5.9. Remaining force field parameters were taken from the general AMBER force field (GAFF) [225], as assigned by the utility program Antechamber [226]. The GAFF force field is fully compatible with AMBER force fields and can therefore be used directly together with the GLYCAM04 force field. The GAFF generated atom types are shown in Figure 5.9. The only modification of the GAFF parameters is reparameterization of the two torsional angles determining the relative orientation of the mannose and the p-nitrophenyl substituent. These two torsional angles, ϕ and ψ , are here defined as O5-C1-O1-C7 and C1-O1-C7-C12 (atom names are given in Figure 5.9). Relaxed scans around these two dihedral angles were performed using the Gaussian03 program at a HF/6-31G* level of theory, and compared to equivalent scans performed with the AMBER molecular dynamics program with the GAFF parameters and the derived partial atomic charges presented above. Rotation around the ϕ angle is sterically hindered due to the atoms in the mannose and has a single minimum around 60°. The parameters of this dihedral angle were adjusted to reproduce the position of the minimum.

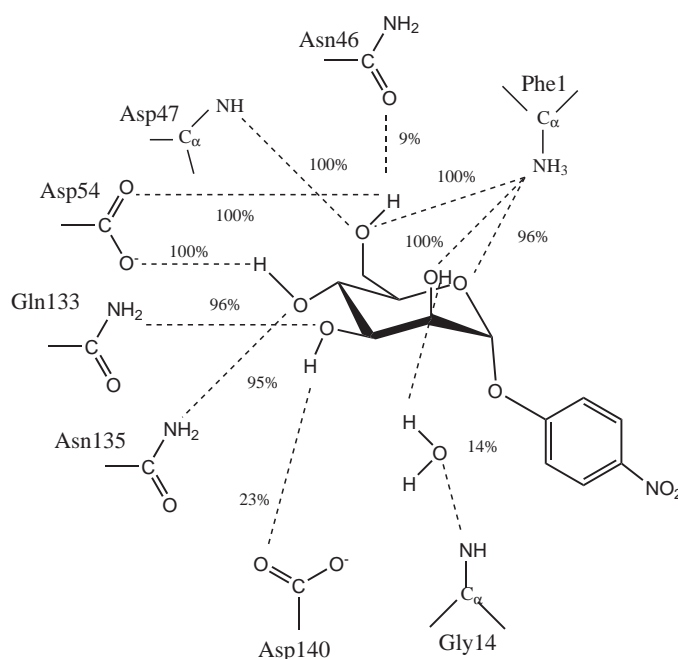


Figure 5.11: Important intermolecular hydrogen bonds formed in the FimH-PNP complex. Hydrogen bond occupancies calculated from the 25 ns trajectory of the complex are given as percentages in the figure.

Parameters of the ψ dihedral angle also had to be adjusted to fit the ab-initio calculations. The results of the ab-initio calculated relaxed potential energy scan around the ψ angle and the corresponding force field results, using the readjusted parameters, are shown in Figure 5.10. Although the curves do not match perfectly, the important energetic features of rotations around this dihedral angle, i. e. the positions of the minima and the height of the barriers, are well reproduced with the force field parameters.

5.6.2 Hydrogen bond analysis

The topology of hydrogen bonds formed between Man α PNP and FimH in the complex, shown in Figure 5.11, is basically the same as that of the FimH-Man α Me complex. The p-nitrophenyl substituent only shows short lived ($< 1\%$ occupancies) hydrogen bonds formed with residues Tyr48, Thr51 and Tyr137 in FimH. The structural water molecule found in the FimH-Man α Me complex is also found here, but bridging Gly14 and the O2 hydroxyl group instead of the O3 hydroxyl group for the majority of the stored

snapshots of the trajectory. Again, several different water molecules can simultaneously mediate hydrogen bonds between FimH and the ligand, and a given structural water molecule can further move around inside the binding pocket, leading to a large number of different water-mediated hydrogen bond combinations.

5.6.3 Binding free energy analysis

Figure 5.12 shows calculated instantaneous $\Delta U_{vac} + \Delta G_{solv}$ values and running averages thereof, calculated with both the MM-PBSA and MM-GBSA approaches. The FimH-Man α PNP complex was stable throughout the simulation and no larger conformational changes were observed. The running average shows a slight drift in the calculated $\Delta U_{vac} + \Delta G_{solv}$ values during the first half of the simulation, but is relatively stable during the last half of the simulation. Again, the results calculated with the two different implicit solvation models are similar, with a negative offset for the the MM-GBSA calculated binding free energy change.

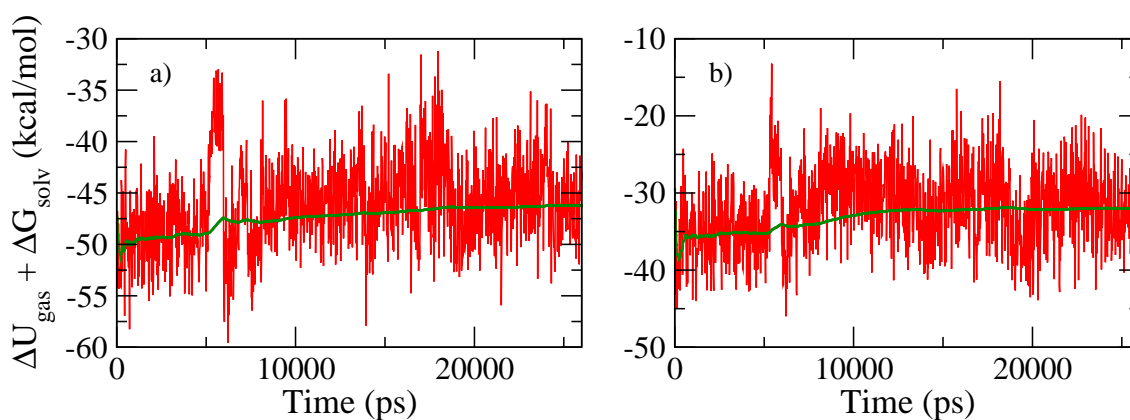


Figure 5.12: Instantaneous $\Delta U_{vac} + \Delta G_{solv}$ values as a function of time for the FimH-Man α PNP ligand a) using the MM-GBSA approach and b) with MM-PBSA approach. Running averages of the calculated $\Delta U_{vac} + \Delta G_{solv}$ values are shown as full lines.

The results of the 1-trajectory MM-PB(GB)SA free energy analysis of the FimH-Man α PNP complex are shown in Table 5.4. Both MM-PBSA and MM-GBSA calculated $\Delta U_{vac} + \Delta G_{solv}$ values, averaged over snapshots separated by 20 ps of the whole 25 ns trajectory, show more favorable contributions to the binding free energy compared to the respective results from the Man α Me complex.

Table 5.4: MM-PB(GB)SA results from the FimH-Man α PNP complex. See Table 5.2 for table captions. All units are in kcal/mol.

	Complex	σ	Receptor	σ	Ligand	σ	Delta	σ
U_{ele} ^a	-4667.3	1.2	-4687.3	1.2	100.7	0.1	-80.7	0.3
U_{vdw} ^b	-603.2	0.5	-589.3	0.5	7.3	0.0	-21.2	0.1
U_{int} ^c	3158.6	0.9	3124.6	0.9	33.9	0.1	0.0	0.0
U_{vac} ^d	-2111.9	1.5	-2151.9	1.5	141.9	0.1	-101.9	0.2
G_{GB} ^e	-1371.4	1.0	-1339.3	1.0	-37.6	0.0	59.6	0.2
G_{PB} ^f	-1261.1	1.0	-1284.8	1.0	-50.0	0.1	73.7	0.2
$G_{GB,np}$ ^g	37.9	0.0	38.6	0.0	2.1	0.0	-2.8	0.0
$G_{PB,np}$ ^h	43.7	0.0	44.0	0.0	3.6	0.0	-3.8	0.0
$G_{GB,solv}$ ⁱ	-1279.4	1.0	-1300.7	1.0	-35.5	0.0	56.8	0.2
$G_{PB,solv}$ ^j	-1217.3	1.0	-1240.8	1.0	-46.4	0.1	69.9	0.2
$U_{vac} + G_{GB,sol}$ ^k	-3391.4	1.0	-3452.6	1.0	106.4	0.1	-45.1	0.1
$U_{vac} + G_{PB,sol}$ ^l	-3329.2	1.0	-3392.7	1.0	95.5	0.1	-32.0	0.1

Calculated average $\Delta\Delta U_{vac} + \Delta\Delta G_{solv}$ values, relative to the FimH-Man α Me complex are -7.2 and -7.6 kcal/mol, for the MM-GBSA and MM-PBSA methods respectively. The more favorable vacuum interaction energy in the FimH-Man α PNP complex is, as seen in Table 5.4, a result of both stronger Coulomb and van der Waals interactions. The origin of this difference is further discussed in Section 5.6.4. The negative $\Delta\Delta U_{vac}$ contribution for the FimH-Man α PNP complex is partly reduced by a positive $\Delta\Delta G_{solv}$ contribution. The negative $\Delta\Delta U_{vac}$ term is, however, larger than positive $\Delta\Delta G_{solv}$, giving a total negative $\Delta\Delta G_{bind}^0$ value for the FimH-Man α PNP complex over the FimH-Man α Me complex.

Solute entropy changes need to be incorporated in order to validate the higher affinity of the Man α PNP ligand seen in the MM-PB(GB)SA results. The solute entropy analysis, using the methods discussed in the previous section, is here repeated for the FimH-Man α PNP complex and presented in Table 5.5. Calculating $T\Delta S_{vac}$ of FimH-Man α PNP complex using normal mode analysis together with ideal gas rigid body contributions, the entropy loss of the complex adds up to 21.3 kcal/mol. Here the normal analysis was performed on 264 different starting structures from the trajectory, over which the calculated vibrational entropies were averaged over. This combination of methods includes both the rigid body entropy loss of the ligand in the complex, as well as, entropy penalties paid from restrictions in internal degrees of freedom of both receptor and ligand in the complex.

Table 5.5: Entropy analysis of the FimH-Man α PNP complex. See Table 5.3 for table captions. All units are in kcal/mol.

	Complex	Receptor	Ligand	Delta
TS _{trans,ig} ^a	16.51±0.0	16.49±0.0	12.89±0.0	-12.87±0.0
TS _{rot,ig} ^b	16.74±0.0	16.71±0.0	10.4±0.0	-10.4±0.0
ΔG_{rbt} ^c				10.2
TS _{vib,nm} ^d	1695.4±1.1	1669.4± 1.0	23.8±0.2	2.1±0.8
TS _{conf} ^e	2675.1	2721.80	44.5	-91.2
TS _{conf,lig} ^f	35.75		44.5	-8.75
TS _{conf,rec} ^g	2638.5	2721.80		-83.3

Comparing to the FimH-Man α Me complex, an additional entropy penalty of 3.6 kcal/mol is paid for the FimH-Man α PNP complex. Together with the MM-PBSA contribution, a total ΔG_{bind}^0 of -10.7 ± 0.8 kcal/mol is obtained. Again, the calculated free energy change is in good agreement with the experimental K_d and calculated ΔG_{bind}^0 values of 44 nM and -10.0 kcal/mol from Bouckaert *et al.* [209]. Together with the MM-GBSA calculated binding free energy contribution, a similar relative increase in binding affinity can be seen, although the calculated ΔG_{bind}^0 value of respective complex is far off compared to experiment.

Using the methods of Schlitter and Swanson *et al.*, neglecting configurational entropy changes of FimH, gives a entropy penalty of of 19 kcal/mol. Combined with the MM-PBSA results, this gives a $\Delta\Delta G_{bind}^0$ of only -0.65 kcal/mol relative to the FimH-Man α Me complex. The configurational entropy change of the Man α PNP ligand is here much larger compared to the Man α Me ligand and adds a positive contribution of approximately 8 kcal/mol for the binding free energy change of the FimH-Man α PNP complex. The time evolution of calculated absolute configurational entropies of Man α PNP, in the complex and in the free state, is shown in Figure 5.13. Both the rigid body term and the configurational entropy are calculated, as previously described, using all atoms in the ligand for superpositions to remove translational and rotational motions of the ligand and to calculate rotational fluctuations of the ligand in the complex. Again, the configurational entropy calculation has not converged in the 25 ns trajectory. The calculated configurational entropy difference between the free and bound states is, however, relatively stable, but still shows fluctuations around 1 kcal/mol during the last 5 ns of the simulation.

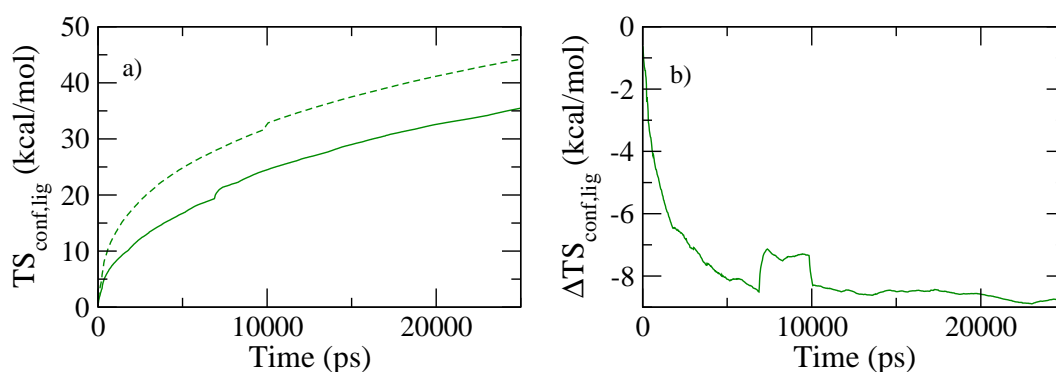


Figure 5.13: a) Calculated $TS_{conf,lig}$ values of the Man α PNP ligand as a function of the simulation time. The absolute configurational entropy of the ligand calculated from the complex trajectory is shown as dashed lines and the corresponding entropy of the ligand in the free state is shown with solid lines. b) $\Delta TS_{conf,lig}$ values calculated from the difference in absolute configurational entropies in the complex relative to the free state.

The calculated $T\Delta\Delta S_{conf,lig}$ of FimH in the FimH-Man α PNP complex, relative to the FimH-Man α Me complex, is greater by approximately 7 kcal/mol, showing that internal motions of FimH are restricted to a greater extent in the FimH-Man α PNP complex. The calculated $T\Delta S_{conf,lig}$ of FimH is again greatly overestimated and these results may not be reliable.

5.6.4 Per-residue free energy decomposition

The per-residue decomposition of $\Delta U_{vac} + \Delta G_{solv,GB}$ is shown in Figure 5.14, for FimH residues in the binding pocket with any atom within 6 Å of the ligand in the first snapshot of the trajectory. These results were averaged over 1250 snapshots, separated by 20 ps, from the complex trajectory. Standard errors of the calculated means are shown as error bars in the figures.

Small shifts can be seen in ΔU_{vac} values for residues in the binding pocket for the two complexes. The Man α PNP ligand interacts more strongly with Asp54 than Man α Me, which in turn, shows more favorable Coulomb interaction with Phe1. Beyond the binding pocket, additional favorable interactions with Man α PNP can be seen, not present for the Man α Me ligand.

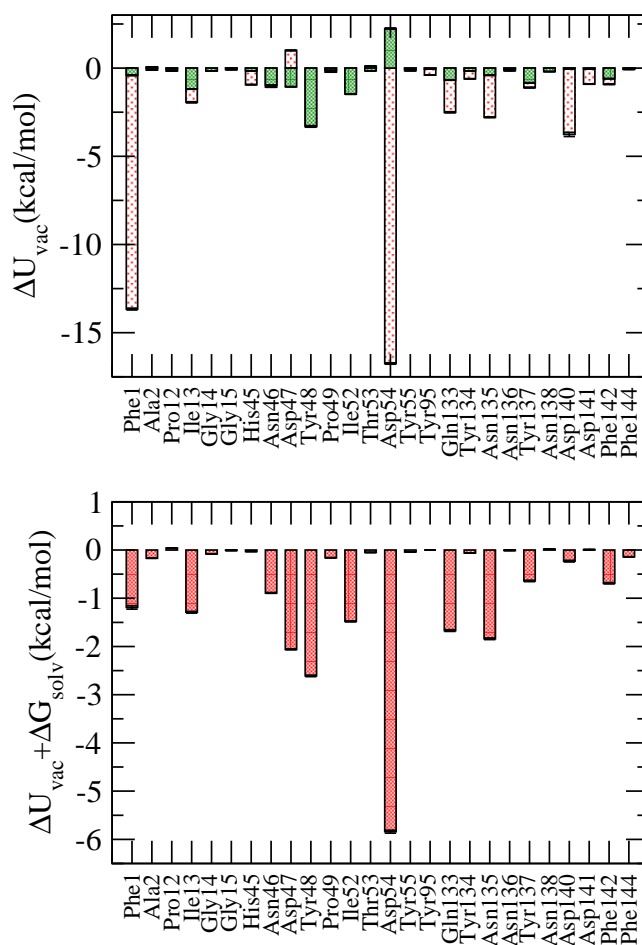


Figure 5.14: a) Decomposition of ΔU_{vac} values of the FimH-Man α PNP complex with respect to contributions from FimH residues. The ΔU_{vdW} contribution from each residue is shown as shaded bars. b) Decomposition of $\Delta U_{vac} + \Delta G_{solv}$ values calculated with the MM-GBSA method.

The two tyrosine residues in the tyrosine gate, Tyr48 and Tyr137, both show more favorable interaction energies, which is mainly a result of van der Waals interactions. Investigations of the complex showed that the aromatic ring of Tyr48 is parallel-stacked with the phenyl ring of the mannoside. An average distance between the ring center of masses of 4.3 Å was calculated. The average distance is not far away from the energy minimum of parallel-stacked benzene which has been calculated to be in the range of 3.5-4.2 Å, depending on the method used [227, 228]. The relative orientation of the phenyl ring of Tyr137 and that of Man α PNP was more flexible in the simulation and the relative orientation can be described by a range of structures ranging from misaligned parallel stacking to T-shaped stacking of different degrees of perturbation. The

ring-ring center of mass distance showed larger fluctuations, with an average distance of 7.4 Å. These geometric differences are reflected in the calculated interaction energies in Figure 5.14a, showing more favorable contributions to the interaction energy from Tyr48. Here the question arises how well these interactions are modeled by the force fields, not explicitly describing these electronic effects. A recent study of Paton *et al.* [229] showed that the interaction energy of the parallel stacked benzene dimer, calculated with AMBER force fields, was overestimated by only 0.13 kcal/mol compared to ab-initio calculations at the CCSD(T)/CBS level of theory. In this study, a mean error of 1 kcal/mol for the AMBER force file was found for a group of complexes interacting predominately with dispersion forces and π - π -stacking. These findings support the use of force fields for studying these interactions which are important for the stability of many lectin-carbohydrate complexes [230].

For comparison with the other complexes investigated in this study, the decomposition of ΔU_{vac} and $\Delta U_{vac} + \Delta G_{solv}$ into contributions of the Man α PNP residues is presented in Figure 5.15. The aglycon makes an additional contribution of -3.9 kcal/mol to the calculated $\Delta U_{vac} + \Delta G_{solv}$ value.

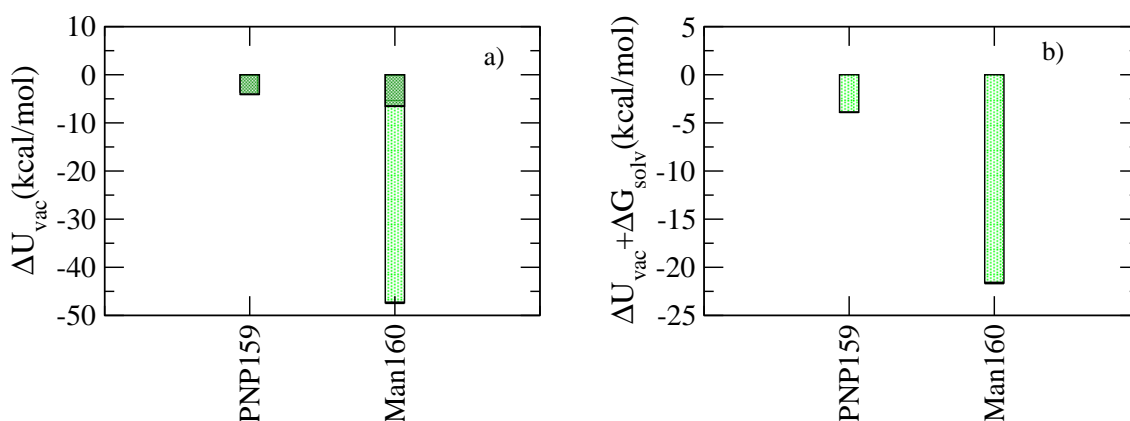


Figure 5.15: a) Decomposition of ΔU_{vac} values of the FimH-Man α PNP complex with respect to contributions from Man α PNP residues. The ΔU_{vdW} contribution from each residue is shown as shaded bars. b) Decomposition of $\Delta U_{vac} + \Delta G_{solv}$ values calculated with the MM-GBSA method.

5.7 FimH-Man₂GlcNac complex

A general feature of mannose-specific bacteria is a preferred binding to short oligomannose chains of N-linked glycoproteins over mannose monosaccharides. The trisaccharide Man α 1,3Man β 1,4GlcNac showed a 20-30 fold increased inhibitory ability for *E. coli* type 1 fimbrial lectin over Man α Me in the studies of Sharon [215]. This trisaccharide is of special interest since it is a common constituent of the N-linked glycoproteins found in the cell surface glycocalyx of many eucaryotic cells. In this section, we investigate the structure and thermodynamics of the ligand Man α 1,3Man β 1,4GlcNac β Me in complex with FimH. This is ligand III in Figure 5.2, and is from hereon abbreviated as Man₂GlcNac.

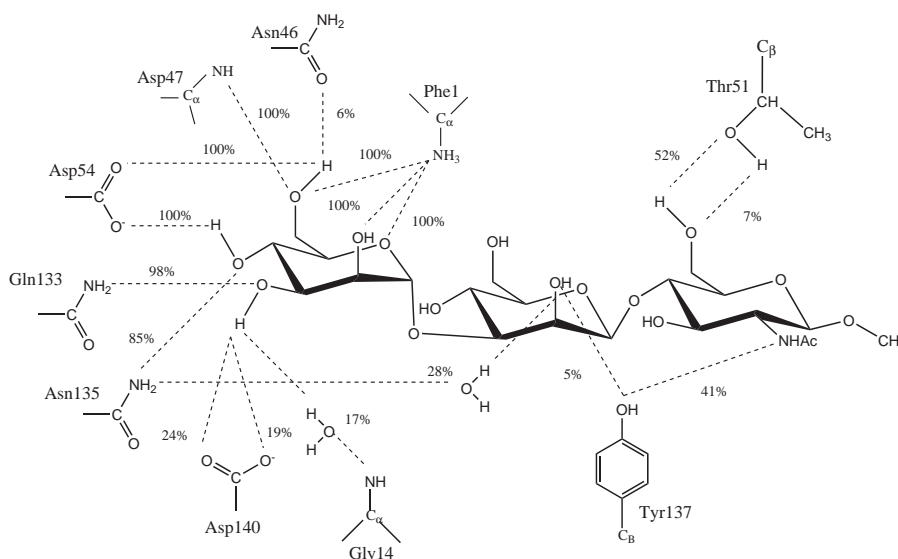


Figure 5.16: Important intermolecular hydrogen bonds formed in the FimH-Man₂GlcNac complex. Hydrogen bond occupancies calculated from the 25 ns complex of the trajectory are shown as percentages in the figure.

5.7.1 Hydrogen bond analysis

Figure 5.16 shows the structure of the FimH-Man₂GlcNac complex and average hydrogen bond occupancies of intermolecular hydrogen bonds, formed in the binding pocket of FimH.

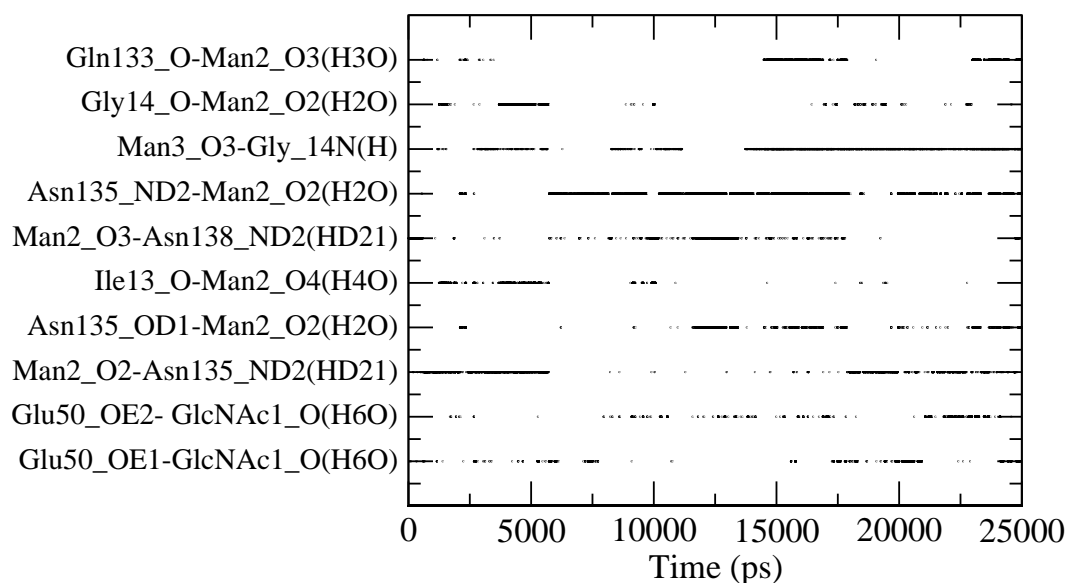


Figure 5.17: Time series of a selected number of water bridged hydrogen bonds formed in the FimH-Man₂GlcNAc complex. Atom names are taken from the PARM99/GLYCAM04 force fields.

The hydrogen bond pattern of Man162 (see Figure 5.2 for residue numbering) is basically the same as that of Man α Me presented in Section 5.5. Additional hydrogen bonds formed by Man161 and GlcNAc160 residues can also be seen. O2 of Man161 forms additional hydrogen bonds with the hydroxyl group of Tyr137 which in turn forms hydrogen bonds with the NAc-arm of GlcNAc160. Oxygen O6 of GlcNAc160 forms additional hydrogen bonds with the hydroxyl group of Thr51. The structure of the complex from these simulations, is very similar to that of the crystallographic structure of a pentasaccharide containing the Man α 1,3Man β 1,4GlcNAc segment in complex with FimH determined by Wellens *et al.* [10]. Analysis of water mediated hydrogen bonds show, at most, five different water molecules hydrogen bonded to FimH and the saccharide at the same time, at a given snapshot of the trajectory. Again, a frequent exchange between structural and bulk water can be seen. Two structural water molecules with relatively high occupancy were observed in this simulation (also shown in Figure 5.16). The first, bridged Gly14 and oxygen O3, and to a lower extent O2 of Man162. The second water molecule bridged O2 of Man161 with the amino group of the Asn135 side chain and was found in almost one out of every three snapshots of the trajectory.

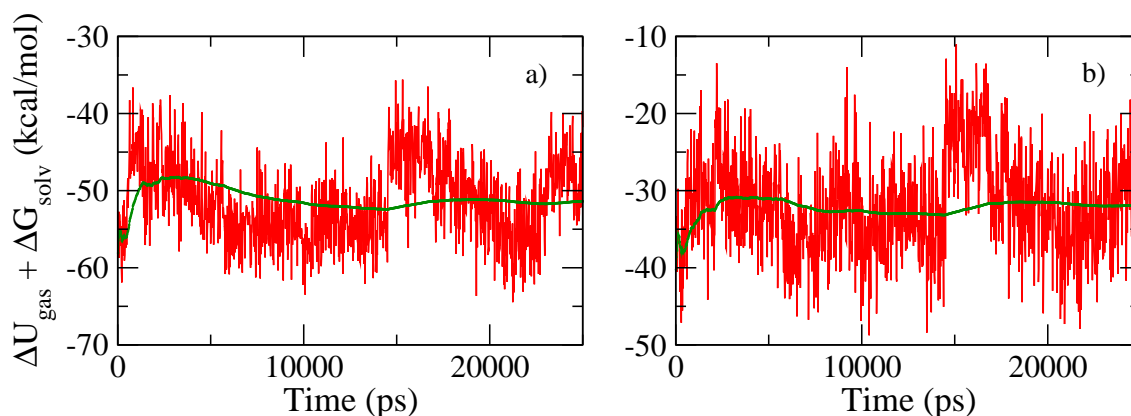


Figure 5.18: Calculated instantaneous $\Delta U_{vac} + \Delta G_{solv}$ values and the running average thereof, for the $\text{Man}_2\text{GlcNAc}$ complex using the MM-GBSA method (a) and the MM-PBSA method (b)

Time series of a selection of water mediated hydrogen bonds with occupancy numbers above 5% are shown in Figure 5.17. The water molecule around GlcNAc160 in the crystallographic structure of Wallens *et al.* was also found in these simulations, here bridging GlcNAc160 and Glu50.

5.7.2 Binding free energy analysis

The FimH- $\text{Man}_2\text{GlcNAc}$ complex is, similarly to the previously investigated complexes, stable during the simulation. The terminal mannosyl unit (Man162) only shows small positional fluctuations in the reference frame of the FimH binding pocket. Figure 5.18 shows instantaneous values and the running average of $\Delta U + \Delta G_{solv}$ calculated with both GB and PB implicit solvation models. Running averages of both methods are in good agreement, except for the shift towards more negative values of the GB implicit solvation model also seen in the previously investigated complexes. Although the fluctuations in instantaneous $\Delta U + \Delta G_{solv}$ values can be large, as a result of large fluctuations in ΔU_{ele} values accompanying relatively small conformational changes, the running averages are relatively stable with time. During the last half of the simulation, fluctuations in the running average of a few kcal/mol can still be seen.

Results from the MM-PB(GB)SA free energy analysis of the FimH- $\text{Man}_2\text{GlcNAc}$ complex are summarized in Table 5.6. Compared to the FimH- $\text{Man}\alpha\text{Me}$ complex, the total

MM-GBSA calculated binding free energy (without including solute entropy changes) is more favorable with both implicit solvation models. The relative free energy difference in the calculated MM-PBSA and MM-GBSA results between the FimH-Man₂GlcNAc and FimH-Man α Me complexes, $\Delta\Delta U_{vac} + \Delta\Delta G_{solv}$, is -12 and -7 kcal/mol respectively. Here the two implicit solvation methods differ and the PBSA method predicts a much higher $\Delta\Delta G_{solv}$ contribution to the binding free energy. Just comparing differences in the ΔU_{vac} contribution, it is clear that both more favorable electrostatic and van der Waals interactions are responsible for the more negative MM-PB(GB)SA binding free energy change of the FimH-Man₂GlcNAc complex. The calculated $\Delta\Delta G_{solv}$ shows a more unfavorable contribution to the binding free energy change in the case of the larger ligand, but the negative $\Delta\Delta U_{vac}$ value outweighs the positive $\Delta\Delta G_{solv}$ contribution, resulting in a more negative binding free energy change of the FimH-Man₂GlcNAc complex.

Table 5.6: MM-PB(GB)SA results from the FimH-Man₂GlcNAc complex. See Table 5.2 for table captions. All units are in kcal/mol.

	Complex	σ_M	Receptor	σ_M	Ligand	σ_M	Delta	σ_M
U_{ele} ^a	-4592.8	1.3	-4692.2	1.2	190.2	0.1	-90.8	0.3
U_{vdw} ^b	-618.0	0.5	-593.7	0.5	5.2	0.1	-29.6	0.1
U_{int} ^c	3179.1	1.0	3126.0	1.0	53.1	0.1	0.0	0.0
U_{vac} ^d	-2031.7	1.5	-2159.9	1.4	248.5	0.2	-120.3	0.2
G_{GB} ^f	-1322.7	1.1	-1330.6	1.0	-66.2	0.1	74.1	0.2
G_{PB} ^g	-1277.0	1.1	-1275.7	1.0	-94.5	0.1	93.3	0.2
$G_{GB,np}$ ^h	37.9	0.0	38.4	0.0	2.9	0.0	-3.4	0.0
$G_{PB,np}$ ⁱ	44.5	0.0	43.86	0.0	5.4	0.0	-4.8	0.0
$G_{GB,solv}$ ^j	-1284.8	1.0	-1292.2	1.0	-63.3	0.1	70.7	0.2
$G_{PB,solv}$ ^k	-1232.6	1.0	-1231.9	1.0	-89.2	0.1	88.5	0.2
$U_{vac} + G_{GB,sol}$ ^l	-3316.5	1.0	-3452.1	1.0	185.3	0.2	-49.6	0.2
$U_{vac} + G_{PB,sol}$ ^m	-3264.3	1.0	-3391.8	1.0	159.3	0.2	-31.8	0.2

The MM-PB(GB)SA free energy analysis is not complete without including the $T\Delta S_{bind,vac}^0$ contribution. Although the MM-PB(GB)SA analysis gave a more negative free energy change of binding the Man₂GlcNAc ligand compared to Man α Me, a large part of this free energy gain is likely to be compensated by a positive $T\Delta S_{bind,vac}^0$ contribution. Absolute vibrational entropies of the solute molecules were calculated and averaged over 253 solute conformations separated by approximately 100 ps from the trajectory of the complex. Using the combination of the ideal gas translational and rotational en-

tropies plus the normal mode calculated vibrational entropy ($T\Delta S_{trans,ig}^0 + T\Delta S_{rot,ig} + T\Delta S_{vib,nm}$), the solute entropy penalty of the FimH-Man₂GlcNAc adds up to 25.7 ± 0.8 kcal/mol, which is 8 kcal/mol larger compared to the corresponding contribution from the FimH-Man α Me complex. Adding this contribution to the MM-PBSA and MM-GBSA calculated free energy changes, gives a total ΔG_{bind}^0 of approximately -6 ± 0.8 and -24 ± 0.8 respectively for the two methods. The MM-PBSA calculated ΔG_{bind}^0 gives a lower affinity of the FimH-Man₂GlcNAc ligand, whereas the MM-GBSA method shows the opposite. This is a direct result of the different ΔG_{solv} contributions from the two methods.

Table 5.7: Entropy analysis of the FimH-Man₂GlcNAc complex. See Table 5.3 for table captions. All units are in kcal/mol.

	Complex	Receptor	Ligand	Delta
$TS_{trans,ig}^0$ ^a	16.52±0.0	16.49±0.0	13.45±0.0	-13.42±0.0
$TS_{rot,ig}$ ^b	16.77±0.0	16.72±0.0	11.52±0.0	-11.5±0.0
G_{rbt}^0 ^c				9.0
$TS_{vib,nm}$ ^d	1724.8±0.7	1670.34± 0.3	55.3±0.1	-0.8±0.8
TS_{config} ^e	2708.0	2721.80	103.2	-116.8
$TS_{config,lig}$ ^f	86.0		102.1	-16.2
$TS_{config,rec}$ ^g	2623.3	2721.8		-98.5

Entropy changes of the solute molecules in the complex were further investigated using the alternative methods of Swanson *et al.* and Schlitter. In both methods, all ring atoms in the saccharide were used in the superposition to remove overall translational and rotational motions of the ligand. Similar to the previously investigated complexes, the calculated change in configurational entropy of the complex is highly overestimated. Table 5.7 shows that the major contribution to the configurational entropy change comes from FimH atomic displacements. Although the calculated loss in configurational entropy of FimH in the complex relative to the free state is poorly estimated by Schlitter’s method, the covariance analysis of FimH atomic fluctuations shows that FimH in the FimH-Man₂GlcNAc complex is more restricted compared to the FimH-Man α Me complex. In order to compare the calculated configurational entropy change to the other ligands, we neglect changes in the configurational entropy of FimH and focus on configurational entropy changes of Man₂GlcNAc. Figure 5.19a shows the calculated configurational entropy of Man₂GlcNAc as a function of the simulation time both in the bound and the free state. It is clear that the configurational

space of the trisaccharide was not fully sampled during the 25 ns trajectory. In Figure 5.19b, the difference in configurational entropy between the bound and the free states is shown. It can here be seen that the calculated ΔS_{conf} has not converged and is still decreasing after 25 ns. Taking the calculated configurational entropy difference after 25 ns, together the rigid body free energy term, ΔG_{rbt}^0 in Table 5.7 gives a entropy loss of approximately 25 kcal/mol for the complex, which is approximately the same as calculated with the NMA-ideal gas treatment. Comparing to the same methods for the FimH-Man α Me complex, both MM-PBSA and MM-GBSA indicate a weaker binding affinity of the Man₂GlcNAc ligand compared to Man α Me.

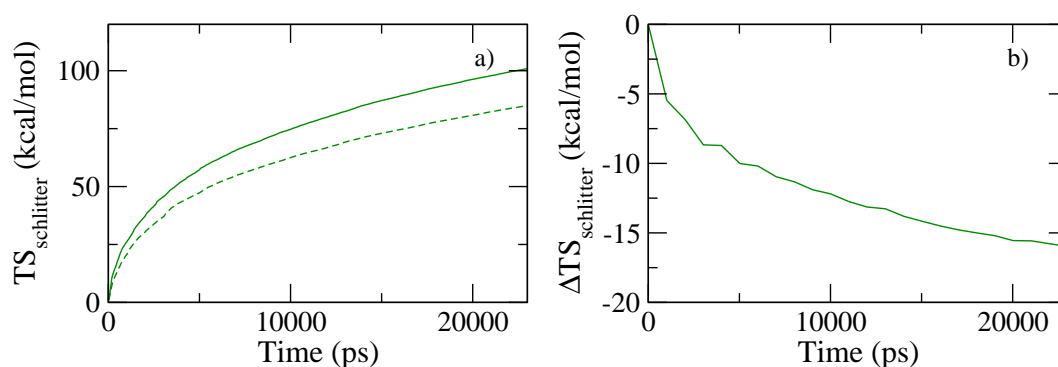


Figure 5.19: a) Calculated TS_{config} values of the Man₂GlcNAc ligand as a function of the simulation time. The absolute configurational entropy of the ligand calculated from the complex trajectory is shown as dashed lines and the corresponding entropy of the ligand in the free state is shown with solid lines. b) Time evolution of the calculated configurational entropy change of Man₂GlcNAc in the complex relative to the free state.

5.7.3 Conformation of Man₂GlcNAc in the complex

Additional energetic penalties that may decrease the binding affinity of Man₂GlcNAc can arise from conformational perturbation of the saccharide in the complex. Since the MM-GBSA binding free energy analysis is performed on complex, receptor and ligand structures taken from the complex trajectory, changes in internal energy on binding will cancel and always add up to zero. In this section, the conformational behavior of the saccharide in both the free and bound state is investigated in terms of fluctuations around the four glycosidic torsional angles, determining the overall conformation of the saccharide. Figure 5.20 shows ϕ - ψ distribution plots calculated from

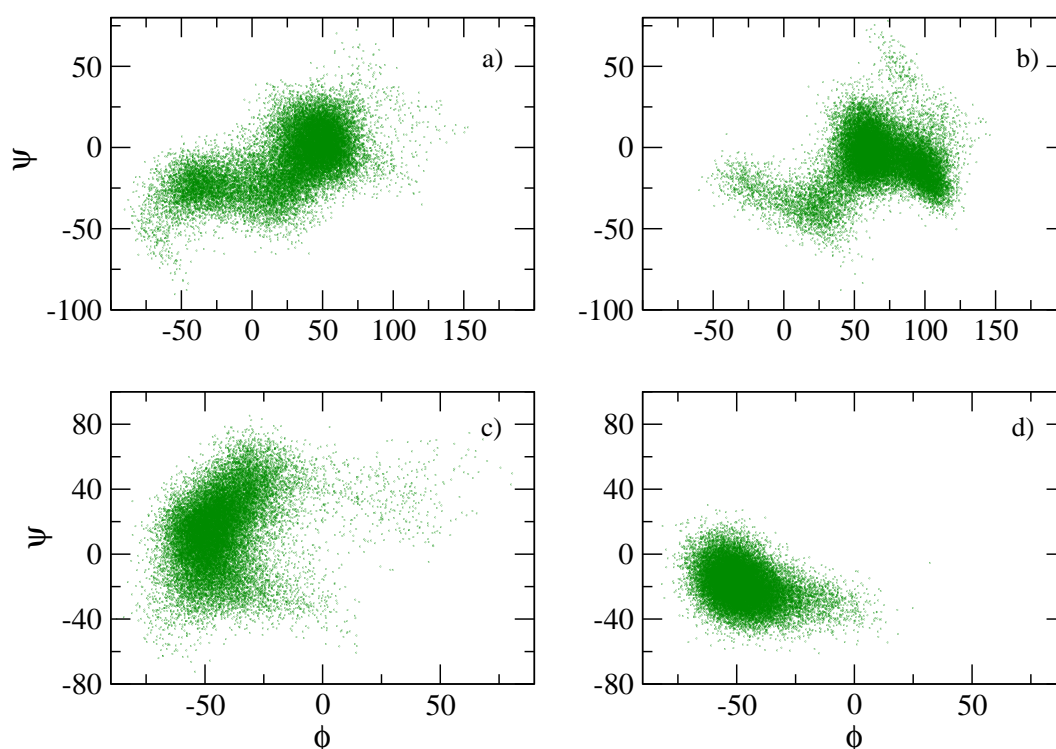


Figure 5.20: ϕ - ψ distribution plots of the $\text{Man}(\beta 1,4)\text{GlcNAc}$ glycosidic linkage calculated for the $\text{Man}_2\text{GlcNAc}$ ligand a) in the free state and b) in the complex. Figures c) and d) show the corresponding ϕ - ψ distribution plots for the $\text{Man}(\alpha 1,3)\text{Man}$ glycosidic linkage. The dihedral angles ϕ and ψ are defined as $\text{Man161:H1-Man161:C1-GlcNAc160:O4-GlcNAc160:C4}$ and $\text{Man161:C1-GlcNAc160:O4-GlcNAc160:C4-GlcNAc160:H4}$ for the $\text{Man}(\beta 1,4)\text{GlcNAc}$ glycosidic linkage and $\text{Man162:H1-Man162:C1-Man161:O3-Man161:C3}$ and $\text{Man162:C1-Man161:O3-Man161:C3-Man161:H3}$ for the $\text{Man}(\alpha 1,3)\text{Man}$ glycosidic linkage. Atom names are taken from the GLYCAM04 force field.

two separate explicit water simulations of the saccharide in the complex and in the free state. Torsional angle distributions of ϕ and ψ (see the caption to Figure 5.29 for angle definitions) for the $\text{Man}\beta 1,4\text{GlcNAc}$ glycosidic linkage are shown in Figures 5.29a (free) and 5.29b (in complex), and similarly for the $\text{Man}\alpha 1,3\text{Man}$ glycosidic linkage in Figures 5.29c (free) and 5.29d (in complex). Although the saccharide still maintains a lot of its conformational freedom in the complex, some reduction in the conformational space can be seen, especially for the $\text{Man}\alpha 1,3\text{Man}$ linkage. Figure 5.29 shows that no major conformational changes of the saccharide take place in the binding process. The ϕ and ψ values sampled in the complex, are well within the distributions observed in the free state. The ϕ and ψ dihedral angles of the $\text{Man}\alpha 1,3\text{Man}$ glycosidic linkage are, in both cases, distributed around the values found in the crystal structure of

Man α 1,3Man β 1,4GlcNac (-58°, -19°) [231]. Calculated ϕ and ψ dihedral angles of the Man β 1,4GlcNac glycosidic linkage are also in good agreement with those found in the crystal structure of (40°, -1°) [231]. These results show that saccharides containing the Man α 1,3Man β 1,4GlcNac motif fit the binding pocket of FimH perfectly and that no major energetic penalties are expected to arise from strain within the saccharide. This is further validated by comparing the difference in internal force field energy (sum of bond, angle and torsion angle terms in the GLYCAM04 force field) of the saccharide in the complex and in the free state. From these simulations, ΔU_{int} is -0.1 ± 0.12 kcal/mol.

5.7.4 Per-residue free energy decomposition

The decomposition of ΔU_{vac} with respect to FimH residues, in Figure 5.21a, shows that the major contribution to the binding affinity comes from interactions of FimH side chains with the mannose residue Man162 in Figure 5.2. Comparing with 5.8a, it can be seen that the calculated ΔU_{vac} values are very similar for the residues in the binding pocket, for the Man α Me and Man₂GlcNac ligands. Slight differences can be seen, indicating minor conformational changes of the terminal mannosyl residue. Most noticeable is the shift towards more negative ΔU_{vac} values for both Asp54 and Asp140. The other differences in Figures 5.21 and 5.8 are due to FimH residues outside the binding pocket. Structural analysis of the FimH-Man₂GlcNac complex revealed that Man161 stacks its B-face onto the phenol ring of Tyr48. Similar stacking could be seen between the B-face of GlcNAc160 and Tyr137 in the tyrosine gate. The nature of these interactions have been attributed to CH/ π , van der Waals or hydrophobic interactions and are a common feature in Lectin-carbohydrate complexes. The study of Spiwok *et al.* [230] calculated interactions energies in the range of -2.8 to -12.3 kcal/mol, for different carbohydrate-aromatic protein side chain interactions, at the MP2/6-311+G(d) level, showing that interactions of this sort can make substantial contributions to the stability of the complex. In this study, it was also shown that GLYCAM/AMBER force field combination could reproduce ab-initio calculated interaction energies with errors within 2 kcal/mol. The phenol ring of Tyr48 and Man161 stacks together in a parallel fashion, with an calculated average center of mass ring distance of 4.5 Å. The relative orientation of Tyr137 and GlcNAc160 is more flexible and can take conformations ranging from a misaligned parallel stacking to T-shaped orientations. The average center of mass distance of the ring atoms in the two residues was 8.0 Å in these

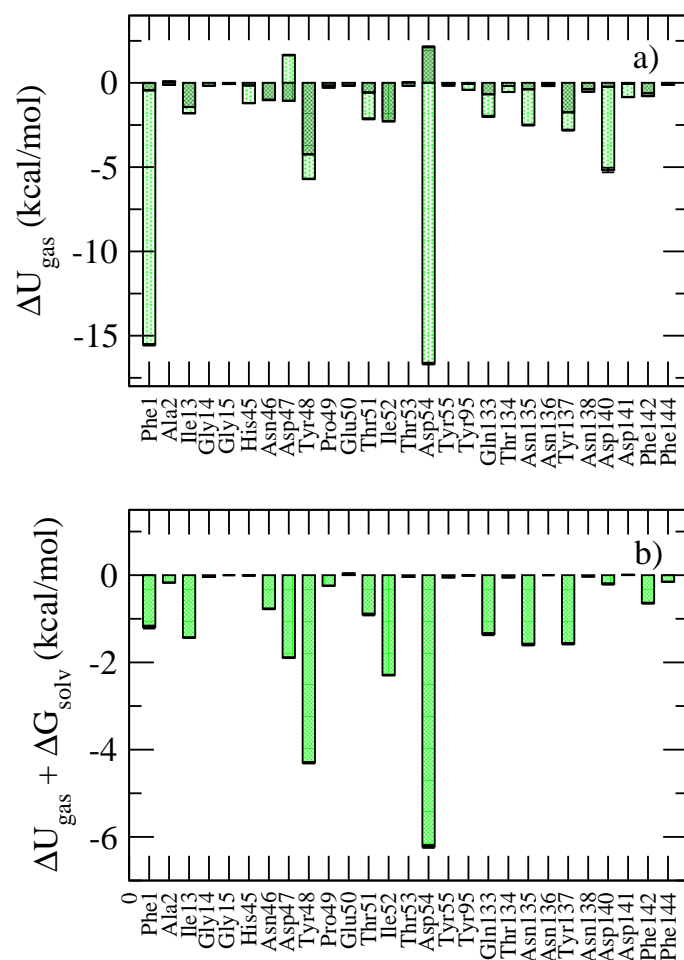


Figure 5.21: a) Per FimH residue decomposition of calculated ΔU_{vac} values of the FimH-Man₂GlcNAc complex. Van der Waals contribution to the interaction energy is shown as shaded bars. b) Per FimH residue decomposition of calculated $\Delta U_{vac} + \Delta G_{solv}$ values using the Generalized Born implicit solvation methodology.

simulations. These differences are reflected in the ΔU_{vac} values in Figure 5.21a, where the decomposed van der Waals interaction energy is more than twice as large for the Tyr48 compared to Tyr137. Comparison with the decomposed MM-GBSA results for the Man α PNP complex, shows a much larger contribution from Tyr48 and Tyr137 for the FimH-Man₂GlcNAc complex, although the π - π stacking, on average, is closer in the Man α PNP complex. Further stabilization of the complex comes from the nonpolar side chain of Ile52. Adding the $\Delta G_{GB,solv}$ contribution to the binding free energy change, it can be seen that interactions between the saccharide and the hydrophobic residues Tyr48, Ile52 and Tyr137 are mainly responsible for the increased MM-GBSA calculated

binding affinity. The large ΔU_{ele} contribution from Asp140 is greatly reduced taking solvation effects into account, showing that this residue is strongly solvated and does not contribute significantly to the binding affinity of the Man₂GlcNAc ligand.

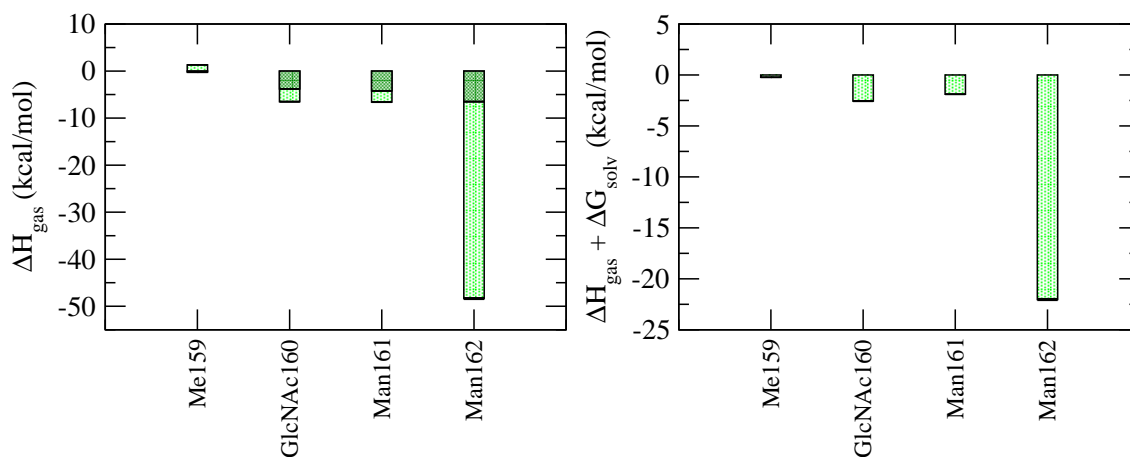


Figure 5.22: Per residue decomposition of calculated a) ΔU_{vac} and b) $\Delta U_{vac} + \Delta G_{solv}$ into contributions from Man₂GlcNAc ligand in the FimH-Man₂GlcNAc complex.

Decomposition of the calculated ΔU_{vac} , with respect to the residues of the saccharide, is presented in Figure 5.22. The terminal Man162, buried in the FimH binding pocket, clearly dominates the negative enthalpy change accompanying the binding process and is very similar to the corresponding value for the FimH-Man α Me complex. Saccharide residues Man161 and GlcNAc160 make smaller, but still favorable, contributions to the overall binding affinity, mainly due to interactions with the hydrophobic residues in the extended binding region. Including the ΔG_{solv} contribution, Man161 and GlcNAc160 make additional contributions of -1.9 and -2.6 kcal/mol respectively to the free energy change of binding.

5.8 FimH-Man₃GlcNAc₂ complex

The branched structure, Man(α 1,6)[Man(α 1,3)]Man(β 1,4)GlcNAc(β 1,4)GlcNAc, is the “core region” of all *N*-glycoproteins with the terminal GlcNAc *N*-glycosidically linked an asparagine residue of the peptide chain in β -mode. A manifold of different glycoprotein structures can form from this uniform core by attachment of additional sac-

charides. The glycoprotein uroplakin Ia has been identified as the urothelial receptor of the FimH lectin [232]. This glycoprotein contains the core region with additional α -mannosyl residues as (1,2)-, (1,3)-, and (1,6)-branches, each in turn terminated by (1,2)-linked mannosyl residues. While the dissociation constant for FimH-uroplakin Ia is approximately 100 nM [232], the affinity for just the core region of the glycoprotein is around 20 nM [210], showing the potential of this pentasaccharide as an effective FimH anti-adhesive. The structure of the ligand is shown in Figure 5.2 and is from hereon abbreviated as $\text{Man}_3\text{GlcNAc}_2$.

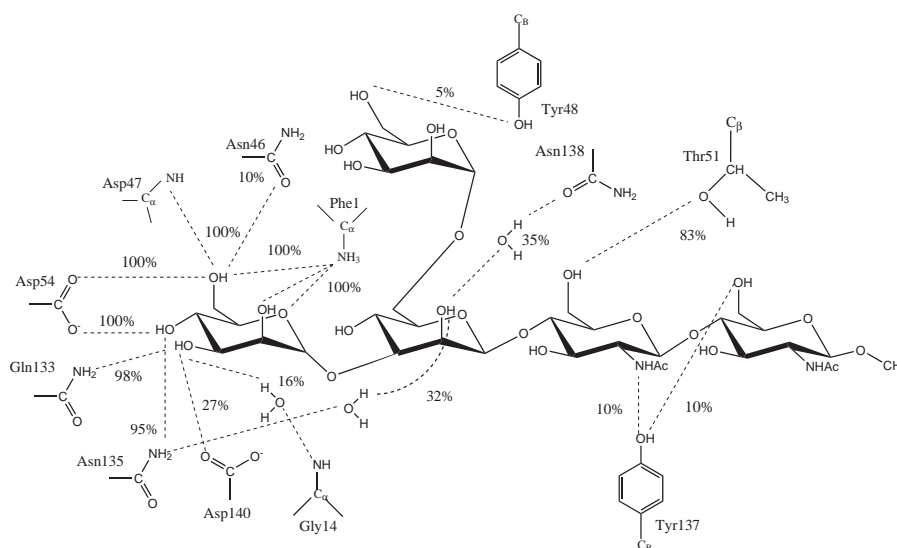


Figure 5.23: Important intermolecular hydrogen bonds formed in the FimH- $\text{Man}_3\text{GlcNAc}_2$ complex. Hydrogen bond occupancies in the 23 ns trajectory are shown as percentages in the figure.

5.8.1 Hydrogen bond analysis

The hydrogen bond analysis of the FimH- $\text{Man}_3\text{GlcNAc}_2$ complex is summarized in Figure 5.23. Hydrogen bond interactions between FimH and the saccharide are very similar to that of the FimH- $\text{Man}_2\text{GlcNAc}$ complex. The hydrogen bonding pattern within the binding pocket is almost the same as for the three previously investigated complexes, and only minor differences can be seen beyond the binding pocket between the FimH- $\text{Man}_3\text{GlcNAc}_2$ and FimH- $\text{Man}_2\text{GlcNAc}$ complexes. Here, Tyr137 forms hydrogen bonds with both the NAc-arm of GlcNAc161 and the O6 hydroxyl group of

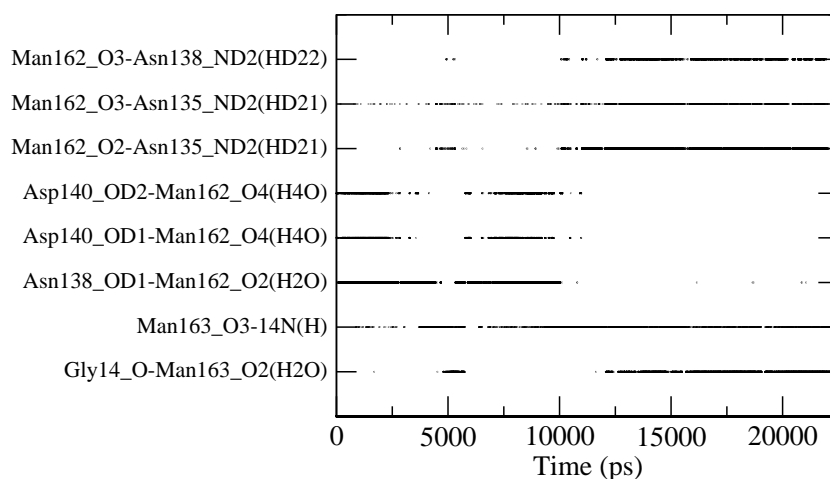


Figure 5.24: Time series of a selected number of water mediated hydrogen bonds observed in the FimH-Man₃GlcNAc₂ complex. Atom names are taken from the PARM99/GLYCAM04 force fields.

GlcNAc160. GlcNAc161, in turn, forms a relatively stable hydrogen bond with Thr51. The α 1,6 linked Man164 is not involved in hydrogen bonding to any greater extent. Spurious and short-lived (with occupancies < 1%) hydrogen bonds were observed with FimH residues Ile13, Tyr137 and Asn138. A longer lived hydrogen bond could be seen between the hydroxyl group of Tyr48 and the O6 hydroxyl group of Man164, with an occupancy of approximately 5% of the trajectory. The two water molecules found in the FimH-Man₂GlcNAc complex were also present in this simulation, together with one additional structural water molecule, bridging the O2 oxygen of Man162 and Asn138. Time series of selected water mediated hydrogen bonds between FimH and Man₃GlcNAc₂ are shown in Figure 5.24. Long residence times, over 10 ns, can be seen for the three structural water molecules.

5.8.2 Binding free energy analysis

Instantaneous values of the calculated MM-PBSA and MM-GBSA results and the running average thereof are plotted in Figure 5.25. Also this complex is relatively stable and does not show any major conformational changes during the simulation. The MM-PBSA and MM-GBSA results are again similar, with a shift towards more negative values for the MM-GBSA results. During the first half of the simulation relatively

large fluctuations in the running averages can be seen. In the last part of the trajectory the running averages are, however, relatively stable. The MM-GBSA free energy analysis of the FimH-Man₃GlcNAc₂ complex is summarized in Table 5.8. The presented results are averaged over 1124 snapshots separated by 20 ps of the 23 ns trajectory.

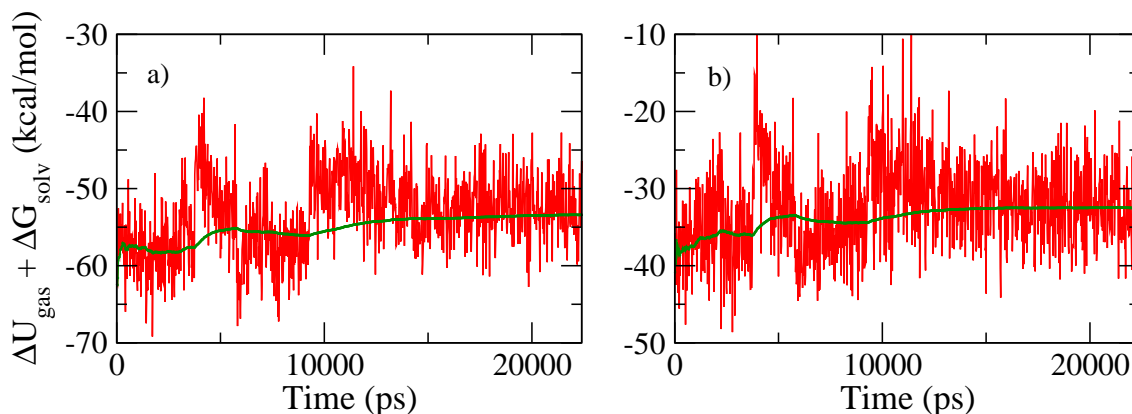


Figure 5.25: Calculated instantaneous $\Delta U_{vac} + \Delta G_{solv}$ values and the running average thereof for the Man₃GlcNAc₂ complex, using the MM-GBSA method (a) and the MM-PBSA method (b).

The similarity in FimH-ligand interactions between Man₃GlcNAc₂ and Man₂GlcNAc, shown in the structural analysis in the previous section, is reflected in the calculated binding free energies. The two additional sugar residues do not make any major contributions to the binding free energy. Calculated $\Delta U_{vac} + \Delta G_{solv}$ values add up to -51.8 ± 0.7 and -32.5 ± 0.5 for the MM-GBSA and MM-PBSA methods respectively, which is approximately 2 and 1 kcal/mol more negative compared to the same calculations for the FimH-Man₂GlcNAc complex. Looking at the contributions to the calculated binding free energy change in Table 5.8, it can be seen that more favorable Coulomb and van der Waals interactions between FimH and Man₃GlcNAc₂ are mainly responsible for the shift towards a more negative free energy change of this complex. The entropy analysis of the solute molecules is presented in Table 5.9. The loss in rotational and translational entropy calculated from ideal gas partition functions, together with the change in vibrational entropy, calculated from a normal mode analysis over 225 snapshot separated by 100 ps of the trajectory, give a entropy penalty of approximately 32 ± 0.9 kcal/mol for the FimH-Man₃GlcNAc₂ complex. Combining the solute entropy change to the MM-PBSA and MM-GBSA results gives ΔG_{bind}^0 values of -0.5 ± 0.9 kcal/mol and -19.61 ± 0.9 respectively, which both are far away from the

expected binding free energy change of -10.6 kcal/mol (at 25⁰C)[210]. Furthermore, these results indicate a lower binding affinity of the Man₃GlcNAc₂ ligand, compared to all previously investigated ligands. Here, the calculated binding affinity is in clear disagreement with experiment, which as seen in Table 5.9, is a result of a much higher entropy penalty for the larger ligand compared to the other ligands. Calculating solute

Table 5.8: MM-PB(GB)SA results from the FimH-Man₃GlcNAc₂ complex. See Table 5.2 for table captions. All units are in kcal/mol.

	Complex	σ_M	Receptor	σ_M	Ligand	σ_M	Delta	σ_M
U_{ele} ^a	-4461.0	1.6	-4670.1	1.6	304.3	0.2	-95.2	0.3
U_{vdw} ^b	-617.3	0.6	-589.1	0.5	4.5	0.1	-32.8	0.1
U_{int} ^c	3208.6	1.1	3121.7	1.0	86.9	0.2	0.0	0.0
U_{vac} ^d	-1869.7	1.9	-2137.5	1.8	395.8	0.3	-127.9	0.3
G_{GB} ^e	1371.1	1.4	-1352.9	1.4	-98.0	0.2	79.9	0.2
G_{PB} ^f	1337.5	1.4	-1297.0	1.4	-141.1	0.2	100.6	2.2
$G_{GB,np}$ ^g	39.0	0.0	38.5	0.0	4.2	0.0	-3.7	0.0
$G_{PB,np}$ ^h	46.2	0.0	44.0	0.0	7.4	0.0	-5.2	0.0
$G_{GB,solv}$ ⁱ	-1332.1	1.4	-1314.4	1.4	-93.8	0.1	76.2	0.2
$G_{PB,solv}$ ^j	-1291.3	1.3	-1253.0	1.3	-133.7	0.2	95.4	0.3
$U_{vac} + G_{GB,sol}$ ^k	-3201.7	1.1	-3451.9	1.0	302.0	0.2	-51.8	0.2
$U_{vac} + G_{PB,sol}$ ^l	-3161.0	1.1	-3390.5	1.1	262.1	0.2	-32.5	0.2

entropy changes, using the methods of Schlitter and Swanson *et al.* gives a different result. Using the combination of $T\Delta S_{conf,lig} + \Delta G_{rbt}$ gives an entropy penalty of 24.5 kcal/mol, which together with the MM-GBSA and MM-PBSA shows a stronger binding affinity of the Man₃GlcNAc₂ ligand compared to Man₂GlcNAc. In the calculation of these results, alignments of the saccharide to remove translational and rotational motions were done by superposing the ring atoms (C1,C2,C3,C4,C5 and O5 in the GLYCAMO4 force field) of all saccharide residues. Figure 5.26 shows that the calculated configurational entropy of the ligand has not converged. Fluctuations of several kcal/mol of $T\Delta S_{conf,lig}$ can also be seen for the last 10 ns segment of the trajectory. In this estimate of the entropy change, the entropy related to internal degrees of freedom of FimH is not included. Comparing the results in Table 5.9 with those in Table 5.7, a much large entropy penalty from the FimH internal motions in the case of the FimH-Man₃GlcNAc₂ complex can be seen. Although the calculated $T\Delta S_{conf,rec}$ value, as previously discussed, most likely is largely overestimated, it shows that atomic fluctuations are restricted to a greater degree in the FimH-Man₃GlcNAc₂ complex, which

Table 5.9: Entropy analysis of the FimH-Man₃GlcNAc₂ complex. See Table 5.3 for table captions. All units are in kcal/mol.

	Complex	Receptor	Ligand	Delta
$TS_{trans,ig}^a$	16.54±0.0	16.49±0.0	13.90±0.0	-13.85±0.0
$TS_{rot,ig}^b$	16.82±0.0	16.72±0.0	12.44±0.0	-12.34±0.0
$G_{rbt}^0^c$				8.0
$TS_{vib,nm}^d$	1761.4±1.4	1671.9± 1.1	95.5±0.4	-5.9±0.9
TS_{config}^e	2748.2	2721.8	183.6	-157.2
$TS_{config,lig}^f$	167.1		183.6	-16.5
$TS_{config,rec}^g$	2592.3	2721.8		-129.5

would further add an entropy penalty to the binding free energy change. Thus, adding the solute entropy contribution adds a large uncertainty to the calculated free energy change and makes it difficult to estimate even relative binding affinities with these methods. Comparing with the results of the previously studied complexes, it can be seen that the agreement with experiment gets worse as the complex gets larger, which seems to be a result of poor estimations of the solute entropy change for the larger ligands.

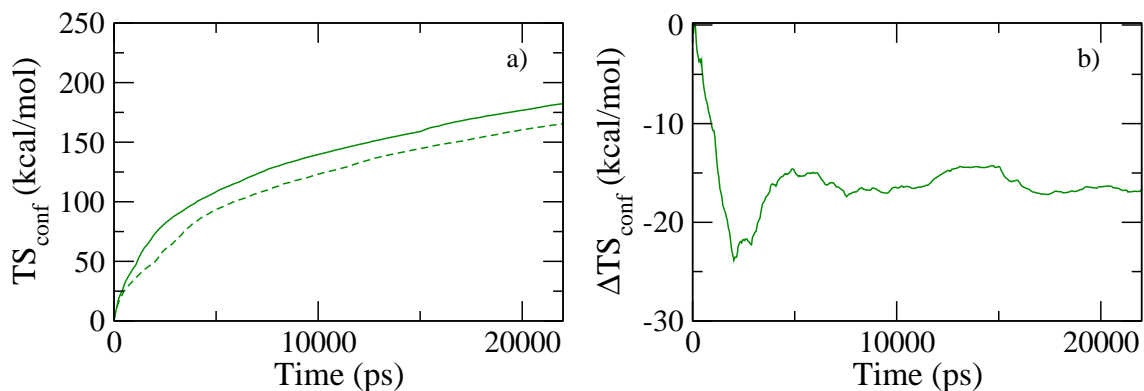


Figure 5.26: a) Calculated TS_{config} values of the Man₃GlcNAc₂ ligand as a function of the simulation time. The absolute configurational entropy of the ligand calculated from the complex trajectory is shown as dashed lines and the corresponding entropy of the ligand in the free state is shown with solid lines. b) Time evolution of the calculated configurational entropy change of Man₃GlcNAc₂ in the complex relative to the free state.

5.8.3 Per-residue free energy decomposition

The decomposition of the MM-GBSA free energy analysis, on to FimH residues with any atom within 6.0 Å from the ligand in the first snapshot of the production run, is shown in Figure 5.27. The structural similarities of the FimH-Man₃GlcNAc₂ and FimH-Man₂GlcNAc complexes are reflected in the MM-GBSA results. Figures 5.27a and 5.21a are very similar for the two complexes. The ΔU_{vac} contribution from residues inside the binding pocket is basically the same as for all the previously investigated complexes. Here, the B-face of Man162 stacks in a parallel geometry against the

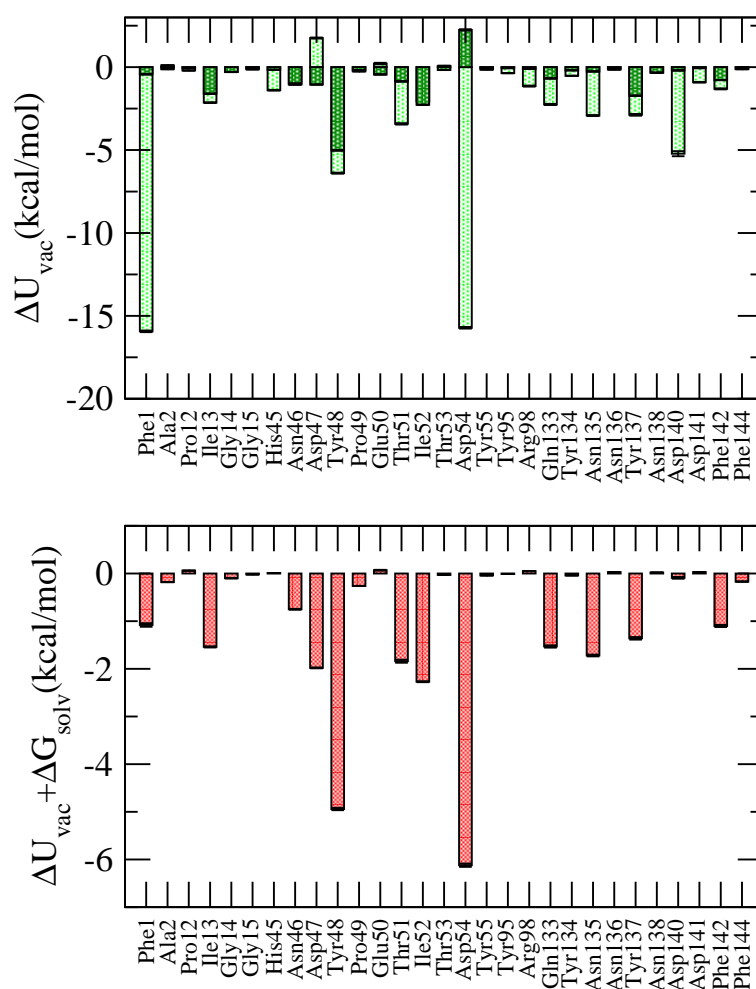


Figure 5.27: Per FimH residue decomposition of calculated ΔU_{vac} values of the FimH-Man₃GlcNAc₂ complex. Van der Waals contribution to the interaction energy is shown as shaded bars. b) Per FimH residue decomposition of calculated $\Delta U_{vac} + \Delta G_{solv}$ values using the Generalized Born implicit solvation methodology.

phenol ring of Tyr48. A slightly larger average center of mass distance of the ring atoms of 4.9 Å is calculated for the FimH-Man₃GlcNAc₂ complex. The stacking of the B-face of GlcNAc161 and Tyr137 is slightly misaligned and is also in this case much more flexible and can take many different geometries. The energetic contributions from these interactions are very similar for the FimH-Man₃GlcNAc₂ and FimH-Man₂GlcNAc complexes. Including the $\Delta G_{GB,solv}$ term in Figure 5.27b, larger differences can be seen between the two complexes. FimH residues Tyr48, Thr51 and Phe142 all make larger and favorable contributions to the calculated binding free energy change. This a result of a larger burial of non-polar surface areas of the non-polar side chains of these residues. Again, although Asp140 is involved in favorable Coulomb interaction with the saccharide, it makes only small contributions to the binding affinity when including the ΔG_{solv} term. The per residue decomposition of interaction energies of Man₃GlcNAc₂, in Figure 5.28, is very similar to that of the Man₂GlcNAc. A slight increase in Coulomb interaction can be seen for GlcNAc161 in Man₃GlcNAc₂ compared to GlcNAc160 in Man₂GlcNAc due to stronger hydrogen bonding for the pentasaccharide with FimH. The two additional saccharide residues Man164 and GlcNAc160 make only relatively small contributions to the overall binding free energy change. Both the α 1,6 linked mannose Man164 and GlcNAc160 point out towards the solvent and do not interact with FimH to any greater extent. Extensions of the saccharide beyond the Man164 residue, commonly found in glycoproteins of the glycocalyx, are not likely to contribute further to the binding affinity to FimH. These residues would be fully solvated and not lower binding affinity due to steric crowding either. Comparing the Man₃GlcNAc₂ and Man₂GlcNAc ligands, these calculations do not show any major enhancement of the binding affinity of the larger ligand that can explain the higher binding affinity of Man₃GlcNAc₂ observed in experiment.

5.8.4 Conformation of Man₃GlcNAc₂ in the complex

Finally, we investigate conformational changes of Man₃GlcNAc₂ in the complex. Figure 5.29 shows distribution plots of all nine glycosidic dihedral angles of the saccharide, calculated from two separate 25 ns explicit water simulations, one in the complex and one in the free state. The glycosidic dihedral angles ϕ and ψ are defined as H1-C1-OX-CX and C1-OX-CX-HX for a (1→X) (X=2,3,4) linkage, analogous to the ϕ_h and ψ_h definition in the IUPAC convention [202]. The additional 1→6 dihedral angle

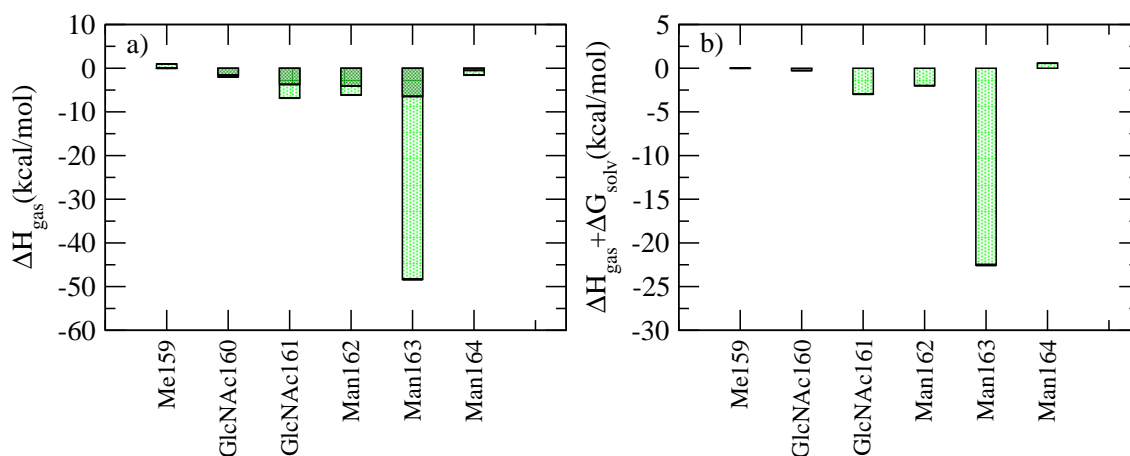


Figure 5.28: Per residue decomposition of calculated a) ΔU_{vac} and b) $\Delta U_{vac} + \Delta G_{solv}$ into contributions from Man₃GlcNAc₂ ligand in the FimH-Man₃GlcNAc₂ complex. The U_{vdW} contribution to the calculated vacuum interaction energies are shown as shaded bars in a).

found in the Man164 α Man162 linkage, ω , is here defined as O6-C6-C5-H5, using atom names from the GLYCAM04 force field. Figure 5.29 shows that the pentasaccharide retains a lot of its dihedral flexibility in the complex. Conformational perturbations can only be seen in two of the dihedral angles. The ψ -Man163(α 1,3)Man162 angle is shifted towards more negative angles, but is still within the boundaries of the range of angle fluctuations observed in the free state. As discussed in Section 5.7.3, this shift is not likely to introduce any larger energetic penalty. The second conformational difference can be seen for the ω -Man164(α 1,3)Man162 dihedral angle. This is a shift into another of the possible staggered conformers of a glycosidic 1-6 linkage and should also not introduce any larger energy penalty. The calculation of differences in molecular mechanics energies from internal force field terms for the saccharide in the complex and in the free state, using two separate trajectories, does, however, give an average internal energy of the saccharide in the bound state of +3 kcal/mol relative to the free state. Given that the conformational space of the saccharide is not fully sampled in both simulations, the relative difference in internal energy might still change with time.

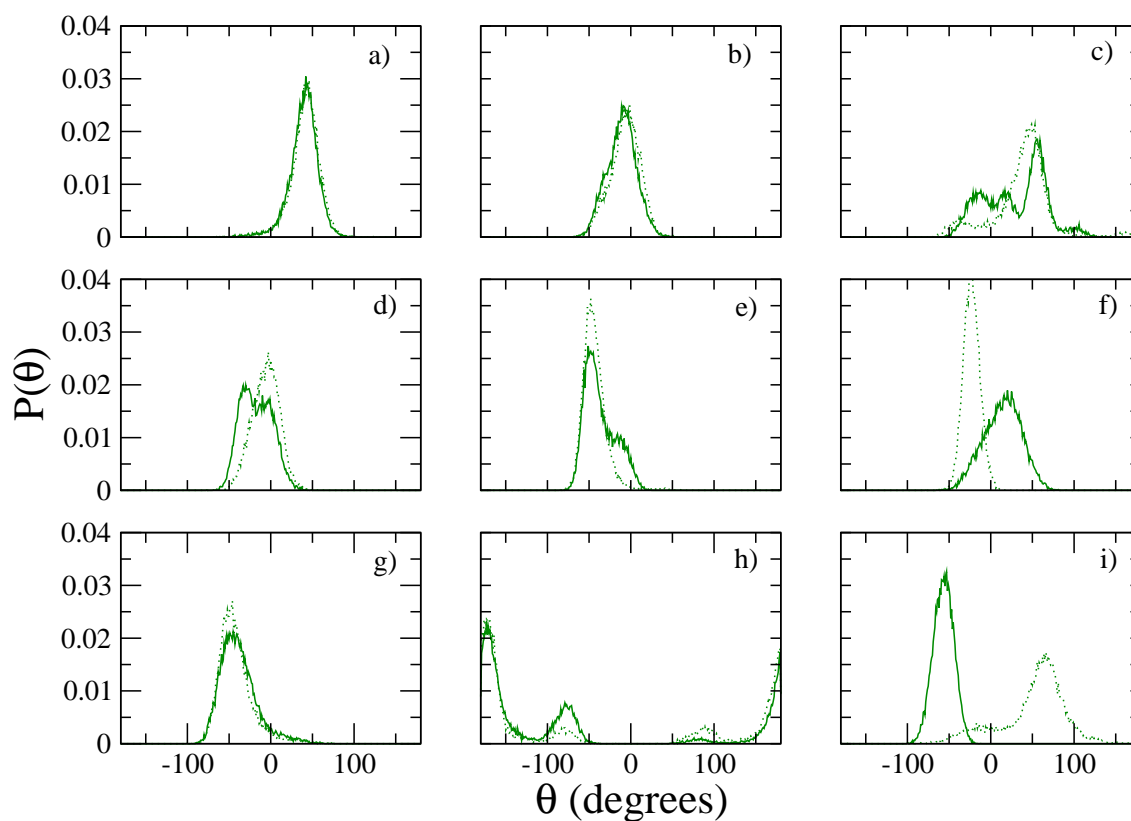


Figure 5.29: Glycosidic dihedral angles of the $\text{Man}_3\text{GlcNac}_2$ ligand in the complex (full lines) and in the free state (dashed lines) a) ϕ -GlcNAc161(β 1,4)GlcNAc160 b) ψ -GlcNAc161(β 1,4)GlcNAc160 c) ϕ -Man162(β 1,4)GlcNAc161 d) ψ -Man162(β 1,4)GlcNAc161 e) ϕ -Man163(α 1,3)Man162 f) ψ -Man163(α 1,3)Man162 g) ϕ -Man164(α 1,6)Man162 h) ψ -Man164(α 1,6)Man162 i) ω -Man164(α 1,6)Man162

5.9 Discussion

In this study, we have investigated the structural and thermodynamic basis of binding of a series carbohydrate ligands to the lectin FimH. The structural investigations revealed a structural similarity of the different complexes in the binding pocket of FimH. The orientation of the terminal mannosyl-group was found to be very similar in all complexes investigated. Due to strong hydrogen bond interactions, only small conformational fluctuations were observed in the binding pocket, for all complexes, during the 23-25 ns long MD-trajectories. The MM-PB(GB)SA approach, together with a series of methods for estimating solute entropy changes, were used to estimate binding free energy changes associated with the complex formation process. Using the combination of MM-PBSA results together with the $T\Delta S_{bind}^0$ contribution estimated by ideal gas rotational and translational entropy calculations, plus changes in vibrational entropy estimated by a normal mode analysis, gives results in good agreement with experiment, at least for the smaller ligands. This combination of methods has in previous studies been shown to outperform traditional docking studies in predicting relative binding free energies of a series of complexes [233]. Calculated ΔG_{bind}^0 values for the FimH-Man α Me and FimH-Man α PNP complexes are -7.1 ± 0.8 and -10.6 ± 0.8 kcal/mol respectively, which compares well to the experimental free energy changes of -7.6 and -10.0 kcal/mol. Using the same methods for the trisaccharide Man₂GlcNAc and the pentasaccharide Man₃GlcNAc₂, in complex with FimH, gives calculated ΔG_{bind}^0 values of -6.0 ± 0.8 and -0.5 ± 0.9 kcal/mol respectively. These results are in disagreement with experimental affinity measurements, showing a higher affinity of both Man₂GlcNAc and Man₃GlcNAc₂ ligands compared to Man α Me. Although the MM-PBSA and MM-GBSA analysis indicated an increased binding affinity for these ligands, the inclusion of the solute entropy change adds increasingly large entropic penalties for the larger ligands. Several different methods available for calculating entropy changes from molecular dynamics simulations have been tested in this study. Convergence problems and inherent errors in the methods can lead to contradictory results when using different methods to estimate solute entropy changes. Calculating entropies of biomolecules is, still to this date, a problem for computational studies and puts a limitation of the predictive power of free energy changes calculated with end-point approaches. The main appeal of the MM-PB(GB)SA approach is the ability to decompose the free energy change into contributions on a residue basis. Contributions from residues in binding pocket revealed ligand-specific interactions responsible for the binding affinity. From

this study, it can be seen that the main stabilization of all investigated complexes come from hydrogen bond interactions between the terminal α -D-mannosyl residue of the ligands and FimH residues in the binding pocket of FimH. The calculated results in this work identify FimH residues Phe1, Asp47, Asp54, Gln133 and Asn135 as the main contributors to the specific binding and recognition of mannose-containing saccharides. Beyond the binding pocket, hydrogen bonding does not make any major contributions to the stability of the complexes investigated. Instead, non-polar interactions between, either non-polar patches of carbohydrate residues or the non-polar aglycons, with non-polar residues in the extended binding site make important contributions to the stability of the complexes. The results presented here show that stacking interactions between carbohydrate residues and aromatic side chains can make substantial contributions to the stability of the complex. The calculated results further indicate a rather large contribution to the binding free energy from structural water molecules in the binding interface of the complex. Structural analysis from explicit water simulations revealed a number of water molecules involved in simultaneous hydrogen bonding with the receptor and the ligand. Including one water molecule explicitly in the free energy analysis, showed that water molecules at the binding interface can make important contributions to the stability of the complex.

In summary, these calculations have revealed several interesting features of FimH-carbohydrate interactions. FimH-mannose binding is enthalpy driven, mainly due to strong hydrogen bonding interactions with a subset of FimH residues in the binding pocket. Stabilizing effects of larger saccharides and mannosides are mainly due to van der Waals interactions with non-polar FimH residues in the extended region of the binding pocket. The free energy analysis presented in this study indicates that mannosides carrying aromatic or aliphatic aglycons, with limited internal degrees of freedom, have great potential as future anti-adhesion drugs against type 1 fimbriated *E. coli* uropathogenic infections.

Bibliography

- [1] T. K. Lindhorst. *Essentials of Carbohydrate Chemistry and Biochemistry*. WILEY-VCH Verlag GmbH, 2007.
- [2] P. M. Rudd, T. Elliot, P. Cresswell, I. A. Wilson, and R. A. Dwek. *Science*, 291:2370, 2001.
- [3] N. Sharon. *Biochim. Biophys. Acta.*, 1760:527, 2006.
- [4] J. Rojo, J. C. Morales, and S. Penadés. *Current Topics in Chemistry*, 218:45, 2002.
- [5] D. Zopf and S. Roth. *Lancet*, 347:1017, 1996.
- [6] D. Neumann, C. M. Lehr, H. P. Lenhof, and O. Kohlbacher. *Adv. Drug Deliv. Rev.*, 56:437, 2004.
- [7] A. G. Barrientos, J. M. de la Fuente, T. C. Rojas, A. Fernández, and S. Penadés. *Chem. Eur. J.*, 9:1909, 2003.
- [8] M. Kleinert, N. Röckendorf, and T. K. Lindhorst. *Eur. J. Org. Chem.*, 2004.
- [9] D. Choudhury, A. Thompson, V. Stojanoff, S. Langermann, J. Pinkner, S. J. Hultgren, and S. D. Knight. *Science*, 285:1061, 1999.
- [10] A. Wellens, C. Garofalo, H. Nguyen, N. Van Gerven, R. Slättegård, J. P. Hernalsteens, L. Wyns, S. Oscarson, H. De Greve, S. Hultgren, and J. Bouckaert. *PLoS ONE*, 3:e2040, 2008.
- [11] S. Weisberger and J. R. Helliwell. *J. Chem. Soc. Faraday Trans.*, 89:2667, 1993.
- [12] J. H. Naismith, C. Emmerich, J. Habash, S. J. Harrop, and J. R. Helliwell. *Acta Cryst.*, D50:847, 1994.

- [13] T. Dam. and F. Brewer. *Chem. Rev.*, 102:387, 2002.
- [14] S. R. Haseley, J. P. Kammerling, and J. F. G. Vliegenhart. *Top. Curr. Chem.*, 218:93, 2002.
- [15] A. Patel and T. K. Lindhorst. *Carbohydr. Res.*, 341:1657, 2006.
- [16] M. Lahmann. *Top. Curr. Chem.*, 2009.
- [17] T. K. Lindhorst. *Top. Curr. Chem.*, 218:201, 2009.
- [18] B. T. Houseman and M. Mrksich. *Top. Curr. Chem.*, 218:1, 2002.
- [19] C. C.M. Appeldoorn, J. A.F. Joosten, F. Ait el Maate, U. Dobrindt, J. Hacker, R. M. J. Liskamp, A. Salam Khan, and R. J. Pieters. *Tetrahedron: Asymmetry*, 16:361, 2005.
- [20] P. Polaverpau and C. Ewig. *J. Comput. Chem.*, 13:1255, 1992.
- [21] M. Dowd, A. French, and P. Reily. *Carbohydr. Res.*, 61:1255, 1994.
- [22] M. Dowd, P. Reily, and A. French. *J. Comput. Chem.*, 13:102, 1992.
- [23] C. Edge, U. Singh, R. Bazzo, G. Taylor, R. Dwek, and T. Rademacher. *Biochemistry*, 29:1971, 1990.
- [24] R. J. Woods, C. J. Edge R. A. Dwek, and B. Fraser-Reid. *J. Phys. Chem.*, 99:3832, 1995.
- [25] F. Corzana, M. S. Motaiwia, C. Hervé du Penhoat, S. Perez, S. M. Tschampel, R. J. Woods, and S. B. Engelsen. *J. Comput. Chem.*, 25:573, 2004.
- [26] S. L. Lee, P. G. Debenedetti, and J. R. Errington. *J. Chem. Phys.*, 112:204511, 2005.
- [27] Q. Liu, R. K. Schmidt, B. Teo, P. A. Karplus, and J. W. Brady. *J. Am. Chem. Soc.*, 119:7851, 1997.
- [28] A. Almond and J. K. Sheehan. *Glycobiology*, 10:329, 2002.
- [29] R. J. Woods, A. Pathiaseril, M. R. Wormald, C. J. Edge, and R. A. Dwek. *Eur. J. Biochem.*, 258:372, 1998.

-
- [30] G. Liang, R. Schmidt, H. A. Yu, D. Cumming, and J. Brady. *J. Phys. Chem.*, 100:2528, 1996.
- [31] R. A. Bryce, I. H. Hillier, and J. H. Naismith. *Biophys. J.*, 81:1373, 2001.
- [32] A. Pathiaseril and R. Woods. *J. Am. Chem. Soc.*, 122:331, 2002.
- [33] C. S. Pereira, D. Kony, R. Baron, M. Müller, W. F. van Gunsteren, and P. H. Hünenberger. *Biophys. J.*, 90:4337, 2006.
- [34] D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, K. M. Merz, B. Wang, D. A. Pearlman, M. Crowsley, S. Bronzell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross, and P. A. Kollman. *Amber 8*. University of California, San Francisco, 2004.
- [35] E. Lindahl, B. Hess, and D. van der Spoel. *J. Mol. Mod.*, 7:306, 2001.
- [36] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. *J. Comput. Chem.*, 26:1781, 2005.
- [37] B. R. Brooks, R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan, and M. Karplus. *J. Comput. Chem.*, 4:187, 1983.
- [38] M. Karplus J. McCammon, J. B. Gelin. *Nature*, 267:585, 1977.
- [39] P. L. Freddolino, A. S. Arkipov, S. B. Larson, A. McPherson, and K. Schulten. *Structure*, 14:437, 2006.
- [40] F. Jensen. *Introduction to Computational Chemistry*. Wiley, 2008.
- [41] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. *J. Am. Chem. Soc.*, 117:5179, 1995.
- [42] M. R. Shirts, J. W. Pitera, W. C. Swope, and V. J. Pande. *J. Chem. Phys.*, 119:5740, 2003.
- [43] T. E. Cheatham III, P. Cieplak, and P. A. Kollman. *J. Biomol. Struct. Dyn.*, 16:845, 1999.
- [44] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura. *J. Chem. Phys.*, 79(2):926, 1983.

- [45] H. J. C. Berendsen, J. R. Grigera, and T. P. Staatsma. *J. Phys. Chem*, 91:6269, 1987.
- [46] P.E. Smith and M. Pettitt. *J. Chem. Phys.*, 95:8430, 1991.
- [47] C. L. Brooks, B. M. Pettitt, and M. Karplus. *J. Chem. Phys.*, 83:5897, 1985.
- [48] P. P. Ewald. *Ann. Phys.*, 64:253, 1921.
- [49] R. W. Hockney and J. W. Eastwood. *Computer Simulations Using Particles*. McGraw-Hill, 1981.
- [50] M. Bhandarkar, R. Brunner, C. Chipot, A. Dalke, S. Dixit, P. Grayson, J. Gullingsrud, A. Gursoy, D. Hardy, J. Héning, W. Humphrey, D. Hurwitz, N. Krawetz, S. Kumar, M. Nelson, J. Phillips, A. Shinozaki, G. Zheng, and F. Zhu. *NAMD User's Guide version 2.6*. 2008.
- [51] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J.R. Haak. *J. Chem. Phys.*, 81:3684, 1984.
- [52] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2002.
- [53] M. Anirban and C. Charusita. *Molec. Phys.*, 102:681, 2004.
- [54] P. Hünenberger. *Adv. Polymer Sci.*, 172:105, 2005.
- [55] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. *J. Comput. Phys.*, 23:327, 1977.
- [56] M.P. Allen and D.J. Tildesley. *Computer Simulations of Liquids*. Oxford Science, 1987.
- [57] M. Yoneya, H. J. C. Berendsen, and K. Hirasawa. *Mol. Sim.*, 13:395, 1994.
- [58] H. Grubmumlller, H. Heller, A. Windemuth, and K. Schulten. *Mol. Sim.*, 6:121, 1991.
- [59] D. Hamelberg, J. Mongan, and J. A. McCammon. *J. Chem. Phys.*, 120:11919, 2004.
- [60] G. M. Torrie and J. P. Wallau. *J. Comput. Phys.*, 23:187, 1977.

-
- [61] R. J. Loncharich, B. R. Brooks, and R. W. Pastor. *Biopolymers*, 32:523, 1992.
- [62] H. S. Ashbaugh, L. R. Pratt, M. E. Paulaitis, J. Clohcy, and T. L. Beck. *J. Am. Chem. Soc.*, 127:2908, 2005.
- [63] T. Astley, G. G. Birch, M. G. B. Drew, and P. M. Rodger. *J. Phys. Chem. A.*, 103:5080, 1999.
- [64] D. Bertolini, M. Cassettari, P. Grigolini, M. Ferrario, and G. Salvetti. *Adv. Chem. Phys.*, 62:277, 1985.
- [65] J. Martí, J. A. Padro, and E. Guàrdia. *J. Chem. Phys.*, 105:639, 1996.
- [66] D.C. Rapaport. *The Art of Molecular Dynamics Simulation - 2nd edn.* Cambridge Univeristy Press, 2004.
- [67] K. N. Swamy and E. Clementi. *Bioploymers*, 28:1497, 1989.
- [68] S.-T. Lin, M. Blanco, and W. A. Goddard. *J. Chem. Phys.*, 119:22, 2003.
- [69] D. C. Rapaport. *Mol. Phys.*, 50:1151, 1983.
- [70] J. Martí, J. A. Padro, and E. Guàrdia. *J. Chem Phys.*, 105:639, 1996.
- [71] G. Ruocco, M. Sampoli, and R. Vallauri. *J. Chem. Phys.*, 96:6167, 1992.
- [72] P. L. Chau and A. J. Hardwick. *Mol. Phys.*, 93:511, 1998.
- [73] J. R. Errington and P. G. Debenedetti. *Nature*, 409:318, 2001.
- [74] M. S. Lee and M. A. Olson. *Biophys. J.*, 90:864, 2006.
- [75] S. Park and K. Schulten. *J. Chem. Phys.*, 120:5947, 2004.
- [76] G. M. Bradbrook, J. R. Forshaw, and S. Perez. *Eur. J. Biochem.*, 267:4545, 2000.
- [77] B. Roux and T. Simonson. *Biophys. Chem.*, 78:1, 1999.
- [78] J. Wang., C. Tan, Y.-H. Tan, Q. Lu, and R. Lou. *Commun. Comput. Phys.*, 3:1010, 2008.
- [79] D. Sitkoff, K. A. Sharp, and B. Honig. *J. Phys. Chem.*, 98:1978, 1994.
- [80] J. A. Reynolds, D. B. Gilbert, and C. Tanford. *Proc. Natl. Adac. Sci. U. S. A.*, 71:2925, 1974.

- [81] A M.L. Connolly. *J. Appl. Cryst.*, 16:548, 1983.
- [82] R. Constanciel and R. Contreras. *Theoret. Chim. Acta*, 65:1, 1984.
- [83] A. Onufriev, D. Bashford, and D. A. Case. *J. Phys. Chem. B*, 104:2000, 2000.
- [84] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, 2000.
- [85] M. Mammen, E. I. Shakhnovich, J. M. Deutch, and G. M. Whitesides. *J. Org. Chem.*, 63:3821, 1998.
- [86] J. M. J. Swanson, R. H. Henchman, and J. A. McCammon. *Biophys. J.*, 86:67, 2004.
- [87] M. K. Gilson, J. A. Given, B. L. Bush, , and J. A. McCammon. *Biophys. J.*, 72:1047, 1997.
- [88] T. Ichiye and M. Karplus. *Proteins*, 11:205, 1991.
- [89] S. Hayward, A. Kitao, and N. Go. *Proteins*, 23:177, 1995.
- [90] J. Schlitter. *Chem. Phys. Lett.*, 215:617, 1993.
- [91] H. Schäfer, A. E. Mark, and W. F. van Gunsteren. *J. Chem. Phys.*, 113:7809, 2000.
- [92] B. Jana, S. Pal, P. K. Maiti, S.-T. Lin, J. T. Hynes, and B. Bagchi. *J. Phys. Chem. B*, 110:19611, 2006.
- [93] S.-T. Lin, P. K. Maiti, and W. A. Goddard III. *J. Phys. Chem. B.*, 109:8663, 2005.
- [94] N. F. Carnaha and K. E. Starling. *J. Chem. Phys*, 53:600, 1970.
- [95] A. Ben-Naim. *Biophys. Chem.*, 101:309, 2002.
- [96] J. R. H. Tame, S. H. Sleigh, A. J. Wilkinson, and J. E. Ladbury. *Nat. Struc. Biol.*, 3:998, 1993.
- [97] Y. Furukawa and I. Morishima. *J. Biol. Chem.*, 279:12983, 2001.
- [98] S. D. Sharrow, K. A. Edmonds, M. A. Goodman, M. V Novotny, and M. J. Stone. *Prot. Sci.*, 14:246, 2005.

-
- [99] J. L. Dashnau, K. A. Sharp, and J. M. Vanderkooi. *J. Phys. Chem. B*, 109:24152, 2005.
- [100] J. W. Brady and R. K. Schmidt. *J. Phys. Chem*, 97:958, 1993.
- [101] S. M. Tschampel and R. J. Woods. *J. Phys. Chem. A*, 107:9175, 2003.
- [102] U. Heugen, G. Schwaab, E. Bründermann, M. Heyden, X. Yu, D. M. Leitner, and M. Havenith. *Proc. Nat. Acad. Sci.*, 103:12301, 2006.
- [103] H. Lis and N. Sharon. *Chem. Rev.*, 98:637, 1998.
- [104] R. U. Lemieux. *Acc. Chem. Res.*, 29:373, 1996.
- [105] B. A. Williams, M. C. Chervenac, and E. J. Toone. *J. Biol. Chem*, 267:22907, 1992.
- [106] H. S. Frank and M. W. Evans. *J. Chem. Phys.*, 13:507, 1945.
- [107] H. S. Ashbaugh and M. E. Paulaitis. *J. Am. Chem. Soc.*, 123:10721, 2001.
- [108] K. R. Gallagher and K. A. Sharp. *J. Am. Chem. Soc.*, 125:9853, 2003.
- [109] N. Smolin and Winter R. *J. Phys. Chem. B*, 108:15928, 2004.
- [110] M. Nakasako. *J. Biol. Phys.*, 28:129, 2002.
- [111] H. M. Chen, V. S. Markin, and T. Y. Tsong. *Biochemistry*, 31:1483, 1992.
- [112] K. A. Dill. *Biochemistry*, 29:7133, 1990.
- [113] Y. Levy and J. N. Onuchic. *Proc. Nat. Acad. Sci.*, 101:3325, 2004.
- [114] C. Chothia. *J. Mol. Biol.*, 105:1, 1976.
- [115] L. X. Dang. *J. Phys. Chem. B*, 102:620, 1999.
- [116] W. L. Jorgensen and J. Tirado-Rives. *PNAS*, 102:6665, 2005.
- [117] W. L. Jorgensen and C. Jenson. *J. Comput. Chem.*, 19:1179, 1997.
- [118] P. Mark and L. Nilsson. *J. Phys. Chem. A*, 105:9954, 2001.
- [119] G. S. Kell. *J. Chem. Eng. Data*, 20:97, 1975.
- [120] K. T. Gillen, D. C. Douglass, and M. J. R. Hoch. *J. Chem. Phys.*, 57:5117, 1972.

- [121] J. Teixeira, M. C. Bellissent-Funel, S. H. Chen, and A. J. Dianoux. *Chem. Phys.*, 1913:31, 1985.
- [122] D. J. Price and C. L. Brooks III. *J. Chem. Phys.*, 121:10096, 2004.
- [123] A. K. Soper and M. G. Phillips. *Chem. Phys.*, 107:47, 1985.
- [124] B. Guillot and Y. Guissani. *J. Chem. Phys.*, 114:6720, 2001.
- [125] W. L. Jorgensen and J. D. Madura. *Mol. Phys.*, 56:1381, 1985.
- [126] M. W. Mahoney and W. L. Jorgensen. *J. Chem. Phys.*, 112:8910, 2000.
- [127] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. *J. Chem. Phys.*, 120:9665, 2004.
- [128] S. W. Rick. *J. Chem. Phys.*, 120:6085, 2004.
- [129] F. Merzel and J. C. Smith. *PNAS*, 99:5378, 2002.
- [130] F. T. Burling, W. I. Weis, K. M. Flaherty, and A. T. Brünger. *Nature*, 271:72, 1996.
- [131] P. E. Mason, G. W. Neilson, J. E. Enderby, M. L. Saboungi, and J. W. Brady. *J. Phys. Chem. B*, 109:13104, 2005.
- [132] U. Essman, L. Petra, M. L. Berkowitz, T. Darden, and L. G. Pedersen. *J. Chem. Phys.*, 103:8577, 1995.
- [133] K. N. Kirschner and R. J. Woods. *PNAS*, 98(19):10541, 1999.
- [134] T. Werder, J. H. Walther, R. L. Jaffe, T. Haliciouglu, and P. Koumoutaskos. *J. Phys. Chem B*, 107:1345, 2003.
- [135] A. Fuchs. *PhD Dissertation, Christian-Albrechts Universität Kiel*, 2005.
- [136] B. Kallies. *Phys. Chem. Chem. Phys.*, 4:86, 2002.
- [137] I. G. Tironi, R. Sperb, P. E. Smith, and W. F. van Gunsteren. *J. Chem Phys.*, 102:5451, 1995.
- [138] R. H. Henchman and J. A. McCammon. *J. Comput. Chem*, 23:861, 2001.
- [139] D. Cremer and J. A. Pople. *J. Am. Chem. Soc.*, 97:1354, 1975.

-
- [140] R. Pethig. *Annu. Rev. Phys. Chem.*, 43:177, 1992.
- [141] G. Otting and K. WÄijthrich. *J. Am. Chem. Soc.*, 11:1871, 1988.
- [142] S. K. Pal, L. Zhao, and A. H. Zewail. *Proc. Natl. Acad. Sci.*, 100:8113, 2003.
- [143] V. A. Makarov, M. Feig, B. K. Andrews, and B. M. Pettitt. *Biophys. J.*, 78:150, 1998.
- [144] R. Kohlrausch. *Ann. Phys.*, 12:353, 1847.
- [145] G. Williams and D. C. Watts. *Trans. Faraday Soc.*, 66:80, 1970.
- [146] J. D. Dunitz. *Chem. Biol.*, 2:709, 1995.
- [147] S. Pal, S. Balaisubramanian, and B. Bagchi. *Phys. Rev. E*, 67:061502, 2003.
- [148] J. D. Dunitz. *Science*, 264:670, 1994.
- [149] B. Lu and C. F. Wong. *Biopolymers*, 79:277, 2005.
- [150] D. Eisenberg and W. Kauzmann. *The Structure and Properties of Water*. Clarendon: Oxford, 1969.
- [151] L. M. Amzel. *Proteins*, 28:144, 1997.
- [152] E. J. Toone. *Curr. Opin. Struct. Biol.*, 4:719, 1994.
- [153] E. Garcia-Hernandez and A. Hernandez-Arana. *J. Am. Chem. Soc.*, 8:1075, 1999.
- [154] J. Zielkewich. *J. Chem. Phys.*, 123:104501, 2005.
- [155] T. S. G. Olsson, M. A. Williams, W. R. Pitt, and J. E. Ladbury. *J. Mol. Biol.*, 384:1002, 2008.
- [156] Y.-K. Cheng and P. J. Rossky. *Nature*, 392:696, 1998.
- [157] B. A. Williams, M. C Chevernac, and E. J. Toone. *J. Biol. Chem.*, 267:22907, 1992.
- [158] T. K. Dam, S. Oscarson, J. C. Sacchettini, and C. F. Brewer. *J. Biol. Chem.*, 273:23826, 1998.
- [159] J. T. Edsall. *J. Am. Chem. Soc.*, 57:1506, 1935.

- [160] P. Jungwirth and D. J. Tobias. *Chem. Rev.*, 106:1259, 2006.
- [161] A. Kohlmeyer, C. Hartnig, and E. Spohr. *J. Molec. Liquids*, 78:233, 1998.
- [162] J. K. Kazimirski and V. Buch. *J. Phys. Chem. A*, 107:9762, 2003.
- [163] F. Schulz and B. Hartke. *Theor. Chem. Acc.*, 114:357, 2005.
- [164] F. Hofmeister. *Arch. Exp. Pathol. Pharmacol.*, 24:247, 1888.
- [165] Y. Zhang and P. S. Cremer. *Curr. Opin. Chem. Biol.*, 10:658, 2006.
- [166] S. Bernèche and B. Roux. *Nature*, 414:73, 2001.
- [167] M. Carrillo-Tripp, H. Saint-Martin, and I. Ortega-Blake. *Phys. Rev. Lett.*, 93:168104–1, 2004.
- [168] Y. Cheng, N. Korolev, and L. Nordenskiöld. *Nucleic Acids Res.*, 34(2):686, 2006.
- [169] L. Vrbka, J. Vondrásek, B. Jagoda-Cwiklik, R. Vácha, and P. Jungwirth. *Proc. Nat. Acad. Sci.*, 103(42):15440, 2006.
- [170] K. D. Collins. *Biophys. J.*, 72:65, 1997.
- [171] F. Schulz and B. Hartke. *Chem. Phys. Chem.*, 3:98, 2002.
- [172] H. Feinberg, S. Park-Snyder, A. R. Kolatkar, C. T. HeiseDagger, M. E. TaylorDagger, and W. I. Weis. *J. Biol. Chem.*, 28:2153, 2000.
- [173] A. Bohne, E. Lang, and C. W. von der Lieth. *J. Mol. Model.*, 4(1):33, 1998.
- [174] D. Berglov and J. Roux. *J. Chem. Phys.*, 100(12):9050, 1994.
- [175] J. Åqvist. *J. Phys. Chem.*, 94:8021, 1990.
- [176] L-A. Näslund, D. C. Edwards, P. Wernet, U. Bergmann, H. Ogasawara, L. G. M. Petterson, S. Myneni, and A. Nilsson. *J. Phy. Chem. A*, 109:5995, 2005.
- [177] E. Guàrdia, D. Laria, and J. Martí. *J. Phys. Chem. B*, 110:6332, 2006.
- [178] F. C. Lightstone, E. Schwegler, M. Allesch, F. Gygi, and G. Galli. *ChemPhysChem*, 6:1475, 2005.
- [179] F. C. Lightstone, E. Schwegler, F. Gygi, and G. Galli. *Chem. Phys. Lett.*, 343:549, 2001.

- [180] S. Varma and S. B. Rempe. *Biophys. Chem*, 124:192, 2006.
- [181] T. Ikeda, M. Boero, and K. Terakura. *J. Chem. Phys.*, 126:034501, 2007.
- [182] M. Cavallari, C. Cavazzoni, and M. Ferrario. *Molec. Phys.*, 102(9-10):959, 2004.
- [183] R. Caminiti, G. Licheri, G. Piccaluga, and G. Pinna. *Chem. Phys. Lett.*, 47:275, 1977.
- [184] D. Jiao, C. King, A. Grossfield, T. A. Darden, and P. Ren. *J. Phys. Chem. B*, 110:18553, 2006.
- [185] K. D. Collins. *Biophys. Chem.*, 119:271, 2006.
- [186] D. R. Linde. *CRC Handbook of Chemistry and Physics, 84th Edition*. 2003-2004.
- [187] P. Mark and L. Nilsson. *J. Phys. Chem. A*, 105:9954, 2001.
- [188] J. E. Enderby. *Electrolyte Monograph, H.H. Willis Physics Laboratory, Bristol, UK*, 1993.
- [189] D. E. Smith and L. X. Dang. *J. Chem. Phys.*, 100:3757, 1994.
- [190] R. Mills and V. M. M. Lobo. *Self-Diffusion in Electrolyte Solutions*. Elsevier: Amsterdam, 1989.
- [191] J. Wang, P. Cieplack, and P. A. Kollman. *J. Comput. Chem.*, 21:1049, 2000.
- [192] M. Patra and M. Karttunen. *J. Comput. Chem*, 25:678, 2004.
- [193] T. Dudev and C. Lim. *Chem. Rev.*, 103:773, 2003.
- [194] A. M. Pyle. *J. Biol. Inorg. Chem.*, 7:679, 2002.
- [195] M. Egli. *Chemistry and Biology*, 9:277, 2002.
- [196] D. Spillmann and M. M. Burger. *J. Cell. Biochem.*, 61:562, 1996.
- [197] W.L. DeLano. DeLano Scientific, Palo Alto, CA, USA., 2002.
- [198] R. J. Woods. *Glycoconj. J.*, 15:209, 1998.
- [199] K. J. Naidoo, D. Denysyk, and J. W. Brady. *Prot. Eng.*, 10(11):1249, 1997.
- [200] C. Höög, C. Landersjö, and G. Widmalm. *Chem. Eur. J.*, 7(14):3069, 2001.

- [201] J. P. Carver. *Pure & Appl. Chem.*, 65(4):763, 1993.
- [202] S. W. Homans. *Biochemistry*, 29:9110, 1990.
- [203] D. van der Spoel and H. J. C. Berendsen. *Biophys. J.*, 72:2032, 1997.
- [204] N. Sharon and H. Lis. *Glycobiology*, 14:53R, 2004.
- [205] M. Aronson, O. Medalia, L. Schori, D. Mirelman, N. Sharon, and I. Ofek. *J. Med. Sci.*, 15:88, 1979.
- [206] C. H. Jones, R. Roth J. S. Pinkner, J. Heuser, A. V. Nicholes, S. N. Abraham, and S. J. Hultgren. *Proc Natl Acad Sci U S A.*, 96:2081, 1995.
- [207] S. N. Amraham, D. Sun, J. B. Dale, and E. H. Beachey. *Nature*, 336:682, 1988.
- [208] C.-S. Hung, J. Bouckaert, D. Hung, J. Pinkner, C. Widberg, A. DeFusco, C. G. Auguste, R. Strouse, S. Langermann, G. Waksman, and S. J. Hultgren. *Mol. Microbiol.*, 44:903, 2002.
- [209] J. Bouckaert, J. Berglund, M. Schembri, E. De Genst, L. Cools, M. Wuhrer, C.-S. Hung, J. Pinkner, R. Slättegård, A. Zavialov, D. Choudhury, S. Langermann, S. J. Hultgren, L. Wyns, P. Klemm, S. Oscarson, S. D. Knight, and H. De Greve. *Mol. Microbiol.*, 55:441, 2005.
- [210] J. Bouckaert, J. Mackenzie, J. L de Paz, B. Chipwaza, D. Choudhury, A. Zavialov, K. Mannerstedt, J Anderson, D. Piérard, L. Wyns, P. H. Seeberger, S. Oscarson, H. De Greve, and S. D. Knight. *Mol. Microbiol.*, 61:1556, 2006.
- [211] P. Aprikian, V. Tchesnokova, O. Yakovenko, C. Larock, B. Kidd, V. Vogel, W. Thomas, and E. Sokurenko. *J. Biol Chem*, 283:7823, 2008.
- [212] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham III. *Acc. Chem. Res.*, 33:889, 2000.
- [213] H. Gohlke and D. A. Case. *J. Comput. Chem.*, 25:238, 2004.
- [214] J. Weiser, P. S. Shenkin, and C. W. Still. *J. Comput. Chem.*, 20:217, 1999.
- [215] N. Sharon. *FEBS Lett.*, 217:145, 1987.

-
- [216] N. Firon, S. Ashkenazis, D. Mirelman, and N. Sharon. *Infect. Immun.*, 55:472, 1987.
- [217] T. K. Lindhorst, C. Kieburg, and U. Krallmann-Wenzel. *Glycoconj. J.*, 15:605, 1998.
- [218] G. Bouvignies, P. Bernadó, S. Meier, K. Cho, S. Grzesiek, R. Brüschweiler, and M. Blackledge. *PNAS*, 102:13885, 2005.
- [219] J. Numata, M. Wan, and E.-W. Knapp. *Genome Inform.*, 18:192, 2007.
- [220] R. Grünberg, M. Nilges, and J. Leckner. *Structure*, 14:683, 2006.
- [221] J. R. Tame, S. H. Sleight, A. J. Wilkinson, and J. E. Ladbury. *Nat. Struct. Biol.*, 3:998, 1996.
- [222] J. E. Ladbury. *Chemistry and Biology*, 3:973, 1996.
- [223] Z. Yu, M. P. Jacobson, Julia Josovitz, C. S. Rapp, and R. A. Friesner. *J. Phys. Chem.*, 108:6643, 2004.
- [224] A. Pigache, P. Cieplak, and F.-Y. Dupradeau. *227th ACS National Meeting, Anaheim, CA, USA*, 2004.
- [225] J. Wang, R. M Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. *J. Comput. Chem.*, 25:1157, 2004.
- [226] J. Wang, W. Wang, P. A. Kollman, and D. A. Case. *J. Mol. Graph. Mod.*, 25:247260, 2006.
- [227] X. Ye, Z-H Li, W. Wang, K. Fan, W. Xu, and Z. Hua. *Chem. Phys. Lett.*, 397:56, 2004.
- [228] G. B. McGaughey, M Gagné, and A. K. Rappé. *J. Biol. Chem.*, 273:15458, 1998.
- [229] R. S. Paton and J. M. Goodman. *J. Chem. Inf. Model.*, 49:944, 2009.
- [230] V. Spiwok, P. Lipovová, T. Skálová, E. Vondrácková, J. Dohnálek, J. Hasek, and B Králová. *J. Comput. Aided. Mol. Des.*, 19:887, 2006.
- [231] W. Warin, F. Baert, and R. Fouret. *Carbohydrate Res.*, 76:11, 1979.

- [232] G. Zhou, W. J. Mo, P. Sebbel, G. Min, T. A. Neubert, R. Glockshuber, X. R. Wu, T. T. sun Sun, and X. P. Kong. *J. Cell Sci.*, 114:4095, 2001.
- [233] B. Kuhn, P. Gerber, T. Schultz-Gasch, and M. Stahl. *J. Med. Chem.*, 48:4040, 2005.

Declaration

I hereby declare that the work presented in this thesis was done by me, under the supervision Prof. Dr. Bernd Hartke, with no other help than the referenced sources in the text. This is my first dissertation and the work has never been used in any other dissertation attempts.

Kiel, 22.08.2009

Mats Eriksson

Acknowledgments

First and foremost, I would like to thank Prof. Dr. Bernd Hartke and Prof. Dr. Thisbe K. Lindhorst for presenting me to this topic and for their guidance and support throughout this project. I also thank Dr. Andreas Fuchs for his kind introduction to the field of glycobiology and modeling of saccharide systems. Furthermore, the generous grant of computer time at the HLRN computing center (Berlin/Hannover), without which this work would not be possible, is greatly acknowledged. Last, but not the least, I would like to thank past and present members of the Hartke working group (in order of appearance) Dr. Bernhard Bandow, Dr. Frank von Horsten, Florian Koskowski, Jan Sielk, Boyke Schönborn, Johannes Dieterich and Ole Carstensen for providing an enjoyable working environment.

Curriculum Vitae

	Mats Henrik Eriksson
Date of birth	04.08.1977
Birth place	Luleå, Sweden
Nationality	Swedish
Address	Bremer Str. 6 24118, Kiel
	Education
1997-2004	Master of Science in Chemistry, Gothenburg University (Sweden) Master thesis in Theoretical Chemistry: "Extended Method for Diabatic Ordering of Vibrational Normal Modes in Reaction Valley Studies"
2000-2004	Master of Science in Computing Science, Gothenburg University (Sweden) Master thesis in the field of Bioinformatics: "ALIGNER: A Toolkit for Protein Structure Comparison"
Dec. 2005-	Ph. D. Student at Christian-Albrechts-Universität zu Kiel under the supervision of Prof. Dr. Bernd Hartke

