Molecular dynamics simulations of glycolipid and carbohydrate binding protein systems in the explicit water environment

Vom Fachbereich Maschinenbau
der Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg
zur Erlangung des akademischen Grades eines Doktor-Ingenieurs genehmigte

DISSERTATION

von

M. Sc. Praveen Konidala
aus Chennai, Indien

Hamburg 2006

1st Supervisor: Prof. Dr.-Ing. Bernd Niemeyer

2nd Supervisor: Prof. Dr.-Ing. Stephan Kabelac

Oral examination: 30.05.2006

Abstract

And the outline of the research work

The carbohydrate derivatives as naturally embedded systems on the cell membrane lipids and proteins are regarded as flexible molecules with wide array of conformations and protein recognition functions. These groups have been awarded an increasing interest in the recent years due to the mounting evidence that they mediate a host of functions including protein-carbohydrate interaction, cell type recognition, cell signaling and development as well as differentiation. Many questions concerning the protein-carbohydrate interactions are therefore associated with conformational behavior of these biomolecules. Even though experiments provide valuable information about the specificity, binding affinities and other equilibrium thermodynamic properties, it is not always possible to exactly characterize the binding region and the forces involved in such bindings. The dynamics involved in these binding processes are of paramount importance, giving way to the flexible adaptive structures in the solution. Because of the intrinsic dynamics bound with the carbohydrate molecules, not all of the required information can be readily obtained from the experimental studies, so theoretical approach is the only alternative to supplement the experimental data.

Molecular dynamics (MD) simulations have been performed extensively in this work to characterize the physico-chemical properties of the glycolipids and their interactions with the carbohydrate binding proteins (also called lectins). The scope of this research was divided into two parts first, the glycolipids were studied in terms of their structure and dynamics in the explicit solvent conditions and second, the lectin-glycolipid interactions were investigated in addition to the structure and dynamics of the ligand-lectin complex systems.

In order to reveal the physico-chemical properties, the glycolipid *n*-octyl-β-D-glucopyranoside (OG) molecule was employed in the simulations, which serve as a model system for the glycolipid structural and dynamical analysis. This type of lipid is classified as a non-ionic surfactant with a glucose head and a small hydrocarbon tail. The experimental application of this has been frequently come across the field of protein solubilization and crystallization (Garavito et al., 1996; Ostermeier and Michel, 1997). These detergents were utilized to isolate the proteins from the cell membrane without denaturation of their biological functions (Stubbs et al., 1976; le Maire et al., 2000). Even though it is a simple glycolipid in

contrast to the more complex combination of lipid structures observed in the cell membranes, the knowledge of their basic interactions with the aqueous solvent and between the lipids themselves can obviously improve our understanding towards the processing and application of glycoconjugates.

The structural packing parameters were calculated for OG from the simulations and compared to the Small Angle Neutron Scattering (SANS) experiments. The geometric property result indicates that OG micelles were more likely to exist in a non-spherical shape, even at the concentration range near to the critical micelle concentration (0.025 M) because of their packing constraints. The micelle shape computed from the principal moment of inertia ratios conclude that the OG aggregate exists in a prolate ellipsoidal form was more favorable than spherical shape. Higher fluctuations in the solution were also observed from the inertial ratios at some instants and found that the OG micelle was more dynamic in nature with a temporary restructuring of its shape to the other forms like cylinders or small bilayer like rods.

The micelle size calculated from the radius of gyration (R_g) and the solvent accessible surface area (ASA) of OG in the solution was constant, despite a couple of infrequent fluctuations in the micelle shape. The interaction of OG atoms with the solvent molecules were analyzed by the radial distribution functions (RDF) calculated between the monomer head group atoms and the oxygen atoms of water molecules. The RDF results proved that due to the axial β stereoisomer conformation the hydroxyl oxygen atoms at the interface to the solvent molecules are more effectively hydrated than, in comparison to the acetal oxygen, ring oxygen and anomeric carbon atoms. The latter atoms were facing in the interior direction towards the micelle hydrocarbon core, thus lacking enough interaction with the water molecules. Estimation of the diffusion coefficients for the OG micelle in the isotropic concentration region is proposed through a modified empirical correlation. The results are agreeing well with the experiments and other theoretical works.

The carbohydrate recognition protein used for the interaction analysis is a legume lectin, which can be easily isolate from the garden pea (*pisum sativum*) by affinity chromatography (Trowbridge, 1974; Einspahr et al., 1986). Several MD simulations were applied in this study to investigate the structure and interaction properties of the pea lectin with various concentrations of OG monomers freely distributed in the aqueous solution. The Ca^{2+} and Mn^{2+} divalent ions needed for the biological activity of the lectin were also added to the simulation cell in

various combinations to observe the influence of carbohydrate bindings on the lectin monomer. The stability of the pea/OG systems was judged by the root mean square deviations and the results state that the flexibility of the lectin conformations in all the simulations was conserved.

The interaction energy patterns show that one or two OG monomers are expected to interact with the lectin surface amino acid residues, and thus, they bind to that lectin site more strongly than the other OG monomers in the solution. This apparently indicates that the lectin has more than one binding site on its surface which accommodates specific carbohydrate conformations to interact strongly in those locations. It was also evident from the simulations that the lectin-ligand complex was stabilized by the Ca^{2+} divalent ion which is located next to the carbohydrate binding site of the lectin. The diffusion coefficients of solvent molecules, free OG monomers and the Ca^{2+}, Mn^{2+} and Cl^{-} ions were calculated and the results were discussed elaborately. Diffusion coefficients of ions were compared with the experimental data. The water diffusion coefficients estimated from the TIP3P water model were in good agreement with the other theoretical results reported previously.

Interestingly, strong binding of OG monomers to the pea lectin have been observed in the system with only Ca^{2+} divalent ions. Absolutely no binding of carbohydrate moieties on the lectin surface was apparent in the simulations without charged ions. These simulations reemphasize the relative importance of divalent ions for the biological activity of the pea lectin and the binding of the carbohydrates wouldn't be possible if they are not present in the solution. Finally, the ligand-lectin complex in the aqueous solution is maintained by the long-lived hydrogen bonds and the water bridges. In addition to the couple of hydrogen bonds between the lectin and OG, the molecular complex was further stabilized by the water bridges. At least one water bridge was involved in the binding of OG to the pea lectin.

Table of contents

1 Introduction

1.1 Brief overview of MD computations

Computational approaches have been instruments in understanding the physical basis of the structure, dynamics and function of the biological macromolecules (Karplus and McCammon, 2002; Saiz and Klein, 2002; Yan et al., 2003; Tavernelli et al., 2003; Woo and Roux, 2005). Simplified models have been applied to understand the inherent principles governing these molecules and will continue to play an essential role in this endeavor. The molecular systems which comprise one or more molecules of interest treated with the implicit or explicit water (e.g. glycolipids or globular proteins in a solvent cell) can be studied by a variety of computational methods. One of these is molecular dynamics, which simulates the movement of all of the particles of a molecular system by iteratively solving Newton's equations of motion (Karplus and McCammon, 2002). Encouraging results have been obtained from this method in the past and the efficient application of these methods and algorithms are developing steadily. All-atom molecular dynamics simulations can provide ultimate detail concerning the physico-chemical properties of the systems as a function of time (Tieleman et al., 2000; Bogusz et al., 2001; Grigera, 2002). Thus, they can be used to address specific questions about the properties of a model system, often more easily than experiments on the actual system. However, validating the results derived from the molecular dynamics simulations with the available experiments is a crucial part in testing the simulation methodology. Comparison of simulation and experimental results serve to test the accuracy of the calculated results and provide criteria for improving the methodology.

The monitoring of the simulated systems as it moves from one conformational state to the next in a phase-space trajectory of molecular dynamics simulations allows one to deduce the timescales for those transitions and to compare the computed properties with the experimental results. It has been more than twenty five years that the group of Karplus and co-workers published their preliminary work on the dynamics of the globular protein bovine pancreatic trypsin inhibitor (BPTI) using the molecular dynamics simulation (McCammon et al., 1977). These studies actually open the channel for mining the vast physiological functions of the biological macromolecule of interest. The simulation was performed in a vacuum with crude potential energy parameters and lasted for only 9.2 ps. They stated that the BPTI was served as a hydrogen molecule of protein dynamics

because of its small size, high stability and relatively accurate X-ray structure available at that time of simulation. The results conclude that the protein interior is fluid-like in that the local atom motions have a diffusion character. Nevertheless, this was an initial foundation for studying such systems with the molecular dynamic simulations. Since then, in the last twenty five yeas of continuous development in this field has brought us to the state where we can treat the systems similar to the way it is handled in the laboratory experimental samples. The reality of the simulations has been exploded in terms of the molecular representations and in the simulation methodologies (Saiz and Klein, 2002; Jang et al., 2004). Interestingly, a microsecond time scale molecular dynamics simulation on a small protein (villin headpiece sub-domain), with an explicit atomic-level representation of both protein and solvent, has marked the beginning of direct and realistic simulations of the folding processes (Duan and Kollman, 2001). Obviously, the future holds a lot for the investigation of these complex systems, besides simultaneously developing computer processor architectures and the efficient software algorithms, such steady improvements will aid us to explore in-depth, the underlying physical mechanisms.

1.2 Glycolipids

The basic biological cell wouldn't exist if it is not protected by the lipid membrane and they might not function proper if it is not mediated by the carbohydrate units on its surface. Clearly, it serves as a marker for the cellular recognition processes (Patthi et al., 1987; Garavito et al., 1996; Gabius, 2000; Bryce et al., 2001; Neumann et al., 2002). Glycolipids are a group of compounds which has a short sequence of carbohydrate residues attached to the lipid chain. They are glycosyl derivatives of lipids such as acylglycerols, ceramides and phenols. To investigate their complex properties in the solution, a simple non-ionic *n*-octyl-β-D-glucopyranoside (OG) structure was used in this work (Konidala et al., 2005). The OG monomer has a single glucose head group attached to the octyl hydrocarbon chain as shown in the Figure 1-1. From the colloidal chemical point of view, they are surfactants with a hydrophilic glucose head group connected to the non-polar hydrocarbon chain via glucosidic linkage. Due to their intrinsic solution properties they are utilized in a broad range of areas starting from household to biology laboratories and to the numerous industrial applications (Gould et al., 1981; Straathof et al., 1988; Lorber et al., 1990; La Mesa et al., 1993; Nilsson et al., 1996; He et al., 2000).

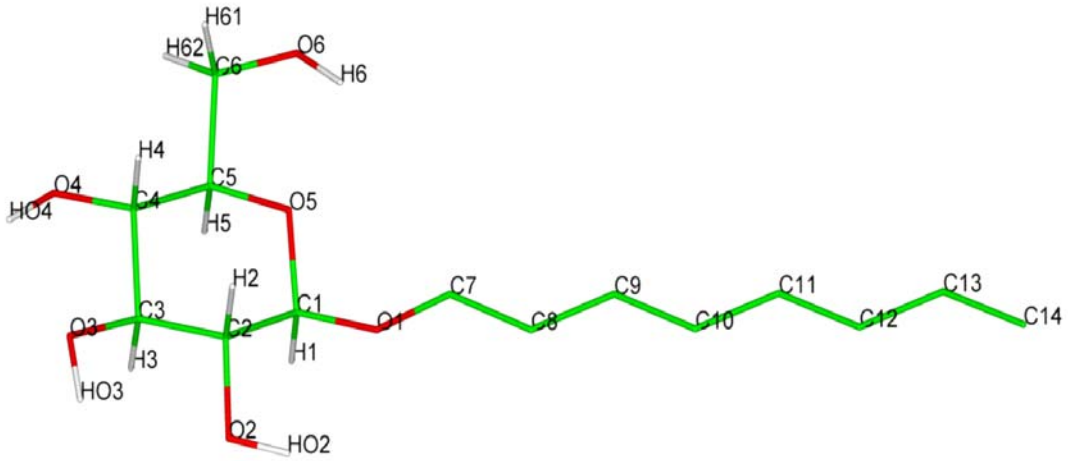

Figure 1-1: Stick representation of the structure of *n*-octyl-β-D-glucopyranoside monomer. Oxygen (red), carbon (green) and hydrogen atoms (white).

The knowledge of the physiochemical properties of OG enables a better understanding of their role in the biological applications as indispensable mild detergents. The experimental evidences have already showed their importance in the field of protein solubilization and crystallization (Lauterwein et al., 1979; Garavito and Rosenbusch, 1986; Eisele and Rosenbusch, 1989; le Maire et al., 2000). The main goal of using these biologically significant detergents in these studies is to conserve the structure of the protein without denaturation and thus investigate their functions in its native state.

Vast collection of experimental and theoretical studies was already available for the OG molecules in spite of their inequalities and differences between them (D'Aprano et al., 1996; He et al., 2000; Bogusz et al., 2000). However, these molecules do have in common that they self-assembled into different structures at above critical micelle concentration (*cmc*) as shown in the Figure 1-2 (Nilsson et al., 1996; Marrink et al., 2000; Tieleman et al., 2000). The structures which are often observed in the solution of the surfactant molecules are spherical (B), ellipsoidal, cylindrical micelles (A), small rod-shaped cylinders, monolayers, bilayers (C), liposome or vesicles (D) and inverted micelles (E) (Figure 1-2). The structural properties are usually examined by the Small Angle Neutron Scattering (SANS) (He et al., 2000), Small Angle X-ray Scattering (SAXS) (He et al., 2002), NMR (Nilsson et al., 1996; Dixon et al., 2002), and Raman and Infrared spectroscopy experiments (Laczko-Hollosi et al., 1992). Most of these methods assume a predefined model to interpret the experimental data. Since the rate of micelle formation and dissociation above *cmc* is quite fast, it is often very difficult to measure these quantities accurately with the experiments (Tanford, 1980;

Nilsson et al., 1996). Moreover, the purity of sample prepared from the commercial process is also expected to have a significant influence on the properties of the OG measured by the experiments (Lorber et al., 1990; He et al., 2000; Thiesen et al., 2006).

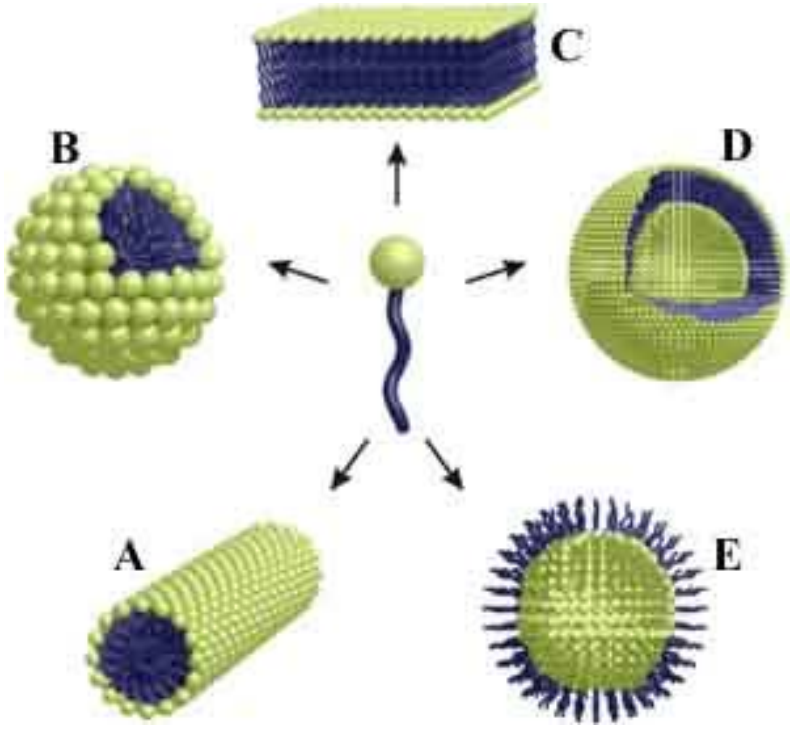

Figure 1-2: Various aggregate structures formed in the aqueous solution at above critical micelle concentration (Elemans et al., 2003). A: cylindrical micelle, B: spherical micelle, C: bilayer, D: vesicles, E: reversed micelle.

Concerning these issues into account molecular models were developed for the OG systems with the incorporation of the reliable and consistent experimental data (La Mesa et al., 1993; D'Aprano et al., 1996; He et al., 2000) into the atomistic models. The phase-space trajectory which stores information on the time evolution of the MD system (with respect to atomic coordinates and velocities) was utilized to characterize the structural and dynamical properties.

1.3 Carbohydrate binding proteins

Lectins are defined as proteins which specifically and reversibly bind carbohydrates (Bridges and Fong, 1979; Debray et al., 1981; Bradbrook et al.,

2000). The term lectin was first coined by Boyd and Shapleigh and is derived from the latin word *legere*, which means 'to choose' (Boyd and Shapleigh, 1954). Their presence is ubiquitous in nature and is found in viruses, bacteria, plants, animals and human beings (Higgins et al., 1983). In general they are sugar specific proteins, when embedded in the membrane cell, recognize the sugar structures in the neighboring cells (Figure 1-3) and tag those cells via glycoprotein and glycolipid molecules (Hamelryck et al., 1998; Sharon and Lis, 2004).

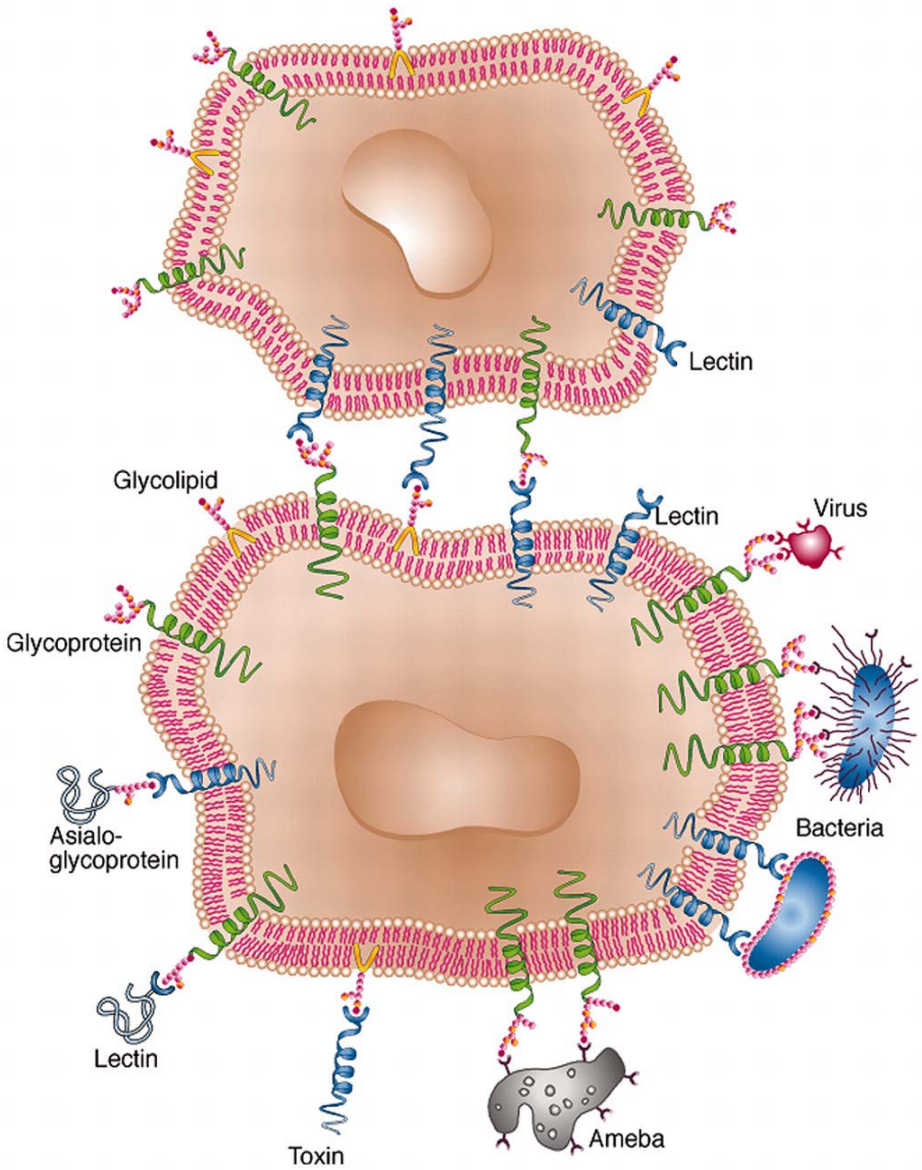

Figure 1-3: Lectins serve as bridge between different cells. They undergo numerous lectin–carbohydrate interactions at the cell surface via glycoconjugates. The attachment of other microorganisms on the cell via surface carbohydrates (Sharon and Lis, 2004) was also shown in the above scheme.

The biological activity of lectins is based on their ability to specifically bind carbohydrates without distorting the covalent structure of the bound sugars, although there are numerous evidences that show no high affinity towards the simple saccharides (Gupta et al., 1997; Bradbrook et al., 2000; Bryce et al., 2001). Moreover they are proteins of non-immune origin unlike enzymes with specific apparent activity, so their physiological functions remain unclear to date (Trowbridge, 1974; Meehan et al., 1982; Rin et al., 1993; Sharon and Lis, 2004), albeit much is known about their carbohydrate binding specificity, sequence of amino acids, and three-dimensional structure of globular proteins. This might be the cause that they are present in a wide spectrum of biological processes. This implies that lectins possess the ability to act as recognition molecules inside of cells, on cell surfaces, and in physiological fluids.

They seem to be involved in a broad range of functions and play an active role in the host defense, infection, control of glycoprotein synthesis, targeting lysosomal enzymes, regulation of cell growth and cycle, modulation of cell-cell interactions and cell-cell interactions in the immune and neural systems (Birdwell and Strauss, 1973; Pletnev et al., 1997; Hamelryck et al., 1998; Gabius, 2001). Initially they were recognized as agglutinins which selectively agglutinate red blood cells based on their blood group type (Sharon and Lis, 1972; Kawai and Takeuchi, 1976). They have also been used for decades as a model system for the study of protein-carbohydrate interactions, because they show remarkable variations in the binding specificities and are easy to obtain and purify (Schwarz et al., 1996; Bradbrook et al., 2000; Neumann et al., 2002). Over the years, a quite impressive amount of structural and experimental data has been generated on the three dimensional structure of lectins (Meehan et al., 1982; Rin et al., 1993 ; Einspahr et al., 1996; Pletnev et al., 1997). The three dimensional crystal structure of the majority of these lectins in particular from the legume family (isolated from plants) are already deposited in the protein data bank (PDB). A hugh collection of these biological macromolecular structures including proteins and nucleic acids deposited in the PDB to date are derived from either X-ray crystallography or NMR spectroscopic methods. The latter method also gives some detail about the dynamics of the macromolecular complex in the solution.

Despite of its increasing number of known structures, solving their structural and dynamic properties in the liquid phase remains a challenging endeavor (Bryce et al., 2001). The legume lectin studied in this work is a pea lectin (Figure 1-4), which is a mitogenic globular protein separated from the seeds of garden pea

(*Pisum sativum*) by affinity chromatography (Trowbridge, 1974; Einspahr et al., 1986). Though the structure of the pea lectin shows very close resemblance to the other leguminous lectins like Concanavalin A (Con A), lentil, and broad bean. Their binding specificity and affinity toward carbohydrate derivatives vary from each other (Schwarz et al., 1993). At neutral pH, pea lectin exist as dimmer of one pair of α-chains each 6,000 Da (magenta color) and one pair of β-chains each 18,000 Da (green color), whereas Con A the only lectin among these exists as a tetramer.

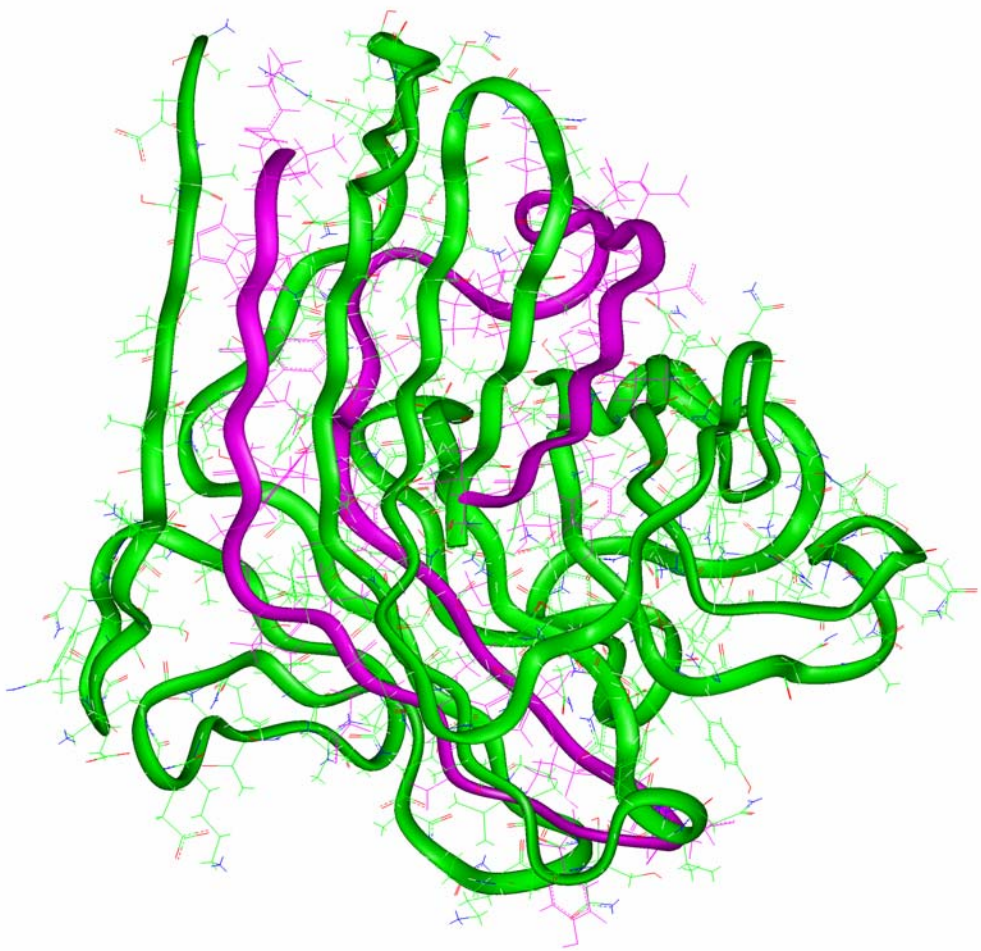

Figure 1-4: Ribbon representation of the backbone pea lectin monomer. The lectin has two chains: β-chain (green) and the α-chain (magenta). The amino acid side chains are shown in lines.

The X-ray crystal structure of pea lectin has been determined long back at 6 Å (Meehan et al., 1982) and 3 Å (Einspahr et al., 1986) resolutions. Also the pea lectin complexed with D-glucopyranose was reported and the atomic coordinates derived from the X-diffraction experiment at 1.9 Å resolution was deposited in the

PDB (entry code 2BQP) (Pletnev et al., 1997). The structural coordinates of pea lectin reported by these authors was employed in the present molecular dynamic (MD) simulations. The ligand molecules and the metal ions were added separately into the simulation cell. Ten different models were performed to simulate the structure, interaction and dynamic properties of the pea lectin accompanied with OG ligands at different concentrations in the aqueous solution. The simulations also reported about the importance of Ca^{2+} and Mn^{2+} metal ions for the biological activity of lectin-ligand complexes (Einspahr et al., 1986; Pletnev et al., 1997), which is in qualitative agreement with these earlier experimental methods. The specific structural and interaction analysis of the pea lectin and OG systems will provide invaluable information of the specific complex interactions at the atomic level and can also be extrapolated to study the properties of other proteins incorporated in the cell membrane.

2 Molecular modeling theoretical background

2.1 Basic quantum chemistry

The quantum mechanics utilizes a set of mathematical formalisms to define a theoretical model for the calculation of molecular properties and geometries. The electronic structure of a molecule is of prime importance in determining these properties. Although molecular mechanics calculations are extremely useful they consider essentially only the position of the nuclei and therefore cannot fully represent the chemical behavior (Leach, 2001). Given the nuclear geometry of a small molecule the complete mathematical description of a system for a time dependent Schrödinger equation can be solved by

$$H(x, y, z, t)\Psi(x, y, z, t) = E_T\Psi(x, y, z, t) \ , \qquad 2.1$$

where E_T is the energy of a molecule, Ψ is the wave function which defines the Cartesian and spin coordinates of the electrons and the nuclei and H is the Hamiltonian operator. In classical mechanics it is a combination of kinetic energy and potential energy of the electrons and nucleic. The Hamiltonian operator acts on the wave function Ψ to determine the behavior or state of the system (Laidler and Meiser, 1982). So when the energy is expressed in terms of momentum, it is said to be a Hamiltonian, which is named after the Irish mathematician Sir William Rowan Hamilton. The Hamiltonian operator can be written as (McQuarrie, 1983),

$$H = \frac{-h^2}{8\pi^2 m_p}\left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}\right) + E_P(x, y, z, t) \ , \qquad 2.2$$

where E_P is the potential energy of the system, h is the Planck's constant and m_p is the mass of a particle. The complete time dependent Schrödinger equation is (Laidler and Meiser, 1982),

$$\left\{\left[\frac{-h^2}{8\pi^2 m_p}\left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}\right)\right] + E_P(x, y, z, t)\right\}\psi = -\frac{h}{2\pi}\frac{d\psi}{dt} \ . \qquad 2.3$$

For stationery states the E_P is independent of time and there is no operator involving time, thus the above equation becomes

$$\left\{\left[\frac{-h^2}{8\pi^2 M}\left(\frac{d^2}{dx^2}+\frac{d^2}{dy^2}+\frac{d^2}{dz^2}\right)\right] + E_P(x,y,z)\right\}\psi(x,y,z) = E\psi(x,y,z) \ . \qquad 2.4$$

This equation is the time independent Schrödinger equation and is frequently written in the compact form

$$H\Psi = E\Psi \ . \qquad 2.5$$

The solution to the Schrödinger is an Eigenvalue problem. As the Hamiltonian operator is the total energy of the system which is composed of the potential energy and kinetic energy components, operated by the wave acting as Eigenfunction to produce an Eigenvalue of the energy for that operator. The kinetic energy of a single electron atom can be described as (McQuarrie, 1983),

$$E_{Ke} = -\frac{\hbar^2}{2m_e}\nabla_e^2 \ , \qquad 2.6$$

$$\nabla_e^2 = \left[\left(\frac{d}{dx}\right)^2 + \left(\frac{d}{dy}\right)^2 + \left(\frac{d}{dz}\right)^2\right] \ , \qquad 2.7$$

where m_e = electron mass (9.11×10^{-31} kg), $\hbar = h/2\pi$ and ∇_e^2 is the Laplacian operator with respect to the position of the electron.

The kinetic energy of the nuclei can be described as (McQuarrie, 1983),

$$E_{KN} = -\frac{\hbar^2}{2m_n}\nabla_N^2 \ , \qquad 2.8$$

$$\nabla_N^2 = \left[\left(\frac{d}{dx}\right)^2 + \left(\frac{d}{dy}\right)^2 + \left(\frac{d}{dz}\right)^2\right] \ , \qquad 2.9$$

where m_n is the mass of the nuclei and ∇_N^2 is the Laplacian operator with respect to the position of the nuclei.

The potential energy (E_p) of a single atom with a charge number Z and the elementary charge *e* is defined as

$$E_P = -\frac{Ze^2}{k_o r} , \qquad 2.10$$

$$k_o \text{ (permitivity)} = 4\pi\varepsilon_o = 1 \qquad 2.11$$

When the molecule has many nuclei and electrons, the kinetic energy of all nuclei and electrons have to be taken into account in addition to the potential interaction (potential energy) of the other atoms surround to it. This makes quantum chemical calculations problematic. For a very simplistic case of single electron (hydrogen) atom with a fixed atomic nucleus of charge number Z at the origin is written (Laidler and Meiser, 1982), in terms of kinetic and potential energies as

$$H(r,R) = E_{Ke} + E_{KN} + E_P \ . \qquad 2.12$$

Because it is assumed that the atom has been pinned at the center, the kinetic energy of the fixed atom is zero, thus the Hamiltonian operator for this simple case (one electron and one nucleus) takes the form

$$H(r,R) = -\frac{\hbar^2}{2m_e}\nabla^2 - \frac{Ze^2}{k_o r} \ . \qquad 2.13$$

The above equation is used with the wave function in the Schrödinger equation to calculate the energy of the system. This equation can be solved accurately for the atoms involving a smaller number of electrons and often fails to the bigger atoms of several electrons. In general these methods utilize a high level of theory to predict the chemical and physical material properties of the system quite accurately by solving the extremely large number of integrals for the electronic motion in a molecule, simply on the basis of one mathematical law i.e. the Schrödinger differential equation and the basic properties of matter formulated therein. The major drawback concerned with the quantum mechanical methods is when dealing with the large systems containing couple of hundreds of atoms, it is very difficult or even impossible to solve the integrals for the motion of thousands of atoms/molecules we are interest in (for example simulation of lectin-ligand complexes, micelles in a solvated cell, interaction of peptides and micelle, solution

conformational properties of proteins, nucleic acids etc.). In such cases the solution to the Schrödinger equation is too complicated and the computations involved in solving the motion of electrons and nuclei in these systems are expensive even to the advanced scientific computer hardware available to date. Therefore many approximations (e.g. Born-Oppenheimer approximation) are invoked to reduce the complexity of solving the equations and thereby introducing new shortcuts or methods to analyze the properties of the system further.

2.2 Born – Oppenheimer approximation

The very first approximation to quantum mechanical systems is made by Born and Oppenheimer. They stated that the mass of the electron is much lighter than the nuclei (approximately 1,837 times than the lightest proton) and the electrons move much faster than the nuclei so they can be decoupled with respect to the motion of nuclei and motion of electrons into two separate equations (McQuarrie, 1983). For a system of two protons and an electron, the motion of the electron was calculated first with the two nuclei fixed at their positions from these equations.

$$E_p = \frac{e^2}{k_o}\left(\frac{1}{r_{AB}} - \frac{1}{r_a} - \frac{1}{r_b}\right), \qquad 2.14$$

$$H = -\frac{h^2}{2m_e}\left(\nabla_a^2 + \nabla_b^2\right) + \frac{e^2}{k_o}\left(\frac{1}{r_{AB}} - \frac{1}{r_a} - \frac{1}{r_b}\right), \qquad 2.15$$

$$\left[-\frac{h^2}{2m_e}\left(\nabla_a^2 + \nabla_b^2\right) + \frac{e^2}{k_o}\left(\frac{1}{r_{AB}} - \frac{1}{r_a} - \frac{1}{r_b}\right)\right]\psi(r;R) = E_e(R)\psi(r;R), \qquad 2.16$$

$$H_e(r;R)\psi(r;R) = E_e(R)\psi(r;R). \qquad 2.17$$

The operator H_e now represents the Hamiltonian without the kinetic energy of the nuclei, because it is considered to be fixed. The energy is calculated with a fixed distance between the nuclei (r_{AB}) and then the new fixed values of (r_{AB}) are chosen and the corresponding energies are determined. The electronic energy E_e here is a function of only the nuclear positions, unlike equation 2.3, where nuclear and electron positions are taken into account explicitly. Since the nuclei are much heavier than the electrons, the electrons adjust instantaneously to the

motion of the nuclei. Thus one can fix the nuclei at some internuclear distance and can solve the electron form of the Schrödinger equation as stated above. This energy can be calculated from the commercial *ab initio* codes (Gaussian, GAMESS, MOPAC, MNDO etc.) for a system of around ~100-1,000 atoms with an affordable cost, and time of computations, respectively. However, it is still a burden to the larger systems, thus often an empirical fit in the form of potential energy surface function (force fields) is usually preferred when dealing with large complex molecules (Leach, 2001; Karplus and McCammon, 2002; Grigera, 2002; Guimaraes et al., 2004).

The second equation represents the motion of nuclei on the potential energy surface setup by the electrons.

$$\left[-\frac{h^2}{2m_n}\left(\nabla_A^2+\nabla_B^2\right)+E_e(R)\right]\phi_n(R_A,R_B)=E_T\phi_n(R_A,R_B)\ , \qquad 2.18$$

$$H_n(R)\phi_n(R) = E_T\,\phi_n(R)\ . \qquad 2.19$$

This approximate separation of the molecular Schrödinger equation such as 2.3 and 2.4 into one for the electron motion and one for the nuclear motion is called the Born-Oppenheimer approximation. The above equation assumes that the nuclei move in an average field of electrons which is averaged through the electron wave function $\psi(r;R)$ in equation 2.17 and the corresponding electron energy E_e is used with the kinetic energy of the nuclei to calculate the total energy of the system. The solution to the above equation 2.18 is referred as quantum dynamics and also requires considerable computing power.

2.3 Molecular dynamics

A great deal of information can be gained from the study of molecular dynamics methods, which solves Newton's second law at each time step to calculate the motion of an individual atom in a system of a very large number of atoms (Allen and Tildesley, 1989; Haile, 1992). The new positions calculated were integrated numerically with the finite difference methods such as velocity Verlet or Leap-frog algorithms.

$$F_i = m_i \times a_i = m_i \frac{dv}{dt} = m_i \frac{d^2 r_i}{dt^2} \quad , \qquad 2.20$$

where F_i is the force exerted on atom i, m_i is the mass of an atom i and a_i is the acceleration of atom i, which is a second derivative of position r_i with time. The force acting on individual atom is termed as a negative gradient of the potential energy surface by the relation

$$F_i = -\frac{dU}{dr_i} = -\nabla_i U \quad . \qquad 2.21$$

Newton's equation of motion can then relate to the derivative of the potential energy (U) and to the changes in the position as a function of time. Combining the above two equation yields

$$-\frac{dU}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \quad . \qquad 2.22$$

To proceed further with the equation of motion the initial positions and the velocities of the system have to be assigned. The initial coordinates for the micelle simulations were taken from the developed model through the glycolipid molecular topology file (Appendix A) and in the case of ligand-lectin simulations it was taken from the x-ray diffraction data through the Protein Data Bank (PDB). The initial velocities were randomly scaled by the Maxwell-Boltzmann distribution

$$f_v(v_x, v_y, v_z)\, dv_x dv_y dv_z = \sqrt{\left(\frac{m}{2\pi k_B T}\right)} \exp\left[\frac{-m(v_x^2 + v_y^2 + v_z^2}{2k_B T}\right] dv_x dv_y dv_z \qquad 2.23$$

where f_v is the probability distribution of atoms having velocities between v and v + dv, v_x^2, v_y^2 and v_z^2 are the velocity components in the three spacial directions and k_B is the Boltzmann's constant. The temperature can be calculated from the velocities using the relation

$$T = \frac{1}{3N} \sum_{i=1}^{N} \frac{|f_v|}{2m_i} \quad , \qquad 2.24$$

where N is the total number of atoms in the system and m_i the mass of an atom i.

Having the coordinates, velocities and the accelerations (from equation 2.20) for each atom at initial time t = 0, it is possible to integrate these quantities for the later time $t' = t + \Delta t$ with the finite difference methods. Leap-frog integrator was used to solve the equation of motion in the simulations with the constant pressure simulation module (NPT ensemble) implemented in the CHARMM (Chemistry at HARvard Macromolecular Mechanics) program (Brooks et al., 1983).

In this algorithm the velocities are calculated at every half time $t + 1/2\,dt$, and then they are used to calculate the position at time $t + dt$ (Allen and Tildesley, 1989; Haile, 1992). In this way the velocities move over the positions and then the position move over the velocities.

$$v_i\left(t + \frac{1}{2}dt\right) = v_i\left(t - \frac{1}{2}dt\right) + \frac{F_i}{m_i}\,dt \quad , \qquad 2.25$$

$$r_i(t + dt) = r_i(t) + v_i\left(t + \frac{1}{2}dt\right)dt \quad . \qquad 2.26$$

The velocities at time t can be approximated by the relation

$$v_i(t) = \frac{1}{2}\left[v_i\left(t - \frac{1}{2}dt\right) + v_i\left(t + \frac{1}{2}dt\right)\right] \quad . \qquad 2.27$$

2.3.1 Force field parameters

The parameters used for the molecular dynamics calculations are derived from the force fields, which are parameterized specifically for small molecular entities with an intention to predict the potential energy interactions for solvated systems of large macromolecules. The parameters for the potential energy surface of a molecule are usually deduced either from quantum mechanical techniques such as *ab initio*, semi-empirical methods or experimental data fitting to the analytical functions. For our simulations we have used CHARMM Carbohydrate solution force field (CSFF) (Kuttel et al., 2002; Palma et al., 2001; Ha et al., 1988) and CHARMM22 (MacKerell et al., 1998; Momany and Rone, 1992) force field

which are parameterized by the experimental data especially for carbohydrates and proteins (Appendix B). The selection of the force field and thus the dynamic programme applied is very important for running the MD simulations (Table 2-1). For example CHARMM (Brooks et al., 1983) and Amber (Cornell et al., 1995; Pearlman et al., 1991) have been explicitly parameterized for biological macromolecules while CVFF (Dauber-Osguthorpe et al., 1998) and Compass (Sun, 1998) have been parameterized specifically for polymers. Extra care should be taken while selecting the force fields prior to the simulation run. The quality of the force field parameters often reflects the end properties calculated from the simulations.

Table 2-1: Selection of various available force field parameters for the MD simulation and their purpose of application

Force field parameters	Purpose and use
CHARMM22 (MacKerell et al., 1998)	General purpose force field parameters for proteins and nucleic acids
CSFF (Kuttel et al., 2002)	Carbohydrates in a solution phase
AMBER/OPLS (Jorgensen et al., 1996)	Biopolymers and carbohydrates
AMBER94 (Cornell et al., 1995)	Proteins and nucleic acids
MMFF (Halgren, 1996)	Biopolymers and drug-like organic molecules
OPLSAA(2000) (Allinger, et al., 1988)	Condensed-phase simulation of peptides
MM2/MM3 (Allinger, 1988; Lii and Allinger, 1998)	Hydrocarbons and molecules with single or remotely spaced functional groups
GROMOS96 (Ott and Meyer, 1996)	General purpose force field parameters for proteins
ESFF	Class II *ab initio* based force field parameters for wide range of atom types
CFF95 (Hagler and Ewig, 1994)	Peptide and protein properties

Moreover the CHARMM force field parameters performs well over a broad range of calculations and simulations, including calculation of interaction and conformation energies, geometries, time dependent dynamic behavior, barriers to rotation, vibrational frequencies, and free energy. There is a wide spectrum of

force field parameters available at present, some of the most important of them for the application of biomolecular simulations are listed in the Table 2-1.

It is clear that the molecular mechanics take a classical approach to calculate the energy of a molecular structure. The molecule is treated essentially as a set of charged point masses which are coupled together with springs (Ha et al., 1988; Kuhn and Rehage, 1997; Reiling et al., 1996). The total energy of a system is calculated using the potential energy surface function also called force field, which is based on the sum of the individual energy components. The form of force field used is CHARMM is

$$U(r) = E_{Internal} + E_{External} \quad , \qquad 2.28$$

$$E_{Internal} = \sum_{bond} \frac{1}{2} k_b (r - r_o)^2 + \sum_{angle} \frac{1}{2} k_\theta (\theta - \theta_o)^2 + \sum_{dihedrals} k_\phi \left[1 + \cos(n\phi - \delta)\right] + \sum_{impr} \frac{1}{2} k_\omega (\omega - \omega_o)^2 + \sum_{u-b} \frac{1}{2} k_u (u - u_o)^2 \quad . \qquad 2.29$$

The potential energy (U) function is mainly divided into bonded (internal) and non-bonded (external) energy terms. The internal energy contributions are further split into bond, angle, dihedral, improper and Urey-Bradley potential functions. The harmonic approximations used in the first two terms in equation 2.29 accounts for the deformation in the bond length and bond angle, with k_b the bond force constant and k_θ the angle force constant. Besides the force constant parameters the ideal bond length r_o and the ideal bond angle θ_o are specific to the type of the atoms involved and must be carefully parameterized (Cornell et al., 1995; Ott and Meyer, 1996; MacKerell et al., 1998). The third term, dihedral angle (or torsion angle) function is a four-atom potential which represents the rotation about an axis defined by the middle bond. The parameters used in this function are the dihedral force constant k_ϕ, n the multiplicity of the function, ϕ the dihedral angle and δ is the phase shift. The fourth term accounts for the out of plane bending with k_ω the improper force constant and ω_o the equilibrium out of plane angle. The improper-dihedral term is applied to maintain planar and chiral configurations in the structures. Additionally, this term provides a better force field near the minimum energy geometry, a consideration that is important for

dynamic calculations and vibrational analysis. The last term comprise of Urey-Bradley harmonic potential, applied to the atoms separated by two bonds (1, 3 interaction), where k_u is the non-bonded distance force constant between 1, 3 atoms and u_o is the ideal distance between 1, 3 atoms. The energies of the bond and angle terms in the above equation can be experimental derived from the spectroscopy and infrared experiments. While for the dihedral energy term, in addition to spectroscopy and infrared experiments it can also be derived from the NMR and empirical data fittings.

The most time consuming part in the molecular dynamic simulations is the calculation of non-bonded interactions which consist of van der Waals and electrostatic interaction energies

$$E_{External} = \sum_{non-bonded} \varepsilon \left[\left(\frac{R_{min}}{r_{ij}} \right)^{12} - \left(\frac{R_{min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} \quad . \qquad 2.30$$

The van der Waals interaction is most often modeled using Lennard-Jones 6-12 potentials for the attraction and repulsion terms (Reiling et al., 1996; Ha et al., 1988). The non-bonded terms are summed for all atom pairs in the system that is separated by 1-4 non-bonded atoms and above. The repulsive force arise at short distances is due to the Pauli exclusion principle. The attractive term arise from the fluctuations in the charge density at longer distances. This attractive term is also referred to as dispersion or London forces. The parameter R_{min} is the van der Waals distance where the Lennard-Jones potential is zero and ε is the energy at the minimum in the potential energy surface. The second term in the above equation is the electrostatic interaction represented by the Coulomb potential. In this term ε_o is the dielectric constant and r is the distance between two atoms having charges q_i and q_j. The third important non-bonded interaction is the hydrogen bonding interaction, which is not treated explicitly in the CHARMM calculation but are accounted implicitly in the van der Waals and electrostatic interactions. These non-bonded interactions are the most important factor for the stability of the biological macromolecule (Kumar and Nussinov, 2002; Allen and Tildesley, 1989; Haile, 1992).

2.3.2 Periodic boundary conditions

Performing the MD simulations in a single solvent cell is usually governed by the surface effects. The interaction or evaporation of the atoms at the cell surface affects the bulk properties of the system weirdly. In order to eliminate the surface effects, periodic boundary conditions (PBC) were applied to the simulations to study the solute properties in a bulk liquid (Heermann, 1986; Haile, 1992) (Figure 2-1).

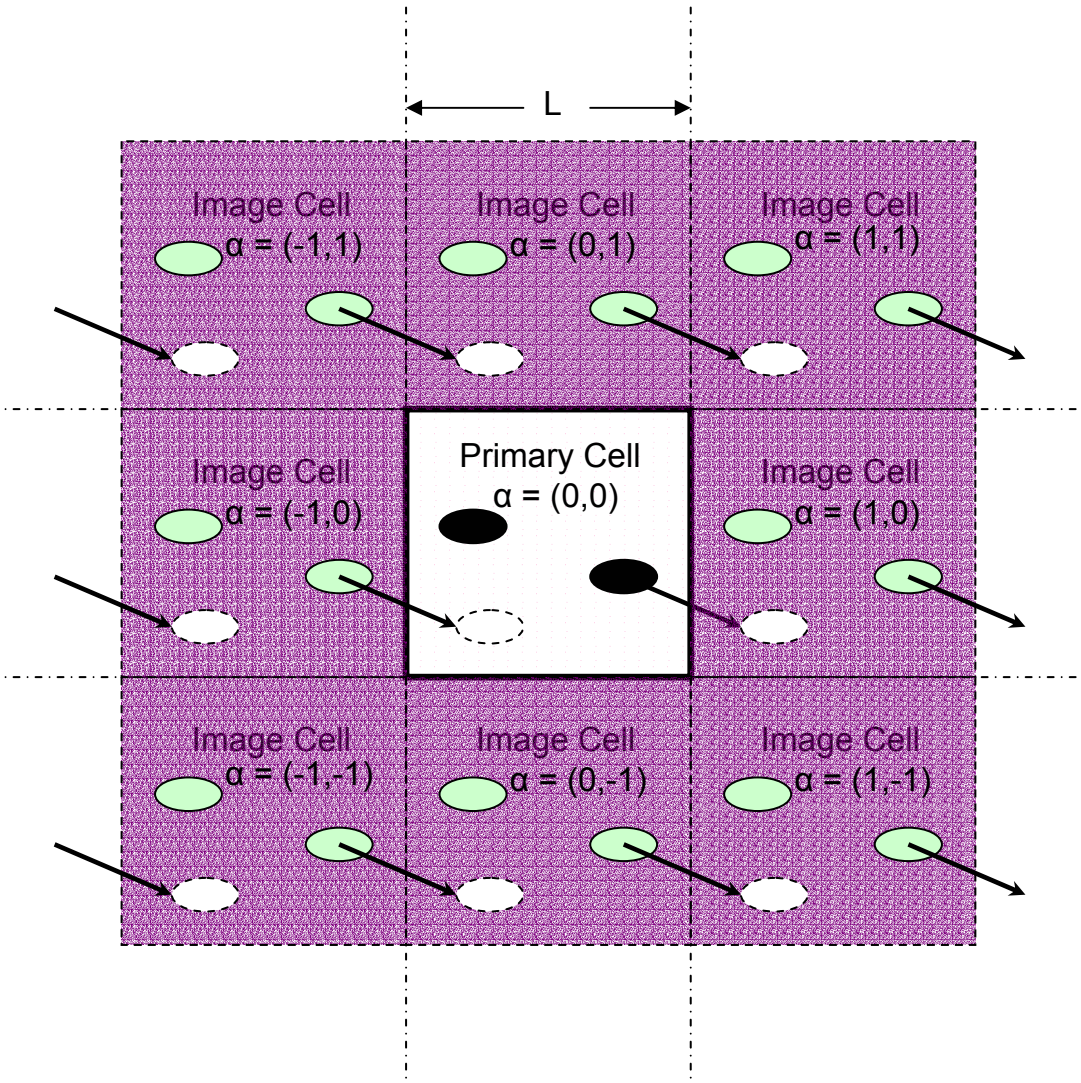

Figure 2-1: Two dimensional periodic boundary conditions applied to the central primary cell.

The description of the system as observed in the macroscopic samples is, however, not possible with the currently envisaged computers so it is realized by the quasi-infinite lattice of the original central cell in all directions, which is considered as a

small volume element representing the sample in the bulk solution as shown in the Figure 2-1. The primary cell is replicated throughout the space to form an infinite lattice. These replicated primary cells are called image cells, which are of the same size and shape as the primary cell. In the course of the simulation, when an atom or molecule moves in the primary cell, its periodic image in every one of the other cells moves with exactly the same orientation in a similar fashion as shown in the Figure 2-1. Thus, as a molecule leaves the primary cell, one of its images will enter from the image cells through the opposite face. There are no walls at the boundary of the central cell, and the system has no surface so the atoms in the primary cell and image cells can freely enter or leave any cell at any time. Nevertheless, the number of atoms (N) in the primary cell should be maintained constant during the simulation run as stated above.

Though the container walls have been removed by the PBC, the position and momentum of the atoms in the image cells are always related to the atoms in the primary cell. Each cell in the lattice is assigned a reference frame which is located at the left corner of the cell. When an atom moves between the cells it is identified with the cell translation vector α, together with the image frame by a vector $R_{\alpha}^{(0,0,0)}$ in three dimensions. Thus for a cube of edge length L, an image atom i in the cells is located with respect to the primary frame vector $r_{i(0,0,0)}^{(0,0,0)}$ as

$$r_{i(\alpha)}^{(0,0,0)} = r_{i(0,0,0)}^{(0,0,0)} + R_{(\alpha)}^{(0,0,0)} , \qquad 2.31$$

$$R_{\alpha}^{(0,0,0)} = \alpha L . \qquad 2.32$$

From this equation we can easily calculate the position of an image atom i with respect to the primary atom i along the primary reference frame vector $r_{i(0,0,0)}^{(0,0,0)}$ and the cell translation vector α.

The potential energy in the primary cell (U_P) can be calculated from the pairwise additive potentials of N atom in the central cell

$$U_p = \frac{1}{2} \sum_{i} \sum_{j>i} U\left(r_{i\,j(0,0,0)}^{(0,0,0)}\right) . \qquad 2.33$$

Beside the atoms in the primary cell, the atoms in the image cells will also contribute to the total potential energy and can be calculated in a similar fashion to

the primary cell but along with the cell translational vector. The total potential energy for a periodic system was then calculated by the pairwise additive function.

$$U_{eff} = \frac{1}{2} \sum_{\alpha} \sum_{i \neq} \sum_{j} U(r_{i\,j(0,0,0)}^{(0,0,0)} - \alpha L) \ . \qquad 2.34$$

2.3.3 Non-bonded cut-offs

A trade-off has been introduced in the form of cut-offs to the potential energy calculations of an individual atom in the cell to speedup the MD computations (Feller et al., 1996; Mark and Nilsson, 2002). The usual non-bonded Lennard-Jones potential has an infinite range. In practical applications, it is customary to establish a non-bond cutoff radius (CUTOFNB) and disregard the interactions between atoms separated by more than the CUTOFNB distance (Figure 2-2). This results in simpler programs and enormous savings of computer resources, because the number of atomic pairs separated by a distance r_{ij} grows as r_{ij}^2 and becomes quickly huge. The atoms beyond the cut-off distance are considered to contribute negligibly small to the potential energy interactions. So the interacting atoms which are higher than the cut-off distance are not included in the potential energy interactions. The difference between the non-bonded list cutoff (CUTNB) (dashed circle) and the CUTOFNB (solid circle) distance is termed neighbour list (Figure 2-2). These neighbour lists are maintained and are updated frequently during the simulations. This ensures a considerable speedup of the computation time.

The present study uses a cutoff distance list of 14 Å and a non-bonded cutoff of 12 Å. All atoms within the CUTNB are kept in the non-bonded list and are checked to see if they are also in the CUTOFNB. If some atoms are not present in the neighbour list then the new list is updated simultaneously and continues in this manner over the dynamic run. The potential energies were forced to zero at the CUTOFNB distance and are not included in the calculations. However the potentials were not abruptly forced to zero at the cutoff distance but are monotonically vanishing to zero by the activated switching functions (Allen and Tildesley, 1987).

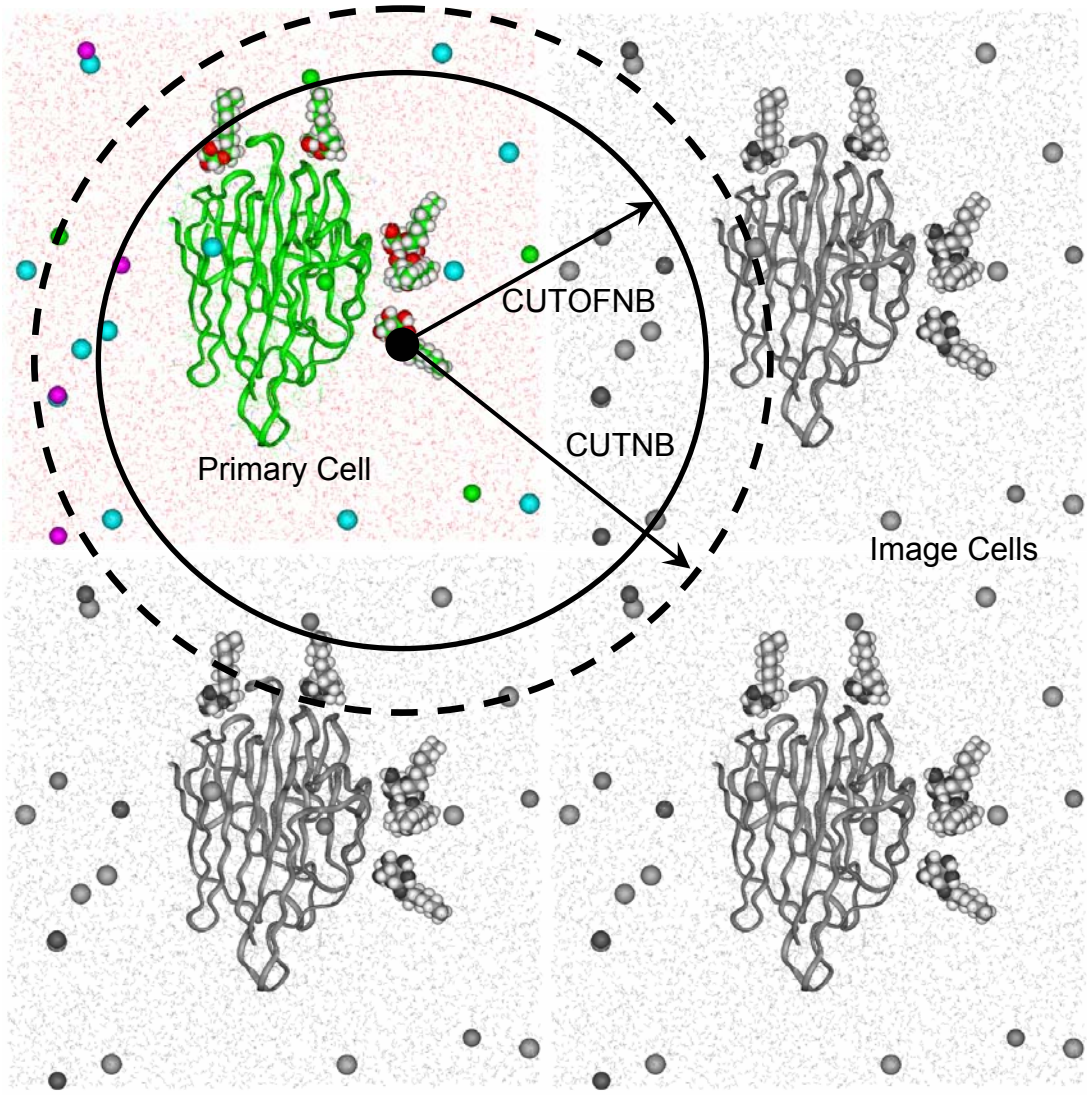

Figure 2-2: Schematic diagram of the non-bonded cut-offs in the MD simulations. The primary cell is shown in color and the image cells were shown in grey color.

The non-bonded interaction applied with the switching function S_W is defined by (Brooks et al., 1983),

$$U_{vdW} = \sum_{i,j=1} \varepsilon \left[\left(\frac{R_{min}}{r_{ij}} \right)^{12} - \left(\frac{R_{min}}{r_{ij}} \right)^{6} \right] S_W(r_{ij}^2, r_{on}^2, r_{off}^2) \ . \qquad 2.35$$

$$U_{Elec} = \sum_{i,j=1} \frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} S_W(r_{ij}^2, r_{on}^2, r_{off}^2) \ . \qquad 2.36$$

$$S_W(r_{ij}^2, r_{on}^2, r_{off}^2) = \begin{cases} 1 & r_{ij} \le r_{on} \\ \dfrac{(r_{off} - r_{ij})^2 (r_{off} + 2r_{ij} - 3r_{on})}{(r_{off} - r_{on})^3} & r_{on} < r_{ij} \le r_{off} \\ 0 & r_{ij} > r_{off} \end{cases} \quad 2.37$$

Here the non-bonded interactions (van der Waals interaction, equation 2.35 and electrostatic interaction, equation 2.36) are treated with the switching function (S_W) to terminate the non-bonded potentials to zero at above CUTNB (r_{off}) distance ($r_{ij} > r_{off}$). The interactions are explicitly accounted with the usual Lennard-Jones and the electrostatic potential functions below the CUTOFNB (r_{on}) distance ($r_{ij} \le r_{on}$) for all atoms within this distance. The distance criterion was applied to shift the potential smoothly to zero between the CUTOFNB and CUTNB distance ($r_{on} < r_{ij} \le r_{off}$).

2.3.4 Water buffer region

Experimental and theoretical evidences state that the water molecules tend to have an uncorrelated density at the surface of the solute (Bruce et al., 2002). The uncorrelated density seems to extend to about three layers in the radial distribution profiles, which is about a radius of 9 Å from the solute surface. So the water layer should be maintained at least this much minimum thickness from the solute atoms in the primary cell in order to avoid any uncorrelated effects. The buffer region can be approximately calculated as two times the CUTOFNB distance between solutes in the primary cell and the image cell as shown in the Figure 2-3. Keeping this in mind glycoconjugate models was developed that have a water layer thickness of above three orders of magnitude as stated above. For the glycolipid simulations the thickness of the water layer was about 12-15 Å around the micelle surface (Konidala et al., 2005) in the primary cell in all the directions. The lectin simulations exploit a much larger water layer thickness of above 20 Å around the pea lectin monomer in all directions. This ensures that the solutes used in the simulations were well solvated with the water molecules for the investigation of their colloidal static and dynamic properties.

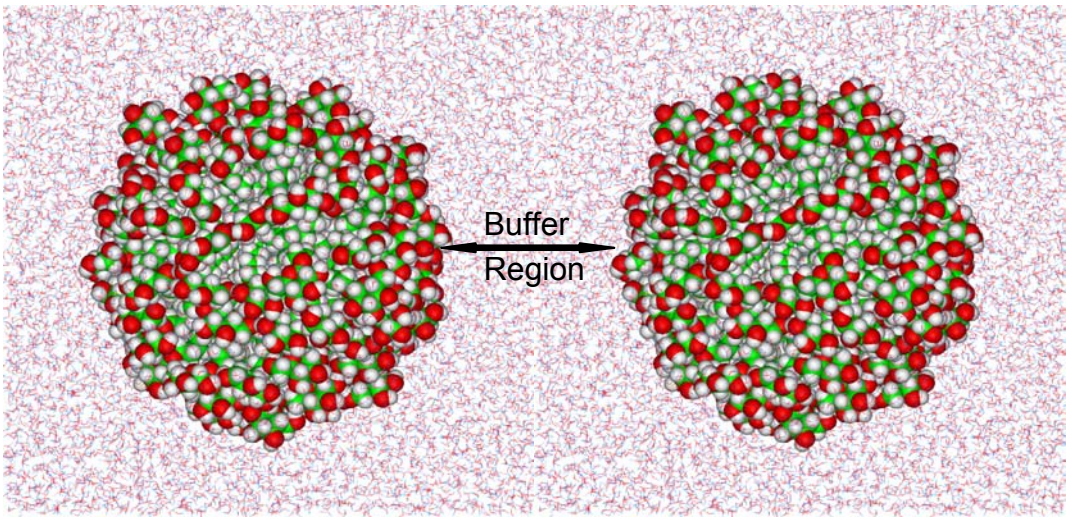

Figure 2-3: Water buffer region between the OG micelles in the primary and image cells. The OG micelle atoms are shown in the center as van der Waals spheres.

2.3.5 Hydrogen bond length constraints

The integration time step is still a tentative process, it should be large enough to sample as much of conformational phase space as possible. At the same time it should be small enough to capture the high frequency motions pertaining to the hydrogen bonds in the model. The motions correspond to the stretching of bonds involving hydrogen atoms, such as C-H, N-H or O-H vibrates with a period of about 10 fs. Consequently, to include these molecular vibrations in the simulations, it is a common practice to go one order of magnitude lower than this limit, thus a time step of at least 1 fs is insisted for the simulations. However, it is often the case that MD simulations of biomolecular systems aim at processes occurring on a much longer timescales, such as folding analysis of proteins, conformational changes in glycoconjugates, or intermolecular interaction studies of globular proteins with the ligand molecules. To address these processes from the present state MD simulations, advancements and if possible some approximations are necessary. In the last few years several numerical integrators have been developed in order to increase the time step for the integration of classical equations of motion, so that both fast and slow motions can be accounted for in the simulations (Tuckerman et al., 1992; Martyna et al., 1996; Cheng and Merz Jr., 1999; Tuckerman, 2000). In any case these developments should not be affected by the efficiency, speed and accuracy of the end results.

Since most phase-space properties which are interesting are expected to be minimally affected by high frequency bond vibrations, it is routinely accepted to constrain some degrees of freedom in the system such that the bonds involving hydrogen atoms are forced to have a constant length. This procedure allows employing a couple of higher time steps. Consequently, longer simulations can be realized for a given computational time. The algorithms most frequently used for performing constrained dynamics are SHAKE (Ryckaert et al., 1977; Tobias and Brooks, 1988; Kraütler et al., 2001), LINCS (Hess et al., 1997) and SETTLE (Miyamoto and Kollman, 1992). These methods differ in their performance and applicability; for example, LINCS is an improved version of SHAKE in that it is compatible with longer time steps, as well as being computationally faster. However LINCS is only applicable to bond lengths and uncoupled bond-angles, whilst the SHAKE method is more general. Finally, SETTLE is an analytical method that constraints the internal degrees of freedom of molecules involving there atoms, such as water. In spite of differences, the basic operation is analogous in all these algorithms. First, inter molecular forces were calculated from the potentials, and perform an unconstrained update of the phase-space coordinates, and finally the constraint algorithm calculates the necessary displacements in the atomic positions so that all constraints should be satisfied simultaneously, within a given tolerance. The resulting phase-space details are then set as initial conditions for the next integration time step to solve the classical Newton's equations in a deterministic approach.

2.3.6 Time averages from the MD simulations

Molecular dynamics simulations are typically applied to systems containing a constant number of atoms (N), volume (V), energy (E), temperature (T) and pressure (P) which describes the thermodynamic state of the system (Allen and Tildesley, 1987). Other thermodynamic parameters like constant density, chemical potential, and heat capacity (C_V) may also be derived through the knowledge of the equations of state and the fundamental equations of thermodynamics. The macroscopic thermodynamic state of a system is defined by a small set of these parameters, usually a combination of three parameters as listed in the Table 2-2. Besides this the microscopic state of the system is defined by the atomic positions, r, and momenta, p. These are considered as coordinates in the multidimensional space called phase space. So for a system of N atoms this space has 6N dimensions (Haile, 1992). The single point in the phase space defines the microscopic state of the system and a collection of points is termed the ensemble

averages of this particular macroscopic state. An MD simulation generates a sequence of points in the phase space as a function of time. These points belong to the same ensemble, but they correspond to the different conformations of the system and their respective momenta.

Table 2-2: Different types of thermodynamic ensembles used in the MD simulations

Ensemble type	Constant parameters
Microcanonical (NVE)	Fixed number of atoms, volume and energy
Canonical (NVT)	Fixed number of atoms, volume and temperature
Isobaric-Isothermal (NPT)	Fixed number of atoms, pressure and temperature
Grand canonical (μVT)	Fixed chemical potential, volume and temperature

The time average properties calculated from the MD simulations are compared to the ensemble averages derived from the experimental sample of an extremely large number of conformations through the statistical averages. The ensemble average for a macroscopic sample is defined as

$$\langle A \rangle_{ens} = \int\int dp^N dr^N \, A(p^N, r^N)\, \rho(p^N, r^N) \quad , \qquad 2.38$$

where $A(p^N, r^N)$ is the property of interest which is a function of momenta and coordinates and $\rho(p^N, r^N)$ is the probability density function. The integrals are over all possible values of momenta and positions. The probability density function with respect to the momentum and coordinates are defined as

$$\rho(p^N, r^N) = \frac{1}{Q} \exp\left[-H(p^N, r^N) \big/ k_B T \right] \quad , \qquad 2.39$$

where Q is the partition function, H is the Hamiltonian operator, k_B is the Boltzmann's constant and T is the temperature of the system

$$Q = \iint dp^N dr^N \exp\left[-H(p^N, r^N) \big/ k_B T \right] . \qquad 2.40$$

In general, the partition function will be applied to calculate the thermodynamic properties like energy (E), entropy (S), density (ρ) and heat capacity (C_v). However, the calculation of these properties is difficult from the partition function because of the double integrals, which accounts for all possible states of the system. So MD simulations were used as an alternative to obtain the time averaged properties from the thermodynamic ensemble.

$$\langle A \rangle_{MD} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \approx \frac{1}{M} \sum_{t=1}^{M} A(p^N, r^N) \ , \qquad 2.41$$

where τ is the simulation time, M is the number of time steps in the simulation and $A(p^N, r^N)$ is the instantaneous value of the property of interest. The calculated time average is then considered to be equal to the experimental ensemble average by the most fundamental assumption, the ergodic hypothesis. This hypothesis states that the time averages calculated from the instantaneous time intervals in MD simulation (equation 2.41) is equal to the macroscopic ensemble average (equation 2.38) of a large number of conformations (Haile, 1992).

2.4 Scene of constant pressure in the MD simulations

Most of the experiments related to the micelle aggregation and/or analysis of proteins were conducted at the normal laboratory conditions of constant pressure and temperature. Consequently the MD simulations performed with the isothermal-isobaric (NPT) ensemble gives a direct comparison to the experimental measured properties (Heermann, 1986). Because of this purpose the constant pressure ensembles are increasingly applied in the recent simulations. Also the application of the NPT ensemble provides improved fluid-solid phase transition properties, which are otherwise, in the case of employing the other ensembles (microcanonical (NVE) and canonical (NVT)), treat as coexistence of the two phases. In order to work with this ensemble the temperature and pressure have to be controlled in the periodic cell during the simulations. This can be handled with the thermostat and barostat control methods.

The temperature is relatively easy to access in a MD simulation. This can be done by the equipartition theorem.

$$T = \left\langle \frac{1}{3Nk_B} \sum_{i=1}^{N} m_i \, v_i^2 \right\rangle , \qquad 2.42$$

where v is the velocity of the atom i, the angular brackets $\langle \cdots \rangle$ denote for the ensemble average over the MD trajectory.

The pressure in the simulation is little more complex and can be measured based on the Clausius virial theorem. The derivation of the theorem is defined by the quantity (W) known as virial for a system of N atoms.

$$W(r) = \sum_{i=1}^{N} m_i \, \vec{r}_i \, \dot{\vec{r}}_i \quad . \qquad 2.43$$

The time derivative of this is

$$\frac{d}{dt} W(r) = \sum_{i=1}^{N} m_i \, \dot{\vec{r}}_i^{\,2} + \sum_{i=1}^{N} m_i \, \vec{r}_i \, \ddot{\vec{r}}_i \quad . \qquad 2.44$$

According to the Newton's law this can be written in term of force acting on each atom i as

$$\frac{d}{dt} W(r) = \sum_{i=1}^{N} m_i \, \dot{\vec{r}}_i^{\,2} + \sum_{i=1}^{N} \vec{r}_i \, \vec{F}_i^{T} \quad , \qquad 2.45$$

where F_i^T is the total force acting on the atom i, which can be divided into two parts, the internal force $\vec{F}_{ij}^{I}$ (due to the interatomic interactions between atom i and j) and external force $\vec{F}_i^{E}$ (force exerted by the simulation cell walls).

$$\vec{F}_i^{T} = \vec{F}_{ij}^{I} + \vec{F}_i^{E} \quad . \qquad 2.46$$

Fitting this to the equation 2.45 gives

$$\frac{d}{dt}W(r) = \sum_{i=1}^{N} m_i \, \dot{\vec{r}}_i^{\,2} + \sum_{i=1,\, j>i}^{N} \vec{r}_{ij} \, \vec{F}_{ij}^{I} + \sum_{i=1}^{N} \vec{r}_i \, \vec{F}_i^{E} \quad . \qquad 2.47$$

The ensemble average of the above equation can be written as

$$\left\langle \frac{d}{dt}W(r) \right\rangle = \left\langle \sum_{i=1}^{N} m_i \, \dot{\vec{r}}_i^{\,2} \right\rangle + \left\langle \sum_{i=1,\, j>i}^{N} \vec{r}_{ij} \, \vec{F}_{ij}^{I} \right\rangle + \left\langle \sum_{i=1}^{N} \vec{r}_i \, \vec{F}_i^{E} \right\rangle \quad . \qquad 2.48$$

The first term on the right hand side is just twice the kinetic energy (K) and the second term is the virial of Clausius (Ξ).

$$\Xi = \sum_{i=1,\, j>1}^{N} \vec{r}_{ij} \, \vec{F}_{ij}^{I} = \left\langle \sum_{i=1,\, j>1}^{N} \vec{r}_{ij} \, \vec{F}_{ij}^{I} \right\rangle \quad , \qquad 2.49$$

$$\left\langle \frac{d}{dt}W(r) \right\rangle = 2K + \Xi + \left\langle \sum_{i=1}^{N} \vec{r}_i \, \vec{F}_i^{E} \right\rangle \quad . \qquad 2.50$$

The last term in the equation 2.47 and 2.48 is the contribution of the external force. For the atoms confined in a cubic box of edge length L and volume V, the external forces are related in a simple way to the pressure exerted by the walls of the cell of surface area A (i.e. $A = L^2$). The external force contribution to the virial in all three dimensions can be evaluated as

$$\left\langle \sum_{i=1}^{N} \vec{r}_i \, \vec{F}_i^{E} \right\rangle = L_x(-P\,L_y\,L_z) + L_y(-P\,L_x\,L_z) + L_z(-P\,L_x\,L_y) \quad . \qquad 2.51$$

For a cubic lattice equation 2.51, simplifies to

$$\left\langle \sum_{i=1}^{N} \vec{r}_i \, \vec{F}_i^{E} \right\rangle = -3\,P\,L\,A = -3\,PV \quad . \qquad 2.52$$

By ergodic hypothesis one may replace the ensemble average of this defined quantity by the time average property, so this can be we written as

$$\left\langle \frac{d}{dt} W(r) \right\rangle = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^{\tau} \left(\frac{dW(r)}{dt} \right) dt$$

$$= \lim_{\tau\to\infty} \frac{1}{\tau} \left[\sum_{i=1}^{N} m_i \, \vec{r}_i \, \dot{\vec{r}}_i \right]_0^{\tau} \qquad 2.53$$

$$= 0$$

The above equality comes from the fact that the sum in the square brackets must always be finite, but the denominator τ (number of time step) becomes infinite. The summing of all the individual terms together results in

$$0 = 2K + \Xi - 3PV \, , \qquad 2.54$$

$$PV = \frac{2K}{3} + \frac{\Xi}{3} \, . \qquad 2.55$$

The above equation is the virial equation and is used normally in the simulations to calculate the pressure inside the system.

$$PV = Nk_B T + \frac{1}{3} \left\langle \sum_{i=1, j>1}^{N} \vec{r}_{ij} \, \vec{F}_{ij}^{I} \right\rangle \, , \qquad 2.56$$

$$P = \frac{Nk_B T}{V} + \frac{\Xi}{3V} \, . \qquad 2.57$$

This can be written in terms of pairwise additive potentials as

$$P = \frac{Nk_B T}{V} - \frac{1}{3V} \left\langle \sum_{i=1}^{N} \sum_{j>i}^{N} \vec{r}_{ij} \left. \frac{dU}{dr} \right|_{r_{ij}} \right\rangle \, . \qquad 2.58$$

2.4.1 Pressure deduced from momentum flux

The derivation of the usual virial theorem is problematic in the constant pressure simulations. This is due to the fact that there are no rigid cell wall boundaries existing in the simulations. As discussed before the cell walls have been removed by the periodic boundary condition, so the atoms can freely move in and out of the

cell at any time (thus the volume of the cell fluctuates). This leads the calculation of the normal virial function with the periodic systems in confusion. An alternative method has been proposed which uses the momentum flux (Haile, 1992) to deduce the virial theorem.

Let us consider a virtual planar surface of area $A = L^2$ inserted perpendicular to the x-axis into the system. The pressure can be defined as the force per unit area acting normal to the surface as

$$P_T = \frac{F_T}{A} = \frac{1}{A}\frac{d(m(v_x + v_y + v_z))}{dt} \quad . \qquad 2.59$$

Thus the pressure can be interpreted as a momentum flux through a unit area of the surface in a unit time. In general, this flux is composed of two parts, the momentum carried by the atoms (P_M) as they cross the surface area during dt, and the momentum transferred (P_F) as a results of intermolecular forces acting between particles that lie on different sides of the virtual planar surface.

$$P_T = P_M + P_F \quad . \qquad 2.60$$

The momentum flux carried by the movement of atoms in the x-direction can be described as

$$\begin{aligned} \langle P_{MX} \rangle &= \left\langle \int m\, v_x\, v_x \frac{N}{V} f(v,t)\, dv_x \right\rangle , \\ &= \frac{Nm}{V}\left\langle \overline{v_x^2} \right\rangle = \frac{2N}{V}\langle E_{KX} \rangle , \end{aligned} \qquad 2.61$$

where E_{KX} is the kinetic energy per atom in the x-direction. Repeating this procedure in y and z directions, we can form the total convective contribution to the pressure and thereby obtain the ideal gas law

$$\langle P_M \rangle = \frac{2N}{3V}\left[\langle E_{KX} \rangle + \langle E_{KY} \rangle + \langle E_{KZ} \rangle\right] = \frac{2N}{3V}\langle E_K \rangle \quad . \qquad 2.62$$

Now the momentum flux caused by the intermolecular forces is considered. Let P_{FX} be the total force per unit area acting normal to the surface A in the x-

direction, where the forces are caused by atoms on one side interacting with atoms on the other side. The forces are pairwise additive and P_{FX} can be written as

$$P_{FX} = \frac{1}{A}\sum_{i}{}' \sum_{j}{}'' F_{ij,X} \quad . \qquad 2.63$$

The first sum runs over all atoms on the left side (single prime) of the surface and the second sum over all atoms which are on the right side (double prime). The primes in the summation indicate that the potential interations are pairwise additive. Now the average over all possible location of the virtual surface is made

$$\overline{P_{FX}} = \frac{1}{AL}\int_0^L \sum_{i}{}' \sum_{j}{}'' F_{ij,X} \quad . \qquad 2.64$$

The integral can be approximated by the sum over all atoms in the following way. Assuming the particles have been labeled sequentially from 1 to N as their x-position increase from zero to L, one can define $X_{k,k+1} = X_{k+1} - X_k$. Note that only the interactions of atoms contribute to $\overline{P_{FX}}$ which are on different sides of the planer surface. So we can write

$$\overline{P_{FX}} = \frac{1}{V}\sum_{k=1}^{N-1}\sum_{i=1}^{k}\sum_{j=k+1}^{N} F_{ij,X} X_{k,k+1} \quad . \qquad 2.65$$

Rearranging the above sum leads to

$$\begin{aligned} \overline{P_{FX}} &= \frac{1}{V}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} F_{ij,X}\sum_{k=i}^{j} X_{k,k+1} \quad , \\ &= \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} F_{ij,X} X_{ij} \quad . \end{aligned} \qquad 2.66$$

Averaging over time gives

$$\overline{P_{FX}} = \frac{1}{V}\left\langle \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} F_{ij,X} X_{ij} \right\rangle \quad , \qquad 2.67$$

$$\langle P_F \rangle = \frac{1}{V}\left\langle \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} F_{ij,X} X_{ij} \right\rangle + \frac{1}{V}\left\langle \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} F_{ij,Y} Y_{ij} \right\rangle + \frac{1}{V}\left\langle \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} F_{ij,Z} Z_{ij} \right\rangle ,$$

$$\langle P_F \rangle = \frac{1}{3V}\left\langle \sum_{i<j}\sum \vec{F}_{ij}\, \vec{r}_{ij} \right\rangle .$$

2.68

This is the virial function for the momentum flux transferred by the intermolecular forces in the system. Combining the equation 2.62 and 2.68 in equation 2.60 gives the complete expression for the pressure

$$P_T = \frac{2N}{3V}\langle E_K \rangle + \frac{1}{3V}\left\langle \sum_{i<j}\sum \vec{F}_{ij}\, \vec{r}_{ij} \right\rangle . \qquad 2.69$$

For system with periodic boundary conditions this becomes

$$P_T = \frac{2N}{3V}\langle E_K \rangle + \frac{1}{3V}\left\langle \sum_{\alpha}\sum_{i<j}\sum (\vec{r}_{ij} - \alpha L)\, \vec{F}_{ij}\, (\vec{r}_{ij} - \alpha L) \right\rangle , \qquad 2.70$$

where α is the cell translational vector and L is the edge length of the periodic cell.

2.4.2 Thermostat and barostat

The microcanonical (NVE) ensemble is native to the molecular dynamics simulation and is still commonly used in practice. As mentioned above, it is more realistic to perform simulations under pressure and temperature control. Several methods have been implemented in CHARMM (Brooks et al., 1983) to control the pressure and temperature in the dynamic simulations. The frequently employed methods are extended system algorithms based on the Lagrangian and Langevin equations and the Berendsen weak coupling method. The latter method rescales the atom velocities by a factor s_T, which is computed from a first order decay of the actual temperature T towards the target value T_O

$$s_T = \left(1 + \frac{\Delta t}{\tau_T}\left(\frac{T_O}{T} - 1\right)\right)^{1/2} . \qquad 2.71$$

The parameter τ_T is the relaxation time constant and should be chosen significantly larger than the integration time step Δt to be in the weak coupling regime. Then the ensemble averages of the Berendsen method are close to the NVT ensemble.

The pressure is controlled using an equivalent form of the above equation, but now scaling has been done to the atom positions and the box sizes. The isotropic scaling of the pressure P to the target P_O is defined as

$$s_P = \left(1 + \frac{\Delta t\, \kappa_p}{\tau_p}\left(P - P_O\right)\right)^{1/3} , \qquad 2.72$$

where τ_P is the pressure coupling constant and κ_P the systems compressibility.

Barostat in the spirit of the Nose-Hoover thermostat have also been proposed. One of such a barostat that generates proper NPT ensemble averages has been proposed by Melchonna et al. 1993. The set of equations of motion for this method is given by

$$\dot{\vec{r}}_i = \frac{\vec{v}_i}{m_i} + \eta_{nh}\left(\vec{r}_i - r_c\right) , \qquad 2.73$$

$$\dot{\vec{v}}_i = \vec{F}_i - (\eta_{nh} + \zeta)\vec{v}_i , \qquad 2.74$$

$$\dot{\zeta} = \frac{1}{\tau_T^2}\left(\frac{T}{T_O} - 1\right) , \qquad 2.75$$

$$\dot{\eta}_{nh} = \frac{1}{N k_B T_O \tau_P^2} V\left(P - P_O\right) , \qquad 2.76$$

$$\dot{V} = 3\eta_{nh} V , \qquad 2.77$$

where r_c is the coordinate of the system center of mass, $\vec{v}_i$ is the velocity of the atom i, η_{nh} is the barostat friction coefficient, ζ is the thermostat friction coefficient, V is the volume of the cell and the dotted variables are the derivatives of these stated variables over time.

2.5 MD setup

Here the procedures one should follow to perform MD simulations are described. For detailed information on the development of the models used in the present simulations see subjected to the next section. The phase-space trajectory which stores information on the positions and velocities of each atom during the MD simulation was obtained from a sequence of steps as shown in the Figure 2-4.

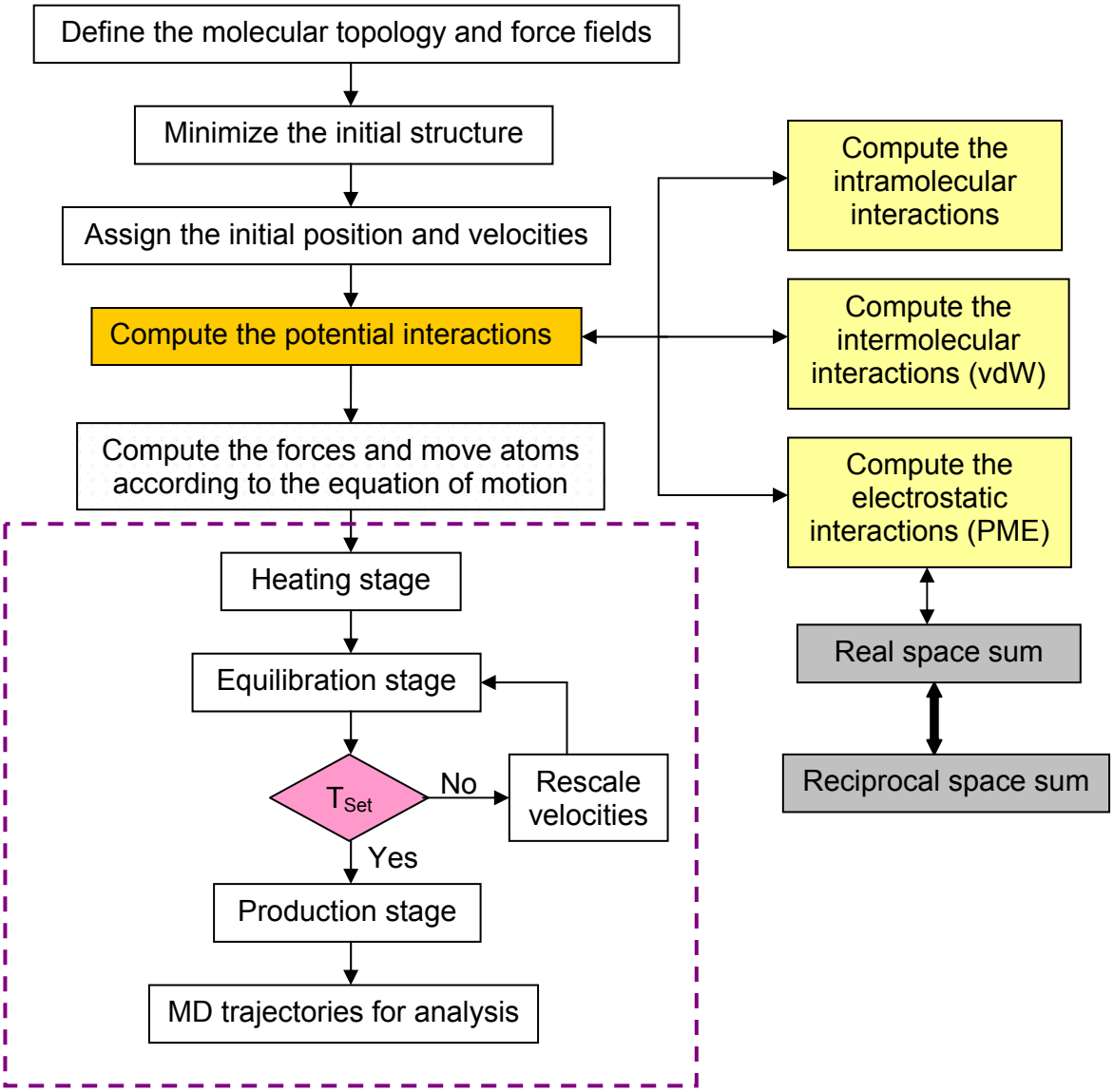

Figure 2-4: General scheme and the CHARMM practical sequential approach of the molecular dynamic simulations.

The initial step for planning the MD simulations is to define the molecular topology and the force field parameters (Appendix A and B). It is emphasized that this is an important step prior to the MD simulations. The starting structure can be developed either from the individual residues of known configurations or can be obtained from the X-ray diffraction crystal structure, NMR, and from recent Electron Microscopy experimental (Grigera, 2002) methods, respectively. Depending on the initial molecular structure, (one should choose configurations that are close to the interested ones to calculate the time average properties) and the quality of the potential energy parameters, the end results will vary, so careful attention at this stage is necessarily required. After defining the initial structure and applying the force field parameters the molecular structures should be energy minimized with the minimization algorithms in order to reduce the extremely high energy such as van der Waals overlaps or any strains encountered during the construction of the models. The simulations might stay in an unstable state or distort the local molecular structure configurations, if such high energy spots are not removed from the initial structure prior to the MD runs. The system in the minimized energy configuration at time zero was obviously static, so initial velocities should be assigned to the molecules or atoms in the simulation cell. The initial velocities are assigned randomly by the Maxwell-Boltzmann velocity distribution functions (Allen and Tildesley, 1987).

The potential energy interaction which is a function of the atomic positions was further initiated from the minimized model coordinates (equation 2.28). The potential interaction energies comprise of two terms: 1) intramolecular interactions and 2) the intermolecular interactions. The intramolecular energies are the bonded energies between atoms which consist of bond stretch, angle, dihedral, improper dihedral energy functions. The intermolecular interactions are the non-bonded energies which are further divided into two important terms, the van der Waals interactions and the electrostatic interactions. The van der Waals interactions can be easily calculated form the Lennard-Jones (12-6) potential model using the non-bonded cutoff scheme. The electrostatic calculations are calculated from the Particle-Mesh Ewald (PME) summation method using Coulomb's law (Essmann et al., 1995). Due to the finite lattice geometry the long range electrostatic interactions with the original Coulomb's law are not converged. For a quasi-infinite lattice periodic systems this laws renders the problem to be intractable. In order to treat electrostatic interactions accurately for large systems in a solvent cell, PME is the only alternative approach to proceed further with these calculations. This method allows one to convert conditionally convergent sum into two rapidly convergent sums. This splits the sum into two portions as

shown in the Figure 2-4. One is evaluated in the real space and the other in reciprocal (Fourier) spaces. Overall the application of PME over the earlier Ewald summation method calculates the long range electrostatic interactions more accurately and in a computationally feasible manner. After the potential energy was determined the forces acting on each atom were calculated with the classical Newton's second law of motion and propagate the system through the numerical finite difference algorithms.

Besides the general scheme of the MD procedures, the practical step-by-step approach frequently followed with CHARMM was also shown in the Figure 2-4, inside the dashed lines. An MD simulation with CHARMM often involves the heating, equilibration and the production stages (Karplus and McCammon, 2002). In the heating stage the temperature of the system is raised from 0 K to the desired temperature in small steps by assigning the velocities. Periodically, new velocities are assigned at a slightly higher temperature and the simulation is allowed to continue until the desired temperature has been reached. After the system has arrived to the desired temperature, the equilibrium phase is monitored to check the structural and dynamical properties, until the values of the thermodynamic parameters were attained stable. The system in the equilibrium stage should be maintained constant over time weighted by the appropriate Boltzmann factor. The third is the production phase which is the actual dynamics stage used to examine the properties of the system. Depending on the systems and the time average properties, respectively, intended to study the MD simulations have to be extended from picoseconds to the time scale of several nanoseconds. The development of the system over time in the phase-space is recorded periodically to the binary trajectory files for analyzing the properties of the biological macromolecules.

Long-range electrostatic interactions

One of the main advantages in using periodic boundary conditions is that it allows treating long-range electrostatic computations more accurately through the Particle-Mesh Ewald (PME) summation method. In general the electrostatic charges are derived from the polarity of the water molecules, ions in the solution, and the charged molecular moieties. The interactions between these charges are important because they are the main long-range forces in the molecular simulations. In order to treat the charges in the system more efficiently, PME algorithm was applied to model the electrostatic particle interactions. As each atom interacts with the other atoms in the system, the time required to perform the long-range interaction goes up to the square of the number of atoms $\Theta(N^2)$, if the

PME method is neglected (Perram et al., 1988; Darden et al., 1993). Specifically, the goal has been to develop a way of modeling electrostatic interaction at an affordable computation time for the biological molecules in a realistic environment. That means the solute molecule should be surrounded by a large slab of water molecules. PME, an improved version over the Ewald method, makes it possible to compute these interactions more quickly and thus permits the solution of large problems that would otherwise has been inaccessible to modern machines on a reasonable time scales (Essmann et al., 1995).

The Ewald summation method allows one to convert the conditionally convergent sum of the electrostatic forces into two rapidly convergent sums. This splits the sum into two portions, one of which is evaluated in real (direct) space and one in reciprocal (Fourier) space (Feller et al., 1996).

$$U_{Elec} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\operatorname{erfc}(\kappa\left|\vec{r}_{ij}+\vec{a}\right|}{\left|\vec{r}_{ij}+\vec{a}\right|} - \frac{\kappa}{2\sqrt{\pi}}\sum_{i=1}^{N}\sum_{j=1}^{N} q_i q_j \delta_{ij} +$$

$$\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \frac{q_i q_j}{\pi L_x L_y L_z} \sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\exp\left(-\frac{\left|\vec{k}\right|^2}{4\kappa^2}\right)\cos(\vec{k}\,\vec{r}_{ij})}{\left|\vec{k}\right|^2} + J(D_{dip}, \varepsilon') \quad , \qquad 2.78$$

where $\vec{k}$ is the reciprocal space lattice vector ($\vec{k} = 2\pi(l/L_x, m/L_y, n/L_z)$ and erfc is the complementary error function. The first and the third terms in the above equation correspond to the real space and reciprocal space summations, respectively. The second term corrects for the self energy of the canceling charge. The last term is a surface correction term that depends on the dipole moment of the unit cell, D_{dip}, and the dielectric constant of the surrounding medium ε'. The Ewald sum works by neutralizing long range forces with the introduction of a Gaussian charge distribution of the opposite sign and arbitrary variance proportional to κ^{-2} around the selected particle. The opposite charge is denoted as the screening distribution. When it is summed in real space along with the original charge distribution, the electrostatic interactions between the charged sites become short ranged. The Fourier reciprocal space term removes the contribution of the

Gaussians, maintains electroneutrality, giving the same value as the original sum. It is summed in the reciprocal space and the total is then transformed back into the real space.

Without the application of the Ewald method, it would be impossible to model the electrostatic interactions of the system because even forces far outside the scope of the model would need to be included. The problem with the Ewald technique is that, to compute the reciprocal space term, one have to compute a Fourier Transform, which is very slow relative to the rest of the program in the order of $\Theta(N^2)$. An operation called Fast Fourier Transform (FFT) exists that can perform the required Fourier Transform much faster in the order of $\Theta(N \log N)$ (Darden et al., 1993; Petersen, 1995). The problem is that a Fast Fourier Transform can only be performed on a function that is defined on a regularly spaced grid. Because the position of the charges in the actual cell is not regular as the atoms are in continuous movement. Thus it was not initially feasible to use the Fast Fourier Transforms. Nevertheless, the introduction of PME solves these issues using an interpolating function to assign pieces (points) the charges to a regularly spaced mesh that could be operated on by a Fast Fourier Transform. The electrostatic energy in PME can be conveniently written as the summation of the energies of the real space, reciprocal space and a correction factor (Essmann et al., 1995).

$$U_{Elec} = U_{real} + U_{reci} + U_{corr} \quad . \qquad 2.79$$

$$U_{real} = \frac{1}{2}\sum_{n}^{*}\sum_{i,j=1}^{N}\frac{q_i q_j \operatorname{erfc}(\beta_c |r_j - r_i + n|)}{|r_j - r_i + n|} \quad , \qquad 2.80$$

$$U_{reci} = \frac{1}{2\pi V}\sum_{m\neq 0}\frac{\exp(-\pi^2 m^2/\beta_c^2)}{m^2} S(m)\, S(-m) \quad , \qquad 2.81$$

$$U_{corr} = -\frac{1}{2}\sum_{(i,j)\in M}\frac{q_i q_j \operatorname{erf}(\beta_c |r_i - r_j|)}{|r_i - r_j|} - \frac{\beta}{\sqrt{\pi}}\sum_{i=1}^{N} q_i^2 \quad , \qquad 2.82$$

where β_c is a constant which control the relative rate of convergence of the real space and the reciprocal space. S(m) is called the structure factor defined by

$$S(m) = \sum q_j \exp(2\pi i\, \vec{m}_r\, r_j) \quad , \qquad 2.83$$

$$= \sum_{j=1}^{N} q_j \exp(2\pi i(\vec{m}_1 s_{1j} + \vec{m}_2 s_{2j} + \vec{m}_3 s_{3j}) \ , \qquad 2.84$$

where $s_{\alpha j}, \alpha = 1,2,3$ in the equation 2.84 are the fractional coordinates of the atom j defined as the product of the conjugate reciprocal vectors a_α^* and position of atom r_j. $\vec{m}_r$ in the above equation is the reciprocal lattice vector.

The asterisk in equation 2.80 denotes for terms with n = 0. The atom pair (i = j or (i, j) $\in$ M) for which non-image non-bond interactions are not calculated, are said to belong to the masked pairlist (M). Since their interaction have been included in the actual Coulombic function so it has to be subtracted from the energy calculated with the real and reciprocal space potentials. The erf is the error function in the correction factor (erf(x) = 1- erfc(x)), V is the volume of the unit cell. The second term in the equation 2.82 is the self interaction energy term. The computation of all these terms introduces additional overhead, but the new operations are in the order of $\Theta(N)$ or less. The overall performance is a significant improvement over the conventional Ewald summation technique. Even though the Lagrangian interpolation method was often employed to approximate the structure factor (equation 2.83), for the simulations an alternative interpolating function was applied: The so called Cardinal B-Splines which is more accurate and differentiable. However the interpolation using the splines is generally more complex than the Lagrangian interpolation. The approximate reciprocal energy ($\tilde{U}_{reci}$) using the B-splines are given by

$$\tilde{U}_{reci} = \frac{1}{2\pi\pi} \sum_{m \neq 0} \frac{\exp(-\pi^2 m^2/\beta_c^2)}{m^2} B(m_1, m_2, m_3)$$

$$\times F(Q)(m_1, m_2, m_3) F(Q)(-m_1, -m_2, -m_3) \ , \qquad 2.85$$

$$B(m_1, m_2, m_3) = |b_1(m_1)|^2 \cdot |b_2(m_2)|^2 \cdot |b_3(m_3)|^2 \ , \qquad 2.86$$

where $F(Q)$ is the discrete Fourier transform and $b_i(m_i)$ is given by

$$b_i(m_i) = \exp(2\pi i(n-1)m_i / K_i \cdot \left[\sum_{k=0}^{n-2} M_n(k+1)\exp(2\pi i m_i k / K_i)\right]^{-1} . \quad 2.87$$

The application of this PME method uses a fixed cutoff in the real space and B-spline interpolation in the reciprocal space structure factors onto a regular grid, permitting the application of Fast Fourier Transforms to calculate the reciprocal space efficiently (Darden et al., 1993; Petersen, 1995). This new methods is substantially more accurate than the original PME and can be improved by adjusting only a few parameters (Essmann et al., 1995). In addition, to facilitate the calculations during the dynamic simulations the PME method should be applied with the neutral charged systems i.e. the overall total charge in the primary cell should be maintained zero.

3 Development of glycoconjugate systems with explicit solvent models

In this chapter attention has been turned to the development of the models for the glycolipid and pea lectin simulations. The model glycoconjugates developed in this study were treated in a fully explicit solvent environment. Since most of the biological processes of the glycoconjugates are happening in the bulk solvent conditions, the presence of water is considered as an essential factor in the simulations which often dictates the activity of these glycoconjugate molecules.

3.1 Solvent models

The water seems to have a very simple structure, but still there is no single water model which can be applied to describe accurately the complete set of physico-chemical properties of a biological macromolecule. In spite of its continuous development over decades there exist several water models as listed in Table 3-1, which have there own advantages and disadvantages (Rahman and Stillinger, 1971; Mark and Nilsson 2002; Nieto-Draghi etal., 2003; Guimaraes et al., 2004).

Table 3-1: Explicit solvent models frequently used in the biomolecular simulations

Water Model	σ (Å)	ε kj/mol	l_1 (Å)	l_2 (Å)	θ^o	ϕ^o	q_1 (e)	q_2 (e)
TIP3P	3.151	0.637	0.9572	-	104.52	-	+0.4170	-0.8340
TIP4P-Ew	3.1644	0.681	0.9572	0.125	104.52	52.26	+0.5242	-1.0484
TIP4P	3.1537	0.648	0.9572	0.15	104.52	52.26	+0.5200	-1.0400
SPC	3.166	0.650	1.0000	-	109.47	-	+0.410	-0.8200
SPC/E	3.166	0.650	1.0000	-	109.47	-	+0.4238	-0.8476
TIP4P/Ice	3.1668	0.882	0.9572	0.158	104.52	52.26	+0.5897	-1.1794
ST2	3.1000	0.317	1.0000	0.80	109.47	109.47	+0.2436	-0.2436
TIP5P	3.1200	0.669	0.9572	0.70	104.52	109.47	+0.2410	-0.2410
TIP5P-Ew	3.097	0.745	0.9572	0.70	104.52	109.47	+0.2410	-0.2410

Careful selection of the water potentials prior to the MD simulations is very important as it reflects the quality of the structural and dynamical properties of the system we are interested in (Bruce et al., 2002; Mark and Nilsson 2002). Further the reality of the biomolecular simulations can only be achieved by the inclusion of the explicit water molecules into the simulation. In Table 3-1 some of the important water models along with their geometric and potential parameters were listed. The column two and three are the Lennard-Jones potentials for the van der Waals distance (σ) of two oxygen atoms from its center and ε the depth of the potential energy minimum. The remaining parameters in the Table 3-1 are the bond length (l_1, l_2), angle (θ^o, ϕ^o) and the charges of hydrogen (q_1) and oxygen (q_2) atoms, respectively.

The potentials of two types of water models (TIP3P (Jorgensen et al., 1983) and TIP4P-Ew (Horn et al., 2004) were applied in the present simulations (see Figure 3-1). The former was used for the investigation of gylcolipid and lectin properties and the latter model was the recently improved version of the standard TIP4P water model which has been used for the preliminary testing of dynamic properties like self-diffusion coefficients of solvent molecules (Wriggers et al., 1998; Bogusz et al., 2001). Investigations of TIP4P-Ew water potentials with different gylcolipid molecules are planned for future work. It should be emphasized here that the calculation of the dynamic properties of the glycolipids in the biochemical applications for studying the solubilization of membrane proteins remains a most promising field to date. As a first step in this regard the diffusion coefficients of these two different water models have been tested within the OG systems performed at constant pressure and temperature simulations.

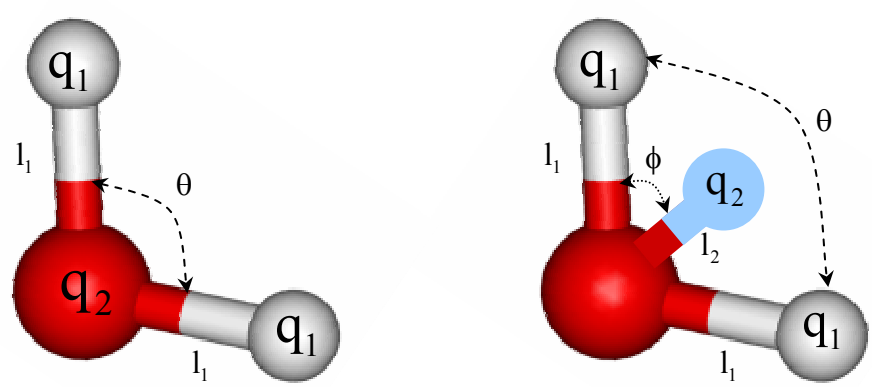

Figure 3-1: Schematic view of the TIP3P and TIP4P-Ew water models.

From the name it implies the TIP4P-Ew (Transferable intermolecular potential 4 point for Ewald method simulations) uses four site interactions in contrary to the TIP3P model. The extra fourth site (q_2) is the fictitious center lying on the bisector of the water H-O-H angle as shown in the Figure 3-1. The new water model has just begun to be employed with glycolipids for the investigation of their dynamic and structural properties because the water potential seems to be well improved compared to the other popular nonpolarizable and polarizable water models (Horn et al., 2004). The structural properties calculated for the OG micelle from these solvent models show very slight differences (Table 3-2) in their aggregate size.

The radius of gyration (R_g) of the OG micelle from the TIP4P-Ew model simulations is more compact ($R_{g\ (TIP4P\text{-}Ew)}$ = 18.4 Å) than calculated by the TIP3P water model ($R_{g\ (TIP3P)}$ = 19.6 Å) (Bogusz et al., 2000; Konidala et al., 2005). Interestingly, the dynamic properties calculated from the two water models show large differences in the diffusion coefficient values. The calculated self diffusion coefficient of 2.4 $\times 10^{-9}\ m^2/s$ from TIP4P-Ew was in good agreement with the experimental and other theoretical works (Essmann et al., 1995; Feller et al., 1996). The experimental self diffusion coefficient for water at 298 K was found to be $2.3 \times 10^{-9}\ m^2/s$ (Mark and Nilsson, 2002).

Table 3-2: MD simulation details of the OG systems with TIP3P and TIP4P-Ew water models

MD Runs	Water Model	System Size (atoms)	Concen-tration (M)	Cell Shape	R_g (Å)	D_{Water} $10^{-9}\ m^2/s$
MD-W1	TIP3P	24,804	0.62	RHDO	19.8	5.1
MD-W2	TIP3P	24,804	0.62	RHDO	19.4	5.1
MD-W3	TIP4P-Ew	43,940	0.45	CUBIC	18.3	2.5
MD-W4	TIP4P-Ew	43,960	0.45	CUBIC	18.4	2.4

As estimated by various authors the TIP3P water model connected with the electrostatic Ewald summation method seems to overshoot the experimental diffusion coefficients by a factor of about 2.2 (Feller et al., 1996; Bogusz et al., 2001; Bruce et al., 2002). The inclusion of an additional interaction site and the re-parameterization of the water potentials improve the estimation of the dynamic

properties of the glycolipids closer to the experiment results. To validate the results of different glycolipid systems with longer MD simulations are necessary however, which are currently prepared to be investigated.

3.2 Solvated OG micelle development

The glycolipid models were developed for the investigation of the structural and dynamic properties of the OG micelle in the bulk solution. The experimental results stated an aggregation number of 90 OG monomers and the micelle structural parameters fitted to the ellipsoidal and cylindrical models were quite reasonable (He et al., 2000). As a consequence the aggregation number of 92 OG monomers was constructed for the MD simulations as detailed in the following section. The presumed size of the micelle is consistent with the published results and the literatures stated therein (Bogusz et al., 2000; He et al., 2000).

3.2.1 Model building procedures for OG systems

The solvated glycolipid model employed in the MD simulations was developed by a couple of systematic procedures. Initially a periodic solvent Rhombic Dodecahedron (RHDO) cell of the lattice parameters *a, b, c* = 70 Å, α, γ = 60° and $\beta = 90^\circ$ was created as shown in the Figure 3-2 (Dixon et al., 2002; Konidala et al., 2005).

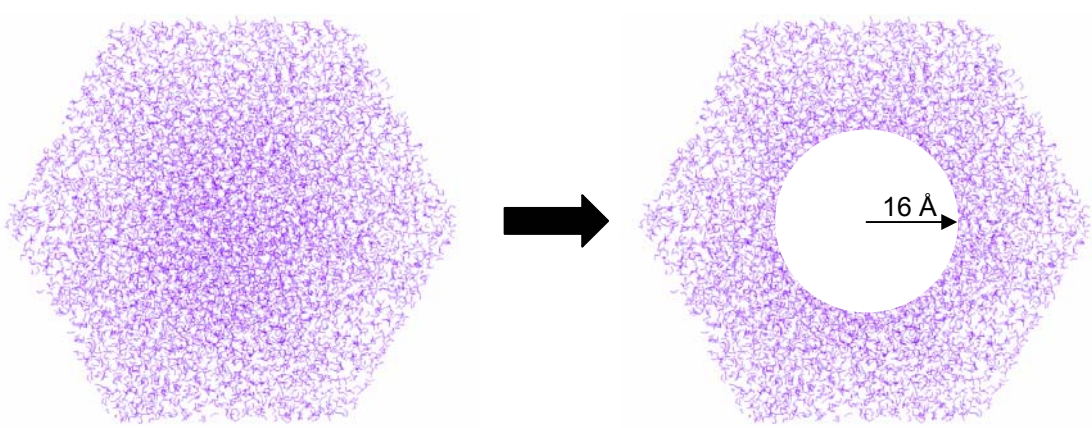

Figure 3-2: Initial rhombic dodecahedron (RHDO) solvent cell.

The TIP3P potential model parameters were applied for the water molecules (Jorgensen, 1983; Durell et al., 1994). The water cell was minimized with 300 steps of Steepest Descent (SD) and 200 steps of Adopted Basis Newton Raphson (ABNR) to remove van der Waals overlaps between water molecules. The resulting coordinates were saved for the heating and equilibration stage followed by a hole of radius 16 Å at the center of the water cell, to place a micelle into the hole. The partially constructed spherical OG micelle as shown in the Figure 3-3 was developed separately with the software program Insight II from Accelry's Inc. (San Diego, California, USA).

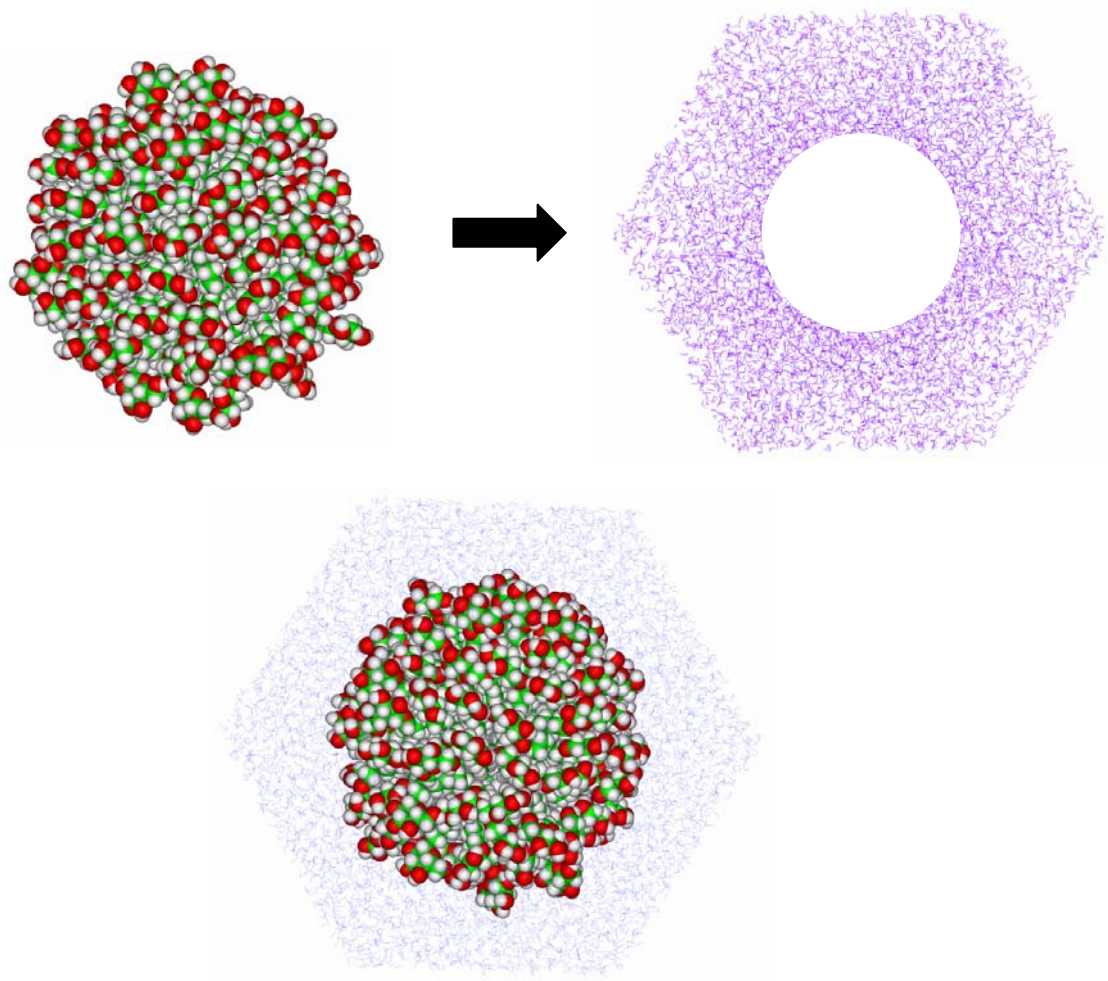

Figure 3-3: Solvated OG micelle at the center of the rhombic dodecahedron (RHDO) solvent cell.

At the beginning one quarter (23 OG monomers) of the micelle was created by placing the glucose head groups extending outwards and the tails facing inwards to the center. Later three copies were taken from the first quarter micelle, systematically rearranged and joined together without any overlaps to obtain a full

spherical micelle of 92 monomers. The partially ordered micelle was then minimized with 1000 steps of SD and 2000 steps of ABNR to relieve from strains during a manual construction. The micelle so created was inserted into the hole of the water cell followed by the constrained and unconstrained minimizations of the OG micelle. The mentioned systematic sequential approach is expected to reduce the equilibration time of the final system to some extent.

3.2.2 OG - Computer experiment in detail

OG parameters were taken from Kuttel et al. 2002, and the references stated therein (Palma et al., 2001; Ha et al., 1988) with the fixed partial charges listed in the Table 3-3 (also refer to Appendix B). The MD simulations were performed with the CHARMM dynamic tool (Brooks et al., 1983). The final system contains 92 OG monomers and 6,796 water molecules, comprising a total of 24,804 atoms present in the RHDO cell. Periodic boundary conditions were applied to the central cell to mimic the influence of the bulk solvent.

Table 3-3: Partial charges of an *n*-octyl β-D-glucopyranoside molecule used in the MD simulations (Kuttel et al. 2002; Palma et al., 2001; Ha et al., 1988)

Atom types	Partial charges
O1	- 0.40
C1	0.200
O5	- 0.40
C2, C3, C4	0.140
O2, O3, O4, O6	- 0.66
HO2, HO3, HO4, HO6	0.430
C6	0.050
C8 – C13	- 0.18
C14	- 0.27
All other hydrogen's	0.090

The van der Waals non-bonded interactions were terminated at 14 Å with a smooth switching function turned on at a distance of 10 Å and a distance dependent dielectric constant. Non-bonded lists were updated automatically when

an atom moves more than 1 Å from the current position. The Particle-Mesh Ewald method was used for the calculation of electrostatic interactions with 64 grip points for the charged mesh and a sixth order B-spline interpolation (Essmann et al., 1995). The width of the Gaussian distribution kappa, κ = 0.34 $Å^{-1}$ has been used with the real space cutoff of 14 Å. The bond lengths of all hydrogen atoms were constrained with the SHAKE algorithm (Ryckaert et al., 1977), thus a higher time step of 2 fs was used for all our simulations. Finally, the potential energy of the whole system was again minimized with the 500 steps of SD and ABNR. The resulting coordinates of the system were used for the 40 ps of heating stage, where the system has been heated up from 0 K to 298.15 K using NPT ensemble. The reference pressure of 1 atm, a pressure piston of 3000 atomic mass units and a collision frequency of 25 ps^{-1} were maintained in the system (Dixon et al 2002). After the system has brought to the desired temperature by coupling to the Hoover thermostat (Hoover, 1985) maintained at 298.15 K, several 100 ps, 200 ps, 1 ns and 2 ns simulations have been performed until 11 ns time scale and the trajectories obtained from these simulations were used for analyzing the structural properties of the OG micelle.

The computations of the solvated OG were performed on a 64 bit HP Itanium2 SuperDome and SGI Octane2 processors. The benchmark testing from our OG model shows the HP Itanium2 processor required only one third of the computation time of that of the Octane2 processor. Because of its excellent performance, we gathered dynamic trajectories for the two large systems of the 11 ns and 6 ns simulation time scales. The calculations took several weeks of computation time with these processors, mainly due to the very large size of the system and the treatment of non-bonded electrostatic calculations in the simulations (Essmann et al., 1995; Feller et al., 1996).

3.3 Development of solubilized pea lectin with OG systems

Several pea lectin models were constructed to investigate the structural properties of the lectin, OG monomer interaction with the amino acid residues of the pea lectin, and dynamics of the ligand monomers in the bulk solution. The OG concentrations in the simulations were varied in addition to the Ca^{2+} and Mn^{2+} metal ions, which are necessary for the binding mechanisms of the lectin-carbohydrate complex (Einspahr et al., 1986; Pletnev et al., 1997). The simulations studied in this work are expected to give better statistics towards the properties of the pea lectin structure and interactions with OG in the solvent phase.

3.3.1 Lectin structure preparation

The crystal structure of carbohydrate binding pea lectin (PDB code: 2BQP) from the X-ray diffraction method was used as an initial structure for the MD simulations with explicit water (Pletnev et al., 1997). The actual lectin is a dimer of two similar monomeric subunits with the same residue sequence and tertiary structure. Simulation of the original lectin dimer with the infinite lattice periodic boundary setup requires a large amount of solvent molecules to treat the protein in an appropriate manner, which leads the computational time to an exponential growth. Because of this and also additionally the two subunits show striking resemblance in their properties, only one subunit was considered for the all-atom MD simulations (Meehan et al., 1982; Einspahr et al., 1986). The PDB structure has a disordered loop of six missing resides (ASN-182:GLU-187) in the subunit chain. This gap was covered by the offset command within the CHARMM program (Brooks et al., 1983). Neutral Histidine residues in the crystal structure of Pea lectin were protonated at the NE2 atom for the MD calculations. The hydrogen atoms were initialized and rebuild with the CHARMM hbuild (hydrogen building) utility. The amino acid sequence of 228 residues with the two chains (β-chain:1-181 and α-chain: 182-228) were present in the final structure (see Figure 3-4a) and the coordinates of the pea lectin molecule were converted and saved as CHARMM Cartesian coordinate format.

Prior to the treatment of the structure with the explicit solvent it was minimized in the vacuum state to reduce the high potential energy barriers inside the structure. Harmonic constraint with a high force constant of 20 kcal/mol had been applied initially to hold atoms together tightly and then minimized the structure with the 100 steps of SD minimization method. The minimized coordinates were compared to the comparison sets and the structure was oriented based on root mean square difference for further relaxations. Subsequently the force constant of 10 kcal/mol, 5 kcal/mol and at the end no harmonic constraints were employed with the minimization of couple of 50 steps of ABNR, and 200 steps of SD methods, respectively before the lectin was completely removed from the constraint (freed lectin) (Dixon et al., 2002).

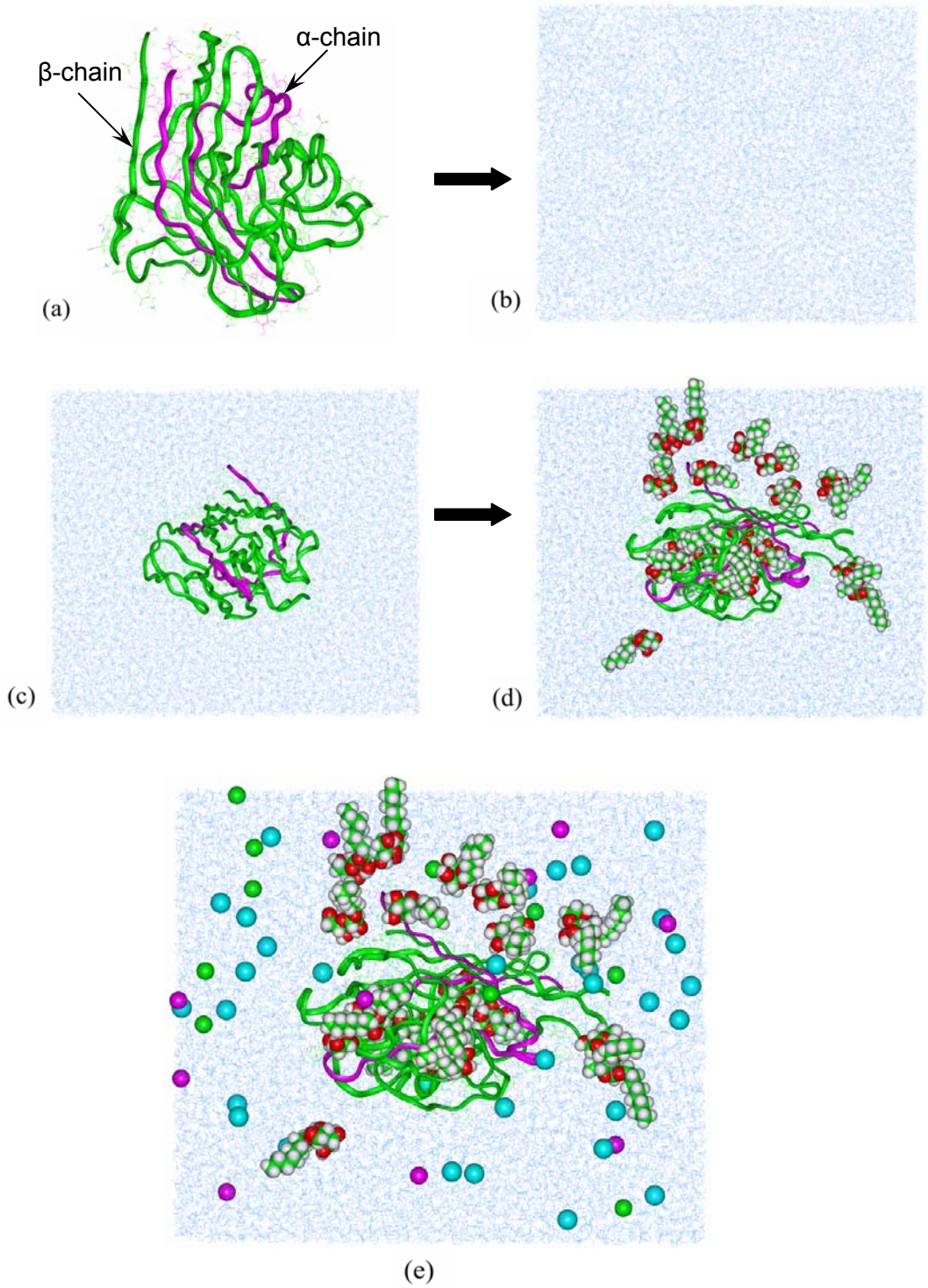

Figure 3-4: Sequence of steps followed in the model development for the pea lectin simulations. (a): pea lectin monomer, (b): explicit tetragonal water cell, (c): solubilized pea lectin in water cell, (d): distributed OG around lectin monomer, (e): complete system with OG and charged ions in the tetragonal water cell.

3.3.2 Solvated pea lectin and glycolipid molecules

The tetragonal (TETR) periodic cell of 90 × 90 × 75 Å was constructed for the water molecules with the TIP3P model potentials (Jorgensen et al., 1983) as shown in the Figure 3-4b. The initial box contains 19,998 water residues and was subjected to 200 steps of SD and 400 steps of ABNR minimizations to relieve bad contact between the residues. The relaxation of the bulk solvent box was performed in the NPT ensemble along with other non-bonded options (see section 3.3.3) for 25 ps of heating and equilibration stage each. After the equilibration of the solvent box a 14 Å hole has been made at the center of the cell to place the lectin monomer inside. The water molecules around 2.0 Å from the lectin were removed after placing the lectin monomer at the center (Figure 3-4c).

The OG lipid molecules were generated separately in this work. Langevin dynamics simulation of the OG monomer in the vacuum state was performed for the OG monomer. The simulation had been carried out for 1 ns and the random monomer conformations were stored for every 1000 ps. Different conformations of the OG were selected randomly from the trajectories produced and distributed around the pea lectin monomer. Keeping the lectin monomer at the center, different number of glycolipid molecules were added to simulations as listed in the Table 1. The OG monomers were first translated from the center and then rotated around the lectin, facing lipid head group near to the lectin surface shown in Figure 3-4d. Solvent molecules that are lower than 1.0 Å distance to the OG monomers were deleted. The OG monomers and the lectin were initially constrained at their respective positions. The water molecules were allowed to move around the constrained segments with the 200 steps of SD and then these segments were unconstrained followed by 300, and 200 steps of SD and ABNR minimizations, respectively.

Finally, depending on the simulations performed several divalent Ca^{2+} and Mn^{2+} ions were added to the system (see Table 3-4) for the biological activity of the lectin macromolecules as determined from the experiments (Einspahr et al., 1986; Pletnev et al., 1997). The water residues were randomly replaced by the ions and the ions were distributed in such a way that they are no closer than 3.0 Å to any of the molecular segments used in the model (Figure 3-4e). The lectin molecule was again constrained for 200 steps of SD and then the whole system was freed from the constraints and the coordinates of the molecular entities were energy minimized with 150 and 200 steps of SD and ABNR minimization methods.

Table 3-4: Description of the ten different MD simulations performed with pea lectin, OG monomers and the charged ions

Simulation ID	System size (No. of atoms)	Molecular segments	OG conc. (M)	Simulation model details
MD1	60,231	20 bog, 10 ca, 10 mn, 33 cl	0.059	20 OG monomers
MD2	59,961	10 bog, 10 ca, 10 mn, 33 cl	0.029	MD2 and MD3: 10 OG monomers, different initial conditions
MD3	59,886	10 bog, 10 ca, 10 mn, 33 cl	0.029	
MD4	60,030	10 bog, 10 ca, 13 cl	0.029	No manganese ions
MD5	59,865	10 bog, 10 mn, 13 cl	0.029	No calcium ions
MD6	59,965	10 bog	0.029	Only OG monomers, ions were removed
MD7	59,805	5 bog, 5 ca, 5 mn, 13 cl	0.015	MD7 and MD8: 5 OG monomers, different initial conditions
MD8	59,790	5 bog, 5 ca, 5 mn, 13 cl	0.015	
MD9	59,748	2 bog, 2 ca, 2 mn, 1 cl	0.0059	MD9 and MD10: 2 OG monomers, different initial conditions
MD10	59,700	2 bog, 2 ca, 2 mn, 1 cl	0.0059	

3.3.3 Details of MD computation

The MD simulations and the analysis of the dynamic trajectories for the lectin systems were performed with the CHARMM version c30b1 (Brooks et al., 1983). The standard CHARMM potential energy surface function and the parameters sets (MacKerell et al., 1998) were applied in the computations along with the other compatible force field parameters for the ligand molecules (Ha et al., 1988; Palma et al., 2001; Kuttel et al., 2002; Li and Lazaridis, 2005). The complete model contains seven segments (two water segments, one pea lectin, OG, calcium, manganese and chloride segment each). The number of OG monomers was varied from 20, 10, 5 and 2 monomers, respectively in addition to the divalent ions as shown in the Table 3-4. On average, the total number of atoms in each simulation was around ~60,000. 3,499 atoms for the Pea lectin were common in all the simulations. An equivalent number of anionic Cl^- ions (see Table 3-4) have also been added to the TETR cell to keep the overall electrical charge of the cell neutral.

The hydrogen atom bond lengths in the simulations were constrained with the SHAKE algorithm (Ryckaert et al., 1977). The long-range non-bonded Lennard-Jones interaction between atom pairs was truncated at 14 Å. The cutoff scheme was handled by a force switch option from 10 Å to 14 Å. The non-bonded interaction pair list was updated automatically when an atom in the current position moved by more than 1 Å. The electrostatic interactions were evaluated from a smooth Particle-Mesh Ewald implementation employing a grid size of $96 \times 96 \times 80$ points in all three directions (Essmann et al., 1995). The width of the Gaussian distribution function, $k = 0.34$ Å^{-1}, was used with the real space cutoff of 14 Å. With all these non-bonded options and image specifications the system was again minimized for 500 steps of SD and ABNR methods prior to the heating step. The integration time step of 2 fs was set in the simulations. The atoms in the system were assigned an initial velocity according to the Maxwellian distribution from 0 K to 289.15 K over a 40 ps heating stage in the NPT ensemble (Karplus and McCammon, 2002). During the equilibration stage the pressure inside the system was maintained at 1 atm using a pressure piston mass of 3000 amu and a Langevin piston collision frequency of 25 ps^{-1}. A Hoover thermostat was applied with a mass of the thermal piston of 20000 kcal ps^2 to keep the temperature of the system at 298.15 K (Hoover, 1985). Different initial coordinates and the velocities were assigned to the simulations prior to the heating and equilibration which gives an excellent scope for the characterization of the structural properties of the molecular components in the simulations in an explicit solvent environment. All the simulations carried out from different coordinates were extended till one nanosecond time scale and the properties from these systems are much more effective than one single long simulation (Karplus and McCammon, 2002).

3.3.4 Force field parameters and trajectory analysis

The energy minimized crystal structure of pea lectin has been applied with the CHARMM22 force field parameters (MacKerell et al., 1998). The partial charges of the OG were taken from Carbohydrate Solution Force Field (CSFF) and appended to the lectin topology and parameter file (Kuttel et al., 2002; Bogusz et al., 2000). The manganese and chloride ions parameters used in the simulations are from the optimized free energy simulations (Beglov and Roux. 1994). The charges and the van der Waals parameters employed for the divalent cations and monovalent ions are listed in the Table 3-5

Table 3-5: Partial charges and the van der Waals parameters for the divalent and monovalent ions

Ligand	Partial charge	ε (kcal/mol)	$R_{min}/2$ (Å)
Ca^{2+}	+2	-0.12	1.71
Mn^{2+}	+2	0.015	1.185
Cl^{-}	-1	-0.15	2.27

The structural stability of the simulations was evaluated by the root mean square (RMS) deviations of the solvated lectin over the initial crystal structure with respect to the simulation time (Bryce et al., 2001; Gerini et al., 2003). The RMS positional fluctuations are calculated for the lectin backbone and side chain atoms per residue basis with reference to the average solution structure. The lectin backbone RMS fluctuations are compared to the X-ray crystal structure through the B- or temperature factor with a simple correlation ($B = (8/3\ \pi^2 \langle \text{RMS-fluctuations}^2 \rangle)$, which reflects the underlying fluctuation of atoms about their average position and measures the important dynamics involved in the amino acid residues of the proteins (Wriggers et al., 1998; Guimaraes et al., 2004; Yuan et al., 2005). Furthermore the solvent accessible surface areas, hydrogen bonding analysis and diffusion coefficients for all molecular components in the simulations were analyzed explicitly and also compared to the available experimental and theoretical investigations reported in the literature (Sen and Nilsson. 1999; Bryce et al., 2001; Reyes et al., 2001). The computations for the lectin systems were performed on the HP parallel computer with 64 bit Itanium2 processors and took several months of computation time due to the very large size of the model and the higher cut-offs of the non-bonded interactions in an explicit solvent periodic cell. The dynamic trajectories were analyzed on the HP Itanium2 and partly on the SGI Octane2 machines.

4 Structural and dynamical property analysis

The phase-space trajectories from the simulations discussed in the preceding sections have been used for the analysis. It is the prime goal to investigate the inherent properties of the glycoconjugate systems in the explicit solvent conditions. This chapter focuses attention on the fundamentals for the calculation of structural and dynamical properties of the OG and pea lectin systems. The structure and dynamics of the OG micelle were characterized by the OG monomer geometric properties, aggregate size and shape, accessible surface areas of the micelle head and tails, radial distribution functions, and diffusion coefficient parameters (Feller et al., 1996; Israelachvili, 1998; Tieleman et al., 2000; Moura and Freitas, 2004). Whereas the pea lectin systems was analyzed by the root mean square deviations, temperature or B-factor dynamics, hydrogen bonding patterns, solvent accessible surface of the lectin, and diffusion coefficients of solvent, OG monomers and the charged ions in the simulations (Wriggers et al., 1998; Kumar and Nussinov, 2002; Mark and Nilsson, 2002; Gerini et al., 2003; Yuan et al., 2005).

4.1 Glycolipid properties

The structural and dynamical properties outlined in the following sections are calculated from the dynamic trajectories of the OG micelle simulations. The calculated properties are averaged over the number of frames in the trajectories. The simulations were performed for a longer nanosecond (11 and 6 ns) time scales and the calculated averages were in a thermodynamic equilibration state verified by the time evolution of the potential energy fluctuations (Kuhn et al., 2002).

4.1.1 Geometric packing

The critical packing parameter (CP_f) is a good indicator for predicting the molecular packing of the lipid monomers in the aqueous solution (Goyal and Aswal, 2001). This gives roughly the shape of the micelle from the ratio of hydrophobic and hydrophilic components in the monomer. The packing parameter can be easily calculated from the geometric properties of the individual OG monomer as

$$CP_f = \frac{v_c}{a_h\, l_c} = \begin{cases} < 1/3 & = \text{spherical} \\ 1/3 < CP_f < 1/2 & = \text{non-spherical} \\ 1/2 < CP_f < 1 & = \text{vesicles or bilayers} \\ CP_f > 1 & = \text{inverted structures} \end{cases} \qquad 4.1$$

where v_c is the volume of the hydrocarbon chain, a_h is the area of the glucose head group, and l_c is the length of hydrocarbon chain.

Molecular geometry plays an important role in the formation of micelles and it is essential to understand how these lipid monomers pack into different structures (He et al., 2000 and 2002). The molecular structures frequently encountered with the lipids upon dissolving in the solution are spherical micelles, ellipsoids, small cylinders, vesicles, bilayers, or inverted micelles. The two main driving forces which control the self association process are hydrocarbon tails in water that favors micelle aggregation and charged or polar head groups that interact with the polar aqueous environment. However, the geometric shape of the aggregate is also dependent on the other environment conditions like temperature, concentration, and pH in the solution (Goyal and Aswal, 2001). From the dynamic trajectories the average OG monomer volume and length, lipid head group areas and volume, and hydrocarbon tail length and volume were calculated. Through these physical dimensions of the monomer the critical packing parameter was estimated for the OG micelle in the aqueous solution. Previous results from the Small Angle Neutron Scattering (SANS) experiments were also compared with the MD results and are discussed in detail in the next chapter (He et al., 2000; Konidala et al., 2005).

4.1.2 Aggregate size and shape

The OG micelle in the solution shows a dynamic performance in nature. So the estimation of the aggregate size is an important variable to monitor during the simulations. This has also been observed to judge the equilibration of the OG lipid systems (Kuhn et al., 2002). An aggregate size of 92 OG monomers consistent with the previously published literature data (He et al., 2000; Bogusz et al., 2000) was employed in the simulations. The radius of gyration (R_g), a measure of the aggregate size was then calculated from the trajectories through the micelle center of mass (Nelson et al., 1997; Bogusz et al., 2000; Moura and Freitas, 2004)

$$R_g = \sqrt{\frac{1}{N_i}\sum_i (r_i - r_{cm})^2} \quad , \qquad 4.2$$

where r_i is the distance of the atom i from the origin, r_{cm} is the micelle center of mass and N_i is the total number of atoms. The center of mass can be calculated as

$$r_{cm} = \frac{1}{N}\frac{\sum_{i=1}^{N} m_i r_i}{\sum_{i=1}^{N} m_i} = \frac{1}{N}\sum \frac{m_1 r_1 + m_2 r_2 + m_3 r_3 + ...}{(m_1 + m_2 + m_3 + ...)} \quad , \qquad 4.3$$

where m_i is the mass of the atom i and $m_i r_i$ is the first moment of the individual atom masses.

The anisotropic shape transformation of a micelle over time was quantified with the moment of inertia tensors (Gao and Wong, 2001; Kuhn et al., 2002). From the inertia tensor, three principle moments of inertia were diagonalised and their ratios are analyzed to observe the shape fluctuations during the simulation. The ratios of these quantities will provide the accurate shape transformations which occurred in the simulations.

The form of the inertia tensor in the volume integral with respect to the Cartesian coordinates can be written as

$$I = \int_V \rho(x,y,z) \begin{bmatrix} y^2 + z^2 & -xy & -xz \\ -xy & z^2 + x^2 & -yz \\ -xz & -yz & x^2 + y^2 \end{bmatrix} dx\,dy\,dz \ . \qquad 4.4$$

So the principle moments of inertia for an ellipsoid of point masses assuming constant density are

$$I_{11} = \sum_{i=1}^{N} m_i (x_i^{\,2} + y_i^{\,2}) \ , \qquad 4.5$$

$$I_{22} = \sum_{i=1}^{N} m_i (x_i^{\,2} + z_i^{\,2}) \ , \qquad 4.6$$

$$I_{33} = \sum_{i=1}^{N} m_i (y_i^{\,2} + z_i^{\,2}) \ , \qquad 4.7$$

where I_{11}, I_{22}, I_{33} are the three diagonalised major, intermediate and minor principle moments of inertias, $I_{11} \geq I_{22} \geq I_{33}$. For simplification the ratios are written further in the text as I_1, I_2, I_3.

The eccentricity e_s of an OG micelle was calculated to observe how the micelle has been elongated over time in the aqueous solution (Bruce et al., 2002).

$$e_s = \sqrt{1 - \frac{I_{min}^2}{I_{max}^2}} \ , \qquad 4.8$$

where I_{min} and I_{max} are the ratios of the minimum and maximum principle moments of inertias.

4.1.3 Solvent accessible regions on the solute molecule

The efforts in regard to the exploration of the three dimensional structure of the lipid aggregates or globular proteins have received increasing attraction over the last few years. The biological macromolecules (lipid membranes, Proteins, DNA, and RNA) show various evolutionary functions or interactions at different regions of their surface with the surrounding solvent molecules (Kumar and Nussinov, 2002). Reports were also shown on the stability and solubility of these macromolecules with respect to the interaction of solvent molecules at their surface (Figure 4-1). The estimation of the aggregate surface in the case of OG micelle and for the pea lectin monomer indicates the roughness of the surface. The high accessible surface areas of the aggregates give clues to the existing surface valleys on the aggregate surface and the ease of interaction with the solvent molecules (Sen and Nilsson, 1999).

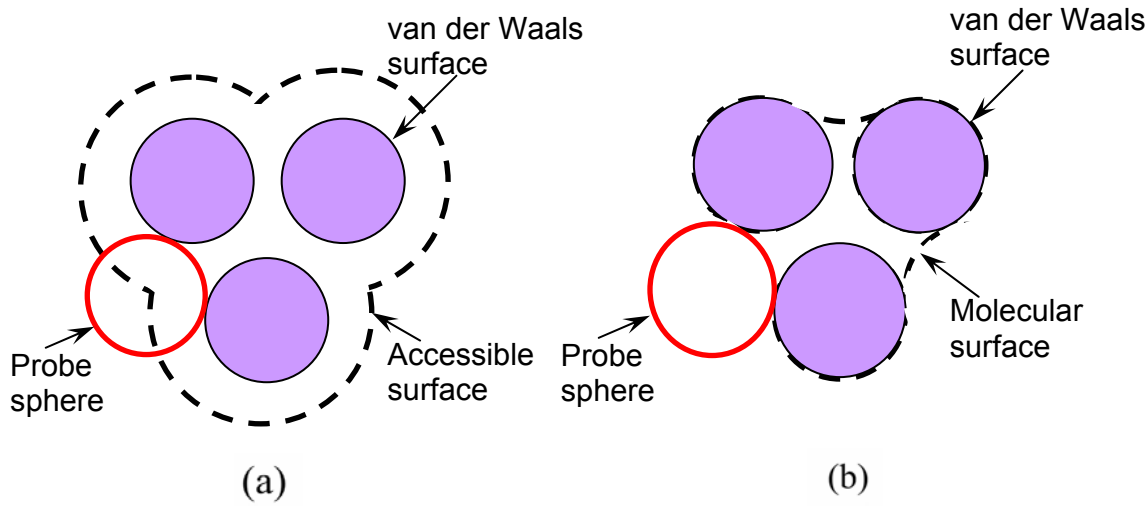

Figure 4-1: (a) Accessible surface of a molecule, defined as the locus of the centre of a solvent molecule as it rolls over the van der Waals surface. (b) Molecular surface of a molecule, defined as the locus of the inward-facing probe sphere to calculate the reentrant surface.

The concept of solvent accessible surface area (ASA) was first introduced by Lee and Richards, 1971 to study the easy water accessible regions on the globular protein residue surfaces. It is defined as the locus of the center of probe sphere (usually the radius of the solvent molecule) when it rolls on the van der Waals surface of the molecule, without penetrating any other atoms of the molecule (Figure 4-1). The radius R of the surface is given by the sum of atomic radii (r_i^{asa}) of all surface atoms and the probe radius (r_p^{asa}) (Hayryan et al., 2005)

$$ASA = \sum_{i=1}^{N} \sum_{s=1}^{m} \frac{R}{\sqrt{R_i^2 - Z_s^2}} D\, L_s \ , \qquad 4.9$$

$$D = \frac{\Delta Z}{2} + \Delta' Z \ , \qquad 4.10$$

where L_s is the length of the arc drawn on a given section s, Z_s is the perpendicular distance from the center of the sphere to the section s, ΔZ is the spacing between the sections and $\Delta' Z$ is $\Delta Z / 2$ or $R - Z_s$, whichever is smaller.

This was later refined by Richards 1977, with the introduction of a new term called reentrant surface, which is nothing but the inward facing part of the probe

sphere in contact with more than one surface atom. The molecular surface is just the sum of the contact surfaces (van der Waals surface that can be touched by a water-sized probe sphere) and the reentrant surface. A probe radius of 1.4 Å was used for the ASA calculations in the micelle simulations and 1.6 Å for the lectin/OG simulations. To compare directly with the experimental results a little higher probe radius was chosen in the pea lectin simulations (Pletnev et al., 1997).

4.1.4 Radial distribution functions (RDF)

The local structure distribution of water molecules around the micelle atoms were analyzed by the radial distribution functions constructed between the oxygen atoms of the OG head and the oxygen atoms of water (Matteoli and Mansoori, 1995; Wymore and Wong, 1999; Gao and Wong, 2001)

$$g(r) = \frac{< N(r, dr) >}{\rho_b \, 4\pi r^2 \, dr} , \qquad 4.11$$

$< N(r, dr) >$ is the average number of atoms in the shell at a distance $r + dr$, ρ_b is the bulk density of water and $4\pi r^2 dr$ is the volume of the shell at $r + dr$.

The RDF is also known as the pair correlation function which is a central quantity in the study of fluids because it determines the average density fluctuations around a reference atom as shown in Figure 4-2. The ordered atoms are shown as white circles immediately around the reference atom (yellow strip circle) up to the distance r. The number density is calculated precisely by counting the number of atoms in each bin (thick black rings) between r and dr and the volume of the bin. At regular intervals the number of atoms in each radial bin is calculated and averaged over the number of trajectory frames in the simulation. The water distribution around the lipid head group and globular proteins is highly affected by the surface charge of these molecules, which can be seen in the Figure 4-2, with a large peak from the reference atom (Bruce et al., 2002). The water molecules are reorganized next to the reference atom with a disruption in the bulk water hydrogen bonding network. The height of the peak at the reference atom is proportional to the number of neighboring water molecules lying next to it. The first peak in this plot is the minimum van der Waals distance between the reference atom and the water atoms that is possible according to the Pauli exclusion principle. The minimum distance in the RDF depends on the nature of

the atoms involved in the interactions. The plot shown here is just to illustrate the idea behind it, however, it is not the actual result calculated from the solution trajectories. Detailed explanations and the RDF plots for the micelle head group atoms are provided in the chapter 5, section 5.5. The water molecules in the bulk are represented as light blue circles which are far away from the reference atom.

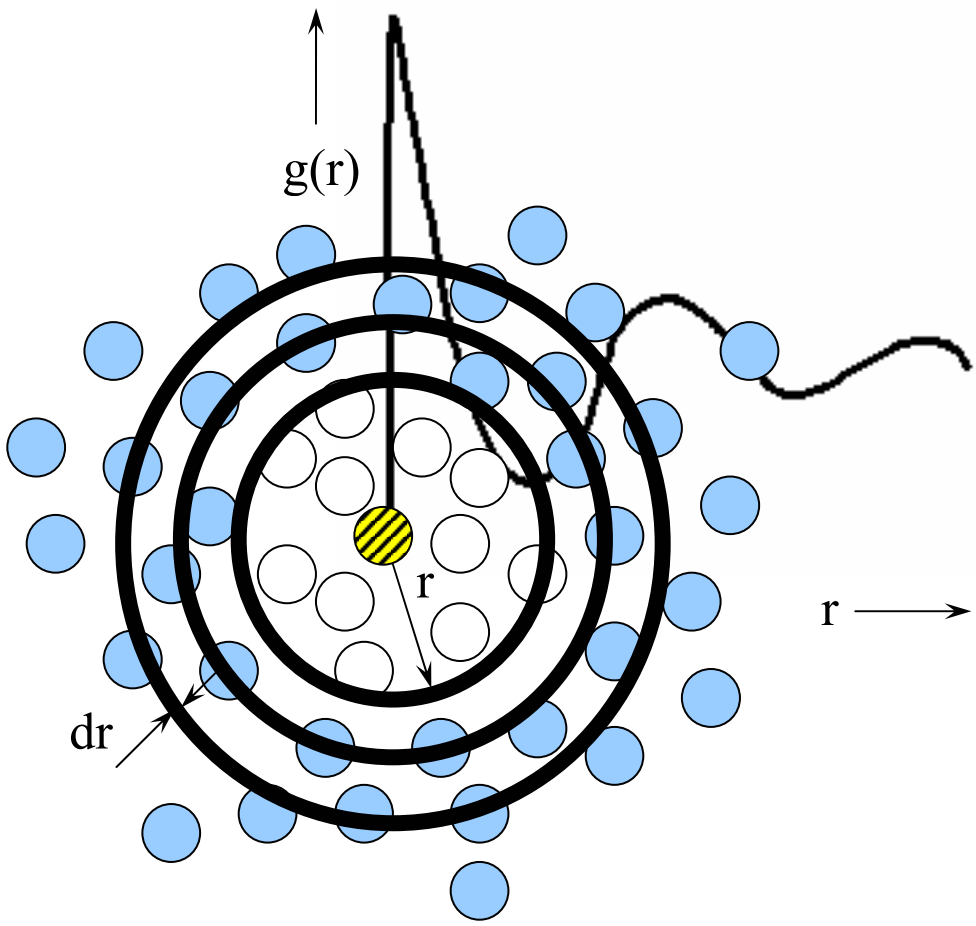

Figure 4-2: Radial distribution function for the reference atom (yellow strip circle,) at the center constructed with a radial distance r. Bulk water is shown in light blue circles .

Analyses of these distribution functions are particularly effective for quantifying the average structure of the disordered molecular systems. As we are dealing with lipid and lectin systems in the solution phase where there is continuous movement of atoms inside the simulation cell, it is extremely useful to construct these distribution functions and understand their primary interactions between the selected atoms.

4.1.5 Diffusion coefficients of OG micelle

The diffusion coefficients from the glycolipid simulations are not calculated from the phase-space trajectories stored during the simulation run. A very recently developed non-orthonormal lattice cell (Rhombic Dodecahedron (RHDO)) type used in the glycolipid simulations prohibit the update of the image centering atoms by the periodic boundary conditions (Dixon et al., 2002; Konidala et al, 2005). The image atoms are not translated in accordance to the primary atoms in the main cell. Also the TIP3P water model potentials applied with the Particle-Mesh Ewald electrostatic summation method seems to overshoot the experimental diffusion coefficient values by a factor of more than twice (Feller et al., 1996; Mark and Nilsson, 2002). Because of these inconveniences encountered with the latest methods employed for the glycolipid simulations, the diffusion coefficients are estimated through several correlations and compared to the experimental values.

The correlation proposed in this work is modified from the derived model published by He and Niemeyer, 2003. The authors used this correlation for different proteins to estimate their diffusion coefficients.

$$D = \frac{6.85 \times 10^{-8}\ T}{\eta \cdot \sqrt{MW^{1/3}\ R_g}} \quad . \qquad 4.12$$

The constant 6.85×10^{-8} in this equation was taken from the literature (Tyn and Gusek, 1990; Durchschlag and Zipper, 1997), which is obtained by fitting to the experimental data, T is the temperature at 298.15 K, η is the solvent viscosity in cP, MW is the molecular weight of the molecule and R_g the radius of gyration of an aggregated molecule in Å. To compare the correlation results, diffusion coefficients calculated with the structural parameters from the MD simulations using the Tyn and Gusek correlation (Tyn and Gusek, 1990) and Einstein-Stokes relationship (Bogusz et al., 2001) was also estimated for the OG micelle. The radius of gyration for the first method (equation 4.13) and hydrodynamic radius (R_H) for the latter (equation 4.14) is the only correlation parameter employed in these equations

$$D = 5.78 \times 10^{-8} \left[\frac{T}{\eta R_g} \right] , \qquad 4.13$$

$$D = \frac{k_B T}{6\pi\eta R_H} \quad . \qquad 4.14$$

Keeping in mind that lipid systems are highly dynamic in nature there is no single correlation existing till now which could estimate correctly the diffusion coefficients of aggregated structures such as OG micelles (Bogusz et al., 2001; Dixon et al., 2002; He et al., 2003). Thus some time was spent to modify the existing empirical correlations. It was taken account of the lipid structural properties obtained in the MD simulations to estimate the diffusion coefficients of the OG systems in the isotropic solution region. The modified correlation (equation 4.15) given here uses a new dimensionless concentration ratio term in addition to the molecular weight and radius of gyration parameters. It is expected that this can be applied successfully to the OG systems above the *cmc* to the entire isotropic region to estimate the diffusion coefficients. For the concentration near to the *cmc* the micelles are assumed to be spherical and the Einstein-Stokes equation can be solved easily (equation 4.14)

$$D = \frac{6.85\times10^{-8}\ T}{\eta\cdot\left[\sqrt{\left(\frac{C_{MD}}{C_{cmc}}\right)\left(MW_{mic}\right)^{0.5}} + R_g\right]} \quad . \qquad 4.15$$

In equation 4.15, C_{MD} is the concentration of the solvated OG in M, C_{cmc} is the concentration of OG at *cmc* (0.025M), and MW_{mic} the molecular weight of an OG micelle. The other quantities in this equation have been already defined. The higher the concentrations in the MD simulation with respect to the *cmc*, the higher the values of the concentration ratio, micelle molecular weight and R_g and thus lower the D value as a function of OG concentration in the simulations.

4.2 Carbohydrate recognition protein-ligand system properties

4.2.1 Lectin conformation analysis

The conformational stability of the proteins and the reliability of the MD simulations were determined by the root mean square deviations (RMSD) of the heavy atoms, backbone and side chain atoms in the lectin/protein monomer (Bryce et al., 2001; Reyes et al., 2001; Yan et al., 2003; Faraldo-Gomez et al., 2003). The

deviations of the protein structures in the simulations from the initial reference structure was calculated by

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i^{md} - r_i^{r})^2} = \left\langle (r_i^{md} - r_i^{r})^2 \right\rangle^{1/2} , \qquad 4.16$$

where r_i^{md} is the position of the atom i in the MD simulation structure, r_i^{r} is the position of the atom i in the reference structure and N the number of atoms in the molecule investigated which is here the lectin monomer. The reference structure in this case is the initially minimized model of the pea lectin derived from the Protein Data Bank (PDB) (Pletnev et al., 1997). In computing the RMS deviations the overall translational and rotational motions of the lectin have been removed by superimposing each configuration of the lectin from the trajectories at regular intervals onto the initial structure (Sen and Nilsson, 1999). In addition to the time evolution of the solution structures from the simulations, the RMS deviations were also calculated for the individual residues in the lectin averaged over all the atoms with respect to the reference structure.

4.2.2 Fluctuations and dynamic regions in the amino acid sequence

The positional fluctuations of the atoms in a protein monomer averaged per residue were calculated from the root mean square fluctuations (RMSF) with respect to the solution average structure along the dynamic trajectories (Wriggers et al., 1998; Chong et al., 1999; Gerini et al., 2003)

$$\mathrm{RMSF} = \sqrt{\frac{1}{N}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(r_{ij}^{t} - \left\langle r_{ij}\right\rangle\right)^2} = \left\langle (r_{ij}^{t} - \left\langle r_{ij}\right\rangle)^2 \right\rangle^{1/2} , \qquad 4.17$$

where r_{ij}^{t} is the position of the atom i in the trajectory frame j, $\left\langle r_{ij}\right\rangle$ is the ensemble average of the lectin structure in the solution trajectories, M is the number of frames in the trajectory and N the number of atoms of the investigated molecule.

The RMSF were calculated for the lectin non-hydrogen atoms, backbone and the residue side chains. The literature survey states that the proteins or protein complexes exhibit a wide spectrum of flexibilities. They are often associated with two main categories of flexible regions in the proteins (Kumar and Nussinov,

2002). First, is called the systemic flexibility which refers to the small scale fluctuations in the amino acid side chains and the back bone residue atoms of the proteins in their native state. The time scale of the systemic protein flexibility is fast and can be easily computed from the MD trajectory averages. The reported averages for the conformational flexibilities of the pea lectin from the MD simulations are taken from the last 400 ps of the dynamic trajectories. Second, there is the segmental flexibility which refers to the motion of one or two sub-regions in the globular proteins with respect to the other. The motions involved in such cases are restricted to a small segment of the protein, which is in a response to a molecular event most often related to the protein function. To observe such motions in the proteins with the MD simulations they need to be equilibrated for a longer period of time, which is intensive in computational time. Also the availability of the new hardware resources to extend the simulations to observe the protein conformational flexibilities at slower time scales is still demanding. The RMSF were also applied for the evaluation of the Temperature- or B-factor profiles (Karplus and McCammon, 2002) through the relation

$$B_i = \frac{8}{3}\pi^2 \left\langle (\mathrm{RMSF})^2 \right\rangle = \frac{8}{3}\pi^2 \left\langle \Delta r_i^2 \right\rangle \ . \qquad 4.18$$

The B-factors (B_i) were estimated for the backbone atoms of the lectin and are compared to the crystallographic results. The distribution of B-factors exhibit more information on the dynamic regions in the tertiary protein structures or specifically in the amino acid sequence (Karplus and McCammon, 2002). It is already well known that the protein structures are not static, and they are, in fact, in constant movement with respect to each other or with the surrounding aqueous solution (Faraldo-Gomez et al., 2003; Guimaraes et al., 2004; Yuan et al., 2005). The polypeptide backbones and especially the amino acid side chains are continuously moving due to the thermal motion and the kinetic energy of the atoms. Partly the internal motions were also correlated with the protein functions such as in the enzymatic processes. Because of these important issues, the B-factor profiles for the lectin structures from the solvent simulations were estimated to reveal the flexible regions and the inherent dynamics involved along its amino acid residue sequence.

4.2.3 Hydration of lectin and water mediated ligand-lectin interactions

Even though the intramolecular peptide or tertiary protein hydrogen bonding contributes largely to the protein structure and stability it is only effective in the absence of accessible water molecules (Sen and Nilsson, 1999; Bryce et al., 2001; Grigera, 2002). In aqueous solution proteins gain their activity by the strong intermolecular hydration forces. The interaction forces are due to the dipole moments from the electronegative atoms and the proton i.e. the higher electronegative oxygen or nitrogen acceptor atoms attracts nearby positive partially charged hydrogen atoms to form a hydrogen bond as shown schematically in the Figure 4-3.

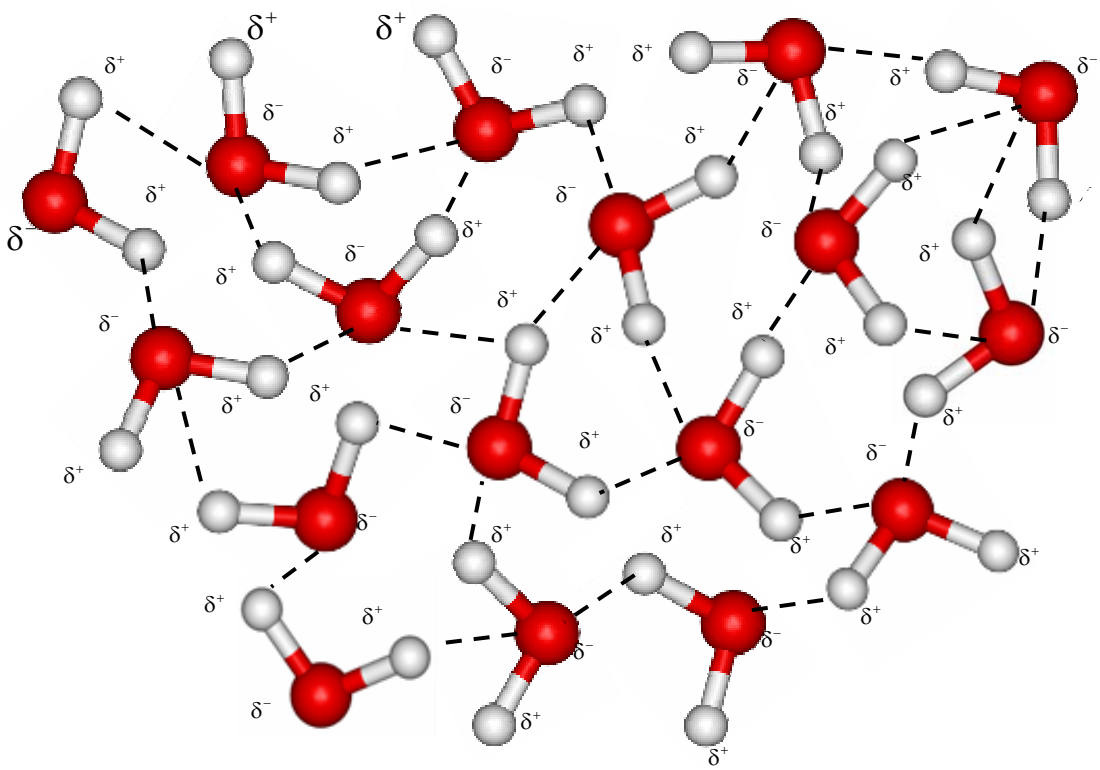

Figure 4-3: Hydrogen bonding network in aqueous solution. Oxygen atoms are shown in red and the hydrogen atoms in light gray color.

A large partial charge difference (dipoles) in the water molecules favors the hydrogen bonding network. The energy of the hydrogen bond is, however; roughly twenty times lower than the covalent bonds. The hydrogen bond is considered to be strong when the distance between the donor atoms and the acceptors atoms are lower and the bond between two electronegative atoms are in a linear fashion (aligned in a straight line). Deviations from the linearity tend to

lower the hydrogen bonding energies. Although the energy of the hydrogen bond is not as large as the covalent bonds, it is the cumulative sum of the hydrogen bonds with the proteins and within solvent molecules that contributes significantly to the protein conformations and also influences the folding processes of proteins in the solution (Reyes et al., 2001; Grigera, 2002; Jang et al., 2004). The presence of these forces causes the protein monomers to visit different conformational states, instantly form hydrogen bonds at their surface next to the water molecules in contrast to the crystal structures or in non-aqueous solutions (Wriggers et al., 1998; Sen and Nilsson, 1999). The hydrogen bonding will, to some extent, loosen up the protein structure and induce fluctuations in the systems. These internal molecular motions in proteins, necessary for the biological activity are partly dependent on the degree of surface hydration (Jang et al., 2004).

The geometric criteria (i.e. the acceptor atom–hydrogen distance and the angle) are the influential parameters for interpreting the results of hydrogen bonding and the water mediated hydrogen bonding (water bridges) (Bryce et al., 2001; Mark and Nilsson, 2002). The water bridge is defined as an interaction between two residues or monomers attached with a common water molecule that makes the hydrogen bonding. The contribution of this term is very important for stabilizing the ligand-lectin complex in the solution. The standard hydrogen bond distance between the donor (D) and the acceptor (A) atoms (D-H---A) is of the order of 2.4-3.5 Å. In the present work hydrogen bonding is recorded when the distance between the acceptor atom and the donor hydrogen is lower than 2.4 Å. Hydrogen bonding interactions and the water mediated hydrogen bonding between the lectin monomer and the ligand monomers were calculated from the last 400 ps of the MD trajectories. In each simulation the OG monomers occupying most of the time by the hydrogen bonding (Reyes et al., 2001; Bryce et al., 2001), and water bridges with the lectin residues were calculated and the results are discussed in detailed in the respective results section, chapter 6.

4.2.4 Water-Ca^{2+}, Mn^{2+}, Cl^{-}-OG residues: Diffusion within lectin systems

The present experience with the glycolipids insists that they are basically flexible molecules in the aqueous solution and posses different conformational isomer states individually and also in the micelle aggregated form. They exhibit considerable conformational complexities in the solution in addition to the overall rotation (tumbling) and translation (Bogusz et al., 2001). In an attempt to assess their dynamic properties we have reported fully unconstrained simulations, i.e.

none of the molecular segments inside the tetragonal cell was constrained, though it is not routinely studied for such big systems with large number of degrees of freedom. As the molecules inside the cell are constantly moving (maintaining a homogeneous mixture in the simulations) with respect to each other, their transport properties such as the rate of diffusion are quite different for each individual molecular segment inside the simulation cell. The diffusion process was analysed for the free OG monomers in the solution and the aggregated OG formed spontaneously over the dynamic run. The translational diffusion coefficients of individual molecular segments (solvent molecules, Ca^{2+}, Mn^{2+} divalent ions, Cl^{-} monovalent ions, and OG monomers) in the lectin simulations was calculated from the Einstein's mean square displacement function (Clapham, 1995; Essmann et al., 1995; Feller, 1996; Wriggers et al., 1998; Bogusz et al., 2001; Mark and Nilsson, 2002)

$$\lim_{t\to\infty}\frac{d}{dt}\left\langle\left[r_i(\tau+t)-r_i(t)\right]^2\right\rangle = \lim_{t\to\infty}\frac{d}{dt}\left\langle\Delta r_i(t)^2\right\rangle = 6D \quad , \qquad 4.19$$

where $\langle\cdots\rangle$ denotes for the average over time origins τ for the segment of interest, $r(t)$ is the position vector at time t, and D is the diffusion coefficient. Using this relation the mean square displacement of the segments is calculated by summing the squares of the distance over the time scale from the solution trajectories and thus determining their diffusion coefficients. As stated earlier the diffusion processes play a vital role in the investigations such as (re)construction of cell membranes embedded with glycoconjugates (le Maire et al., 2000; Lauterwein et al., 1979), ligand as a drug target carrier and in other biotechnical applications.

5 OG micelle simulations – results and discussion

The structural properties reported for the OG micelle are derived from the well equilibrated simulations. The potential energy of the systems was conserved at the very early stage of the dynamic simulation (Figure 5-1). In general, it is been assumed through the ergodic hypothesis that the time average properties from the MD simulation is the same as the ensemble average obtained by the Monte Carlo simulations or experiments, albeit the time averages from the MD simulations are in the nanoseconds scale (Haile, 1992; Leach, 2001). The average time dependent properties of a micelle in the aqueous solution are computed from two different initial velocities and coordinates. It will be certainly of great interest either to perform one very long dynamic simulation run for more than few hundred nanoseconds or many simulations with different initial velocities and coordinates to thoroughly characterize all the relevant time dependent properties of the micelles. However, performing such simulations will obviously require enormous computing power. This, in spite of the development in computer hardware resources, is rendered in this work with the two large simulations, one till 11 ns (bog1) and the other till 6 ns (bog2).

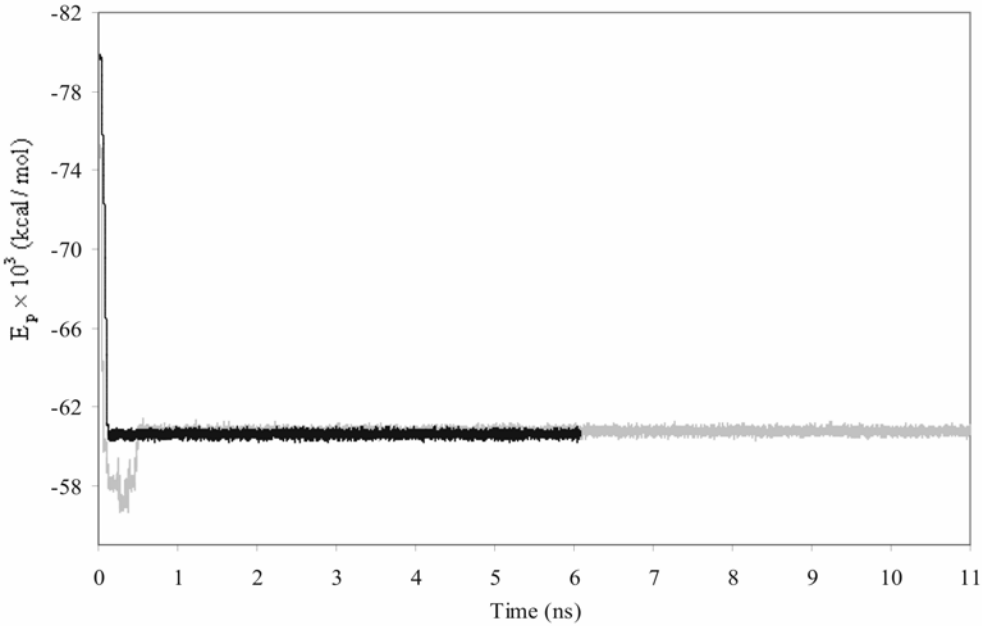

Figure 5-1: Potential energy of the bog1 (light gray) and bog2 (black) system with respect to the simulation time. The energy unit is in kcal/mol. Refer to the text for more information.

As seen in Figure 5-1, both simulations have converged to a constant average potential energy over the nanosecond time scale. The energies reported in this study are in kcal/mol because CHARMM utilizes this distinct system of energy unit for the MD calculations. The computed energies can be converted to SI unit system as 1 kcal/mol equals 4,186 J/mol. In addition, the several parameters applied during the model building and in the simulations were in kcal/mol, which can be compared directly with the other MD theoretical works. Since these systems have been relaxed, the structural properties investigated are considered to be in a thermodynamic equilibrium state. The bog2 simulation started from different microscopic initial conditions reached steady state immediately after the heating stage ($\approx$40 ps) (black color in Figure 5-1). Also the structural properties analyzed were more stable than bog1 simulation with little deviations in the micelle shape. Because of these issues it was contemplated the bog2 simulation has attained equilibrium faster than bog1 and has been terminated at 6ns time scale. The completely diffused OG monomers from the central cell are re-centered with respect to the solvent molecules before revealing the dynamic trajectories for cluster analysis. The structure properties obtained from the simulations and experiments are outlined in the following sections. The diffusion coefficients of the OG micelle was estimated from the modified empirical correlation and compared with the theoretical and experimental data.

5.1 OG monomer structural packing parameter

The conformation of a monomer is one of the main driving force which leads to the formation of different aggregates in the solution. In addition to the geometry of the monomer various other factors like temperature, concentration and ionic strength influence the overall structure of the molecular aggregate. Several geometrical properties (monomer volume, head group area, hydrocarbon chain volume, etc.) were calculated for the OG from the last 1 ns trajectories and averaged (Table 5-1). The OG head group area in the SANS experiment was taken from Nilsson et al. measured at the hydrophilic/hydrophobic interface (Nilsson et al., 1996) and the hydrocarbon tail length of 11.6 Å (maximum length that can be extended) from Tanford, 1980. The head group area in MD was calculated from two perpendicular sides of the outermost glucose head group atoms and averaged over the trajectories with the following valid assumptions: 1) the head groups are heavy and difficult to vary in their dimension, 2) the solution temperature is constant, and 3) there were no ions present in the solution that could influence the area of the head group [Nilsson et al., 1996; Goyal and Aswal, 2001).

Table 5-1: Structural geometric properties of an OG monomer calculated from MD simulations, SANS experiment (He et al., 2000) and the literatures stated therein measured at 25 ˚C

	MD simulations		SANS / Literature data
	bog1	bog2	
Monomer volume (Å^3)	404.5	403.7	422.0[a]
Head group volume (Å^3)	194.5	194.3	179.4
Head group area, a_h (Å^2)	42.9	43.5	40[b]
Hydrocarbon chain volume, v_c	210.0	209.4	242.6[c]
Hydrocarbon chain length, l_c (Å)	8.2	8.3	11.6[d]
Monomer length (Å)	14.6	14.8	14.8
Packing factor	0.59	0.57	0.52

[a] Monomer volume calculated from the apparent molar volume (He et al., 2000).
[b] Head group area of the OG at the air-solution interface (Nilsson et al., 1996).
[c] Calculated from Tanford, 1980 at 25 ˚C.
[d] Maximum possible hydrocarbon chain length that can be extended at 25 ˚C (Tanford, 1980).

MD investigations on micelle aggregates demonstrated in the present work and elsewhere (Laaksonen and Rosenholm, 1993; Kuhn and Rehage, 1997; Bogusz et al., 2000) demonstrate that the hydrocarbon chains were not fully extended inside the micelle core. The existence of trans-gauche dihedral angles causes the hydrocarbon chains to decrease from the extended state. The glucose head group areas and the hydrocarbon tail length calculated from the simulations were reasonable, and the resulting packing factor indicates that the decrease in head group area or hydrocarbon chain length will increase the packing factor value. The critical packing factors of 0.59 (bog1), 0.57 (bog2), and 0.52 calculated from MD simulations, and SANS, respectively suggest that the OG micelle is not spherical. The dimensionless packing factor obtained from the SANS experiment performed at *cmc* and the MD simulations at 0.62 M concentration agree well with each other and prove that the spherical micelles are impossible to form with their molecular geometry (Table 5-1), even at the very low concentration region measured by the

SANS (D'Aprano et al., 1996; He et al., 2000). According to Israelachvili 1998, the packing factor of ~0.50 leads to cylindrical micelles in the solution. The author calculated a packing factor of 0.37 for the sodium dodecyl sulfate and 0.85 for the phosphatidylcholine aggregates with the energetically favoured spherical shape for the first and flexible bilayer or vesicle shape for the latter. In summary, because of its bulky rigid head group and small single hydrocarbon (octyl) chain, the packing factor of above 0.50 for the OG micelle calculated from both MD simulations is more favourable to a cylindrical shape in accordance with the SANS experiments (Lorber et al., 1990; D'Aprano et al., 1996). In spite of some small differences in the geometric properties listed in Table 5-1, good agreement within $\pm 10\%$ has been obtained between simulations and experiment.

5.2 Shape of an OG aggregate in solution

The characterization of the micelle dynamic shape and surface is important particularly when the aggregates in the solution are highly anisotropic, deviating significantly from the spherical/symmetric shape. The moment of inertia tensors were calculated from the coordinates of the system. The corresponding Eigenvalues are diagonalized to obtain three principal moments of inertia. Analyzing the ratios of the three principal moments of inertia (I_1/I_3, I_2/I_3, and I_1/I_2) provides most accurate shape transformations of the aggregate in the solution (Figure 5-2) (Tieleman et al., 2000 ; Gao and Wong, 2001; Kuhn et al., 2002). It is interesting to see from these ratios that large fluctuations in the OG system have occurred at a couple of infrequent intervals over the 11 nanosecond timescale. Nevertheless, the three average moment of inertia ratios calculated from the entire dynamic trajectories were 1.4, 1.3, 1.1 (bog1) and 1.3, 1.2, 1.1 (bog2). The far lowest principal moment of inertia ratio (I_1/I_2) of 1.1 (light gray in Figure 5-2) compared to the other two ratios in the bog1 simulation indicates that the micelle is a prolate ellipsoid (semi axes a = *b*). The micelle in the bog2 simulation was more compact with the ratios a little closer to each other (i.e. the monomers in the micelle stay closer and remain intact). One or rarely two monomers have diffused out of the micelle during the course of the dynamic run in the bog2 simulation in contrast to the bog1 simulation (see Figure 5-2).

Similarly, Tieleman et al. calculated the moment of inertia ratios of 1.4, 1.3, and 1.0 from the dodecylphosphocholine (DPC) simulations performed on the aggregate size of 40 monomers (Tieleman et al., 2000). They also observed from their longer simulation that a major shape change of a micelle occurred for the 54

DPC monomers, leading to less spherical shape on the time scale of above 8 ns (see Figure 3 in Tieleman et al., 2000). The same behavior was observed in our bog1 simulations at around 3, 4.5, and 8-9.5 nanoseconds (Figure 5-2). At these time scales, the two moment of inertia ratios (I_1/I_3 and I_2/I_3) deviate significantly from the third lower ratio. The OG micelle was most often in a prolate ellipsoidal form but sometimes also have a small bilayer-like or rod shaped cylindrical form. The deviations seem to be that the OG micelle tries to restructure its shape into a more complex shape at these instants, but the governing thermodynamic conditions, in particular, the concentration region which were simulated, the geometric constraints of a monomer, and the fixed glucose head group areas at the hydrophobic/hydrophilic interface did not facilitate the formation of such a large asymmetric structure (Nilsson et al., 1996), and so, it starts fluctuating at these time scales. It is concluded from these results that major shape fluctuations in the micelle are always there in the isotropic concentration region, which keeps the OG micelle a small sized ellipsoid or cylindrical rod (D'Aprano et al., 1996).

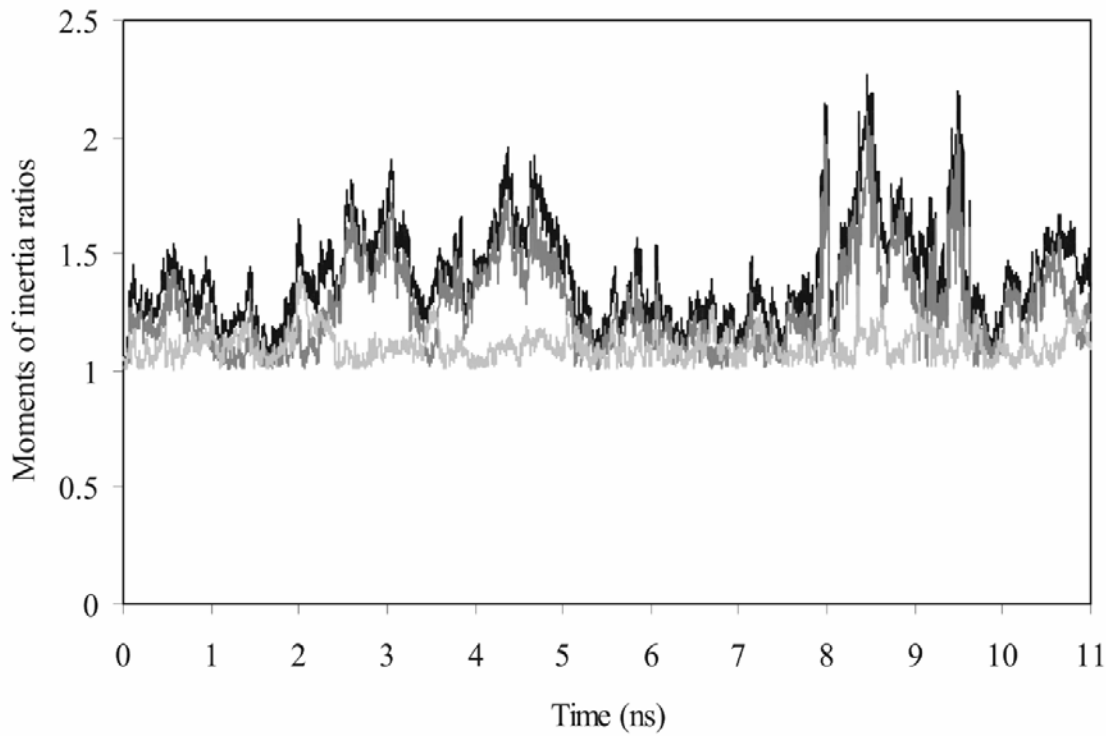

Figure 5-2: Principal moment of inertia ratios of the OG micelle from bog1 simulation. I_1/I_3 - ratio of largest to the smallest principal axis (black), I_2/I_3 - ratio of intermediate to the smallest principal axis (dark gray), I_1/I_2 - ratio of largest to the intermediate principal axis (light grey).

In addition, visual inspection of the trajectories from the bog1 simulation at the 9 ns time scale gave further clues to the overall micelle diffusion and rotation in the solution (Bogusz et al., 2001). A small part of the micelle has diffused out of the RHDO cell at one corner and the completely escaped monomers reentered into the cell in the opposite direction (see Figure 5-3c). In spite of the fact that the micelle monomers were re-centered in the central cell for the analysis of moment of inertia ratios, larger fluctuations in the system were observed. Because of these translational and rotational motions of the micelle, it is suspected that at some point of time, the whole micelle may diffuse out of the RHDO cell (longer MD simulation expected) and reenter from the opposite side of the cell.

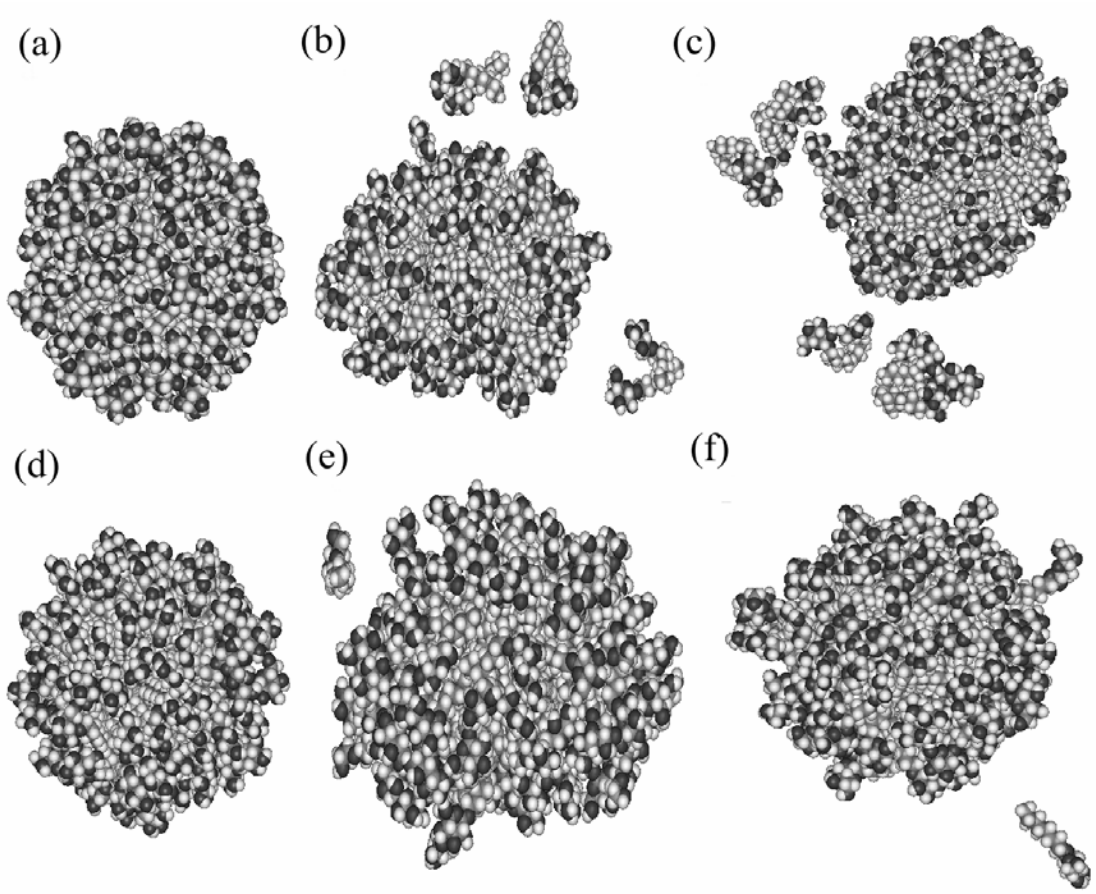

Figure 5-3: Snapshot of the OG micelle from bog1 simulation at time a) t = 0 ns, b) t = 6 ns, c) t = 11 ns and bog2 simulation at time d) t = 0 ns, e) t = 3 ns f) t = 6 ns time scales. The micelle is represented as van der Waals spheres; oxygen (black), carbon (grey) and hydrogen (white). Water molecules were blanked.

The eccentricity (e_s) of an OG micelle was calculated from the principal moment of inertia ratios to observe the elongation of the micelle. The e_s value of 0.62

calculated in this work is consistent with $e_s = 0.60$ from the Bogusz et al., 2000, simulations performed on the same molecule. The results prove that the micelle was elongated, but it did not extend to a bilayer or needlelike rod ($e_s = 1.0$), at least in the present simulations of 92 OG monomers. The aggregate size studied here is much higher than the simulations reported by the aforementioned authors. The observed temporary large shape fluctuations encountered in the solution at a nanoseconds range with the increase in concentration (0.62 M). La Mesa et al., 1993, have shown a similar trend through their dielectric and viscosity experiments that the micelle shape has a slight and continuous influence on the surfactant concentration (La Mesa et al., 1993). Increasing the aggregate size to well above 100 OG monomers might result in the formation of bilayer or complex asymmetric structures in the solution with increasing eccentricities, which is a subject beyond the scope of this work.

5.3 Aggregate surface and roughness

Most of the glycoconjugate molecular aggregates interact with the solvent molecules through the hydrogen bonding scheme. Their surface interaction with the surrounding environment is crucial for many dynamic processes. In the simulations the surface change of a micelle from the ordered spherical shape (Figure 5-3a and d) to the partial asymmetric structure (Fig. 5-3c and f) over time is reflected by the increase in solvent accessible surface areas (ASA) at the beginning of the simulation for the glucose heads and hydrocarbon tails as shown in Figure 5-4. The method developed by Lee and Richards was used with a probe radius of 1.4 Å (to mimic water) rolled over the surface of the micelle to calculate the ASA (Lee and Richards, 1971).

The total ASA of the OG head was nearly constant over the 11 ns simulation (bog1), but the ASA of hydrocarbon tail shows some deviations at the 9.5 ns time scale. As already seen from the preceding section, there were major changes in the micelle shape that occurred at this time (Figure 5-2). The fluctuations in the micelle shape increased the hydrophobic patches. As a consequence, a higher fraction of hydrocarbon chains are exposed to the hydrophilic environment (dark gray in Figure 5-4). These results ensure that very large fluctuations (moment of inertia ratios of above 2.0) in the shape will affect the hydrophobic surface area of the micelle, whereas the smaller fluctuations would not affect the local structure of the monomers significantly (Tieleman et al., 2000). After a few hundred

picoseconds, the ASA for the OG head, tail, and the total monomer remained at an average value of 12800 Å², 3200 Å², and 16100 Å² for the bog1 simulation and 12700 Å², 3100 Å², and 15800 Å² for the bog2 simulation, respectively. The major contribution of the glucose head (139.5 Å²/glucose head (bog1)) and (138.4 Å²/glucose head) (bog2)) to the total monomer ASA (175.2 Å²/OG monomer (bog1)) and (172.0 Å²/OG monomer (bog2)) is due to the aggregate surface roughness and the micelle elongation (Bruce et al., 2002). The rough surface causes water molecules to stay on the micelle surface and interact with the glucose head group atoms. As expected, the contribution of the hydrocarbon tail (35.7 Å²/tail (bog1)) and (33.6 Å²/tail (bog2)) (Figure 5-4) to the total ASA is low because large portions of the chains are facing towards the micellar core thereby avoiding contact with water. The tail ASA in the bog2 simulation is slightly lower than in the bog1 because the structure of the micelle is more compact and prevents hydrocarbon chains from coming into contact with the water molecules.

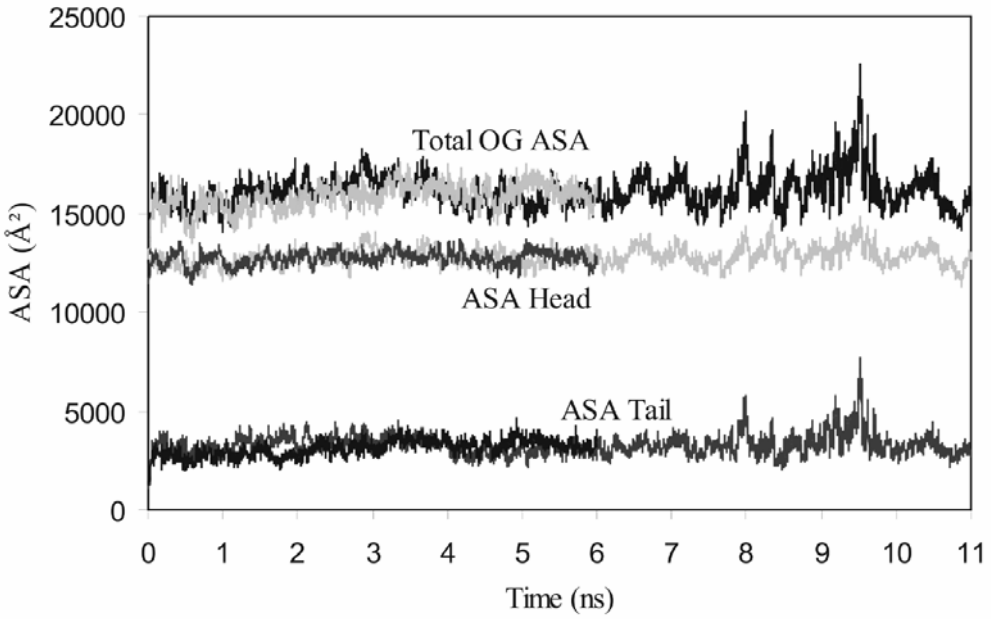

Figure 5-4: Total accessible surface areas (ASA) of an OG micelle (black), glucose heads (light grey) and hydrocarbon tails (dark grey) from the bog1 simulation (11.0 ns). Also shown in the same plot the ASA's of bog2 simulation till 6.0 ns time scale with different color contrast.

The ratio of tail/total ASA obtained for the 92 OG monomers is 20%, which is higher than the ratio of 12% reported for the 67 OG monomers (Dixon et al., 2002). The authors presumed that increasing the size of the OG micelle would reduce the hydrophobic patches exposed to the solvent molecules. The presented results confirm that increasing the OG aggregate size (around 90 monomers) will

certainly increase the exposure of hydrocarbon tails to the water molecules as long as the micelles are asymmetric in structure. Because of its non-spherical shape and partially elongated micelle structure, a higher fraction of hydrocarbon chains were exposed to the water molecules (Bruce et al., 2002).

5.4 Radius of gyration (R_g)

The radius of gyration was also monitored in this MD study to check the equilibration of the system in addition to the usual potential energy fluctuations (Kuhn et al., 2002; Moura et al., 2004). OG monomers those were fully diffused from the RHDO box had been re-centered for the gyration radius cluster analysis. An average R_g value of 19.8 Å (bog1) (black in Figure 5-5) and 19.4 Å (bog2) (light grey in Figure 5-5) was calculated for the OG micelle from both trajectories. The average size of the micelle was constant over the simulation time scales except at around 9 ns where larger fluctuations in the system were observed (Figure 5-5). At these time scales, the OG micelle is more disordered, and the monomers are broadly distributed into the solution, thus increasing the size of the micelle instantaneously.

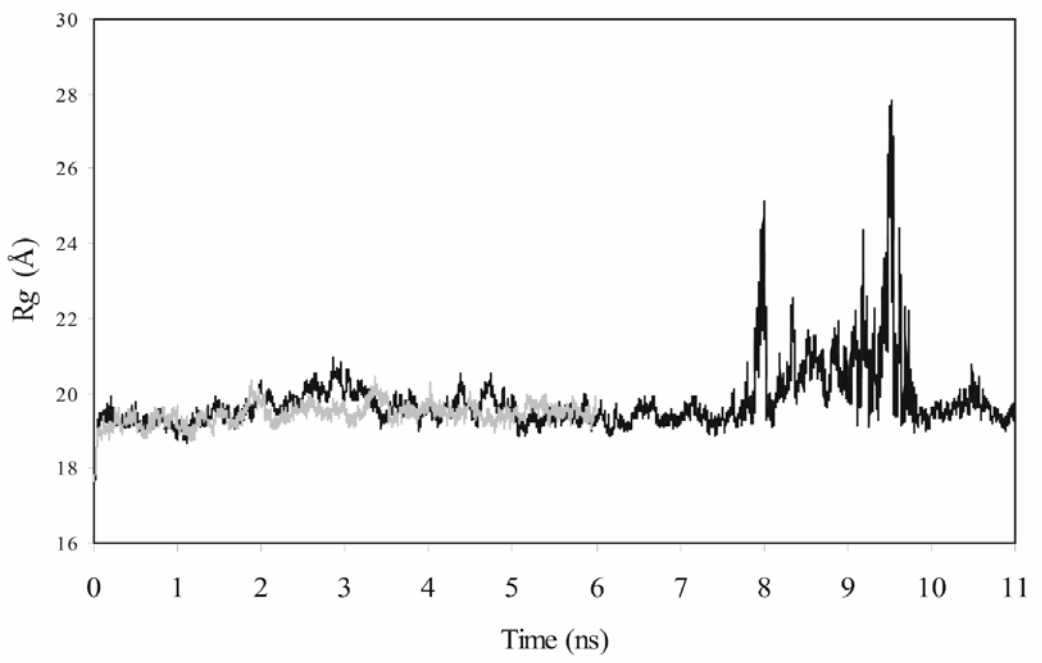

Figure 5-5: OG micelle radius of gyration calculated from the bog1 (black) and bog2 (light grey) simulations as a function of time.

In summary, the radius of gyration calculated from the 11 ns (bog1) trajectories showed that the micelle size in the solution did not increase with time. Because of its polar nature and the fixed geometric areas of the head groups (Nilsson et al., 1996), it exerts a strong hydration force from the aqueous solution (La Mesa et al., 1993; He et al., 2002), which prevents the formation of large anisotropic micelle. The micelle grows very quickly near to the *cmc*. Once it covers the surface with enough glucose head groups consistent with the molecular geometry and free energy requirements of the system, an increase in the number of monomers will not increase the size of the micelle in the isotropic solution region (D'Aprano et al., 1996). Instead, it starts releasing monomers back into the solution, similar to the snapshot of the trajectory shown in Figure 5-3. This proves that the micelle tries to be small and short ranged; no ordering of the micelle to form a bilayer has taken place in the concentration region simulated in this present work.

The comparison of the MD simulations R_g results with the SANS experiments shows a significant difference in these values. The gyration radius of 29.4 Å and 29.1 Å have been calculated using ellipsoidal and cylindrical models from the SANS experiments measured at 0.1 M concentration (He et al., 2002). On the other hand, Nilsson et al. calculated an R_g value of 40 Å from SAXS experiments and 17 Å applying the two-shell model with a spherical micelle core of 11.7 Å and a shell thickness of 5 Å (Nilsson et al., 1996). These variations are often observed between experiments themselves and the theory due to different experimental operating conditions, sample purity in the experiments, and the various treatments of the parameters in the potential model. Previous MD simulations reported an average R_g of 17.6 Å for the aggregate size of 75 OG monomers (Bogusz et al., 2000). Their results show a linear increase of R_g for the different aggregate sizes they studied, which is consistent with these simulations discussed here. In addition, the size of the micelle shows very good agreement with the SANS experiments of D'Aprano et al., 1996. They calculated the R_g of 19 Å for the OG micelle with the cylindrical shape model at the concentration close to the *cmc*.

5.5 Micelle-water interactions

The interaction of solvent molecules with the monomer reference atoms is deduced by the radial distribution functions (RDF) constructed between different atoms of the OG monomer and oxygen atom of the water molecules (Wymore and Wong, 1999; Gao and Wong, 2001). The hydroxyl oxygen atoms (see Figure 1-1) of the glucose head group and oxygen atom of the water molecules show sharp

peaks at a distance of 2.8 Å in the RDF's shown in Figure 5-6. This is the minimum van der Waals diameter between any two oxygen atoms separated by a distance called contact distant *d* (Matteoli and Mansoori, 1995. The second peak appears at a distance of approximately 5.6 Å and reflects the sum of the diameters of two oxygen atoms. It is clear that these atoms are interacting directly with the aqueous environment without any obstructions from the remaining micelle atoms. The presence of micelle atoms at the interface of the aqueous region causes a disruption of the water-hydrogen bonding network (Wymore and Wong. 1999; Bruce et al., 2002). The disturbed water molecules then form hydrogen bonds to the head group atoms and interact with the micelle surface.

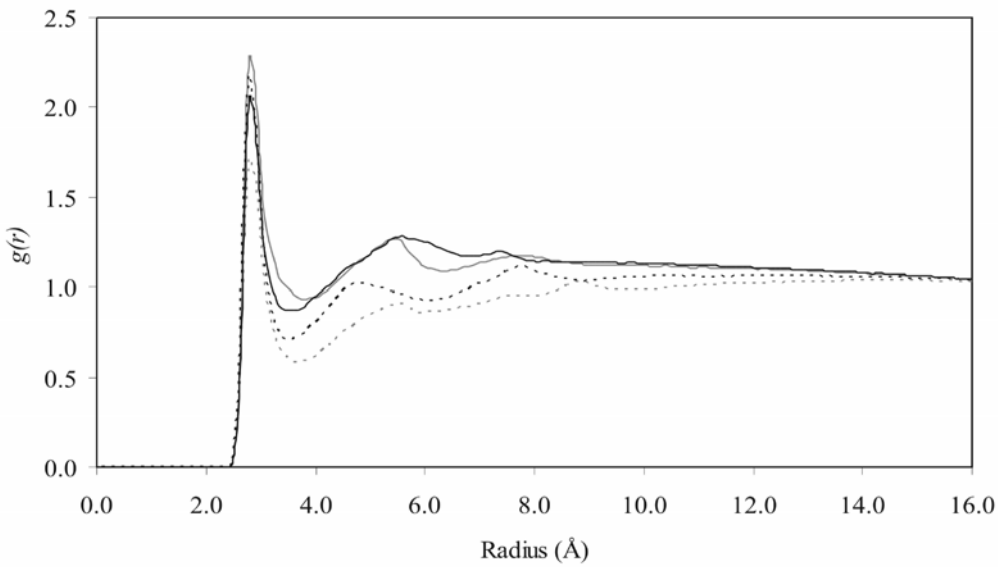

Figure 5-6: Radial distribution functions of hydroxyl oxygen atoms in the sugar head with the water oxygen (O_w) atoms. Refer to Figure 1-1 in the introduction chapter for the atom numbering. O2 - O_w (grey dotted), O3 - O_w (grey), O4 - O_w (black), O6 - O_w (black dottted).

The number of water molecules interacting with the ring oxygen (O5) atom and the acetal oxygen (O1) atom were reduce considerably. The magnitude of the peaks for the ring oxygen (O5) atom (grey line in Figure 5-7) and the acetal oxygen (O1) atom (black line in Figure 5-7) shows an approximately one third reduction of water molecules in the first hydration shell compared to the hydroxyl oxygen atoms. A broader peak area is obtained for these atoms in the second hydration shell due to the ordering of the water molecules near the surface of the micelle. The steric hindrance caused by the other micelle atoms lowers the contact of water in the first hydration shell where most of these oxygen atoms are facing inwards (away from the bulk water) in the direction of the micellar hydrocarbon

core. A similar trend has also been observed for the anomeric carbon (C1) atom (black line in Figure 5-8) and the C7 carbon atom in the hydrocarbon chain connected to the glucose head (black dotted line in Figure 5-8).

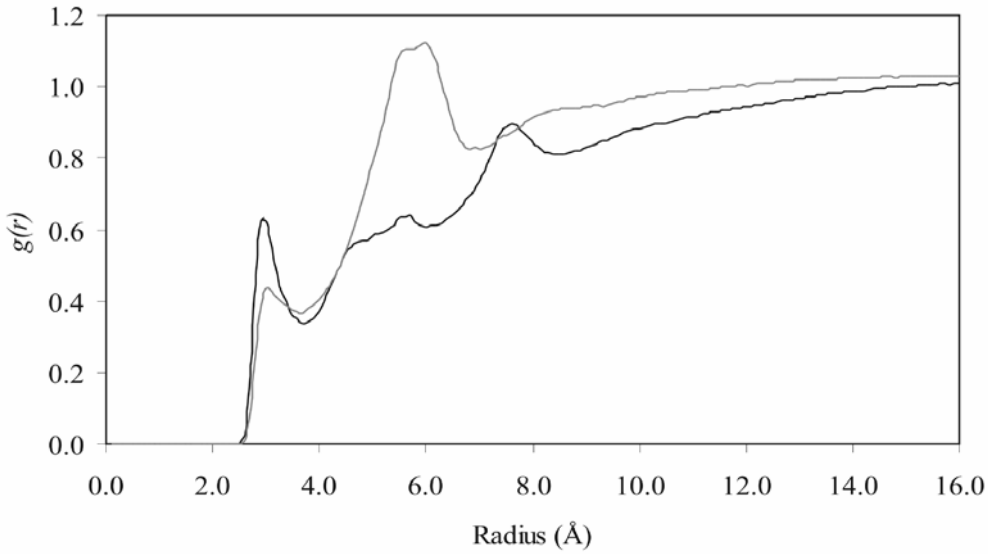

Figure 5-7: Radial distribution functions of the ring oxygen (O5) and acetal oxygen (O1) atoms with the water oxygen (O_w) atoms. Refer to Figure 1-1 in the introduction chapter for the atom numbering. O1 - O_w (black), O5 - O_w (grey).

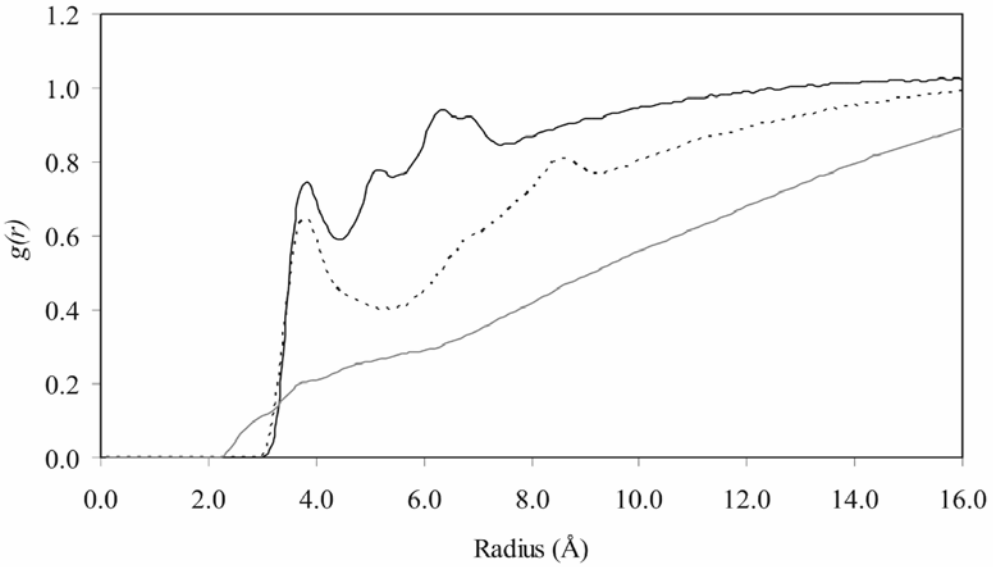

Figure 5-8: Radial distribution functions of the anomeric carbon (C1), C7 carbon atom in the hydrocarbon chain connected to the glucose head and the whole hydrocarbon chain with the water oxygen (O_w) atoms. Refer to Figure 1-1 in the introduction chapter for the atom numbering. C1 - O_w (black), C7 - O_w (black dotted), Octyl chain - O_w (grey).

In addition, the RDF for the whole hydrocarbon chain and water is shown in the Figure 5-8 (grey line), where all the peaks are diffused meaning that there was no considerable number of water molecules involved in the interaction with the hydrocarbon chains. Of course, there may be few water molecules near the acetal oxygen atoms in some OG monomers, which lead to small peaks, but they are negligible compared to the head group atoms. Furthermore, analyzing the RDF plots of oxygen atoms of OG concludes that all hydroxyl atoms in the OG head are primarily involved in the hydration of the micelle. The appearance of sharp pronounced peaks at 2.8 Å (Figure 5-6) for these atom types in the RDF plots are due to the β conformation of the OG monomer (Nilsson et al., 1996) (see Figure 1-1) and the surface roughness of the micelle (Bogusz et al., 2000). In general, the micelle elongation and the presence of an anisotropic rough surface allow water molecules to penetrate into the gaps and interact with the head group atoms efficiently, as compared to the other amphiphilic molecules (Laaksonen and Rosenholm, 1993; Bruce et al., 2002). As seen in Figure 5-6 ($g(r) \approx 2$ at r = 2.8 Å), the number of water molecules near the OG hydroxyl oxygen atoms is about two times higher than the number in bulk water. The water molecules are organized onto the micelle surface and interact with the OG head with a disruption in the hydrogen-bonding network (Wymore and Wong, 1999)

Table 5-2: Average hydration number per OG calculated for the glucose oxygen atoms, anomeric carbon atom (C1), and C7 carbon atom within a radius of 3.0 Å and 5.0 Å

Atom Number	Avg. Hydration No./OG (r = 3.0 Å)		Avg. Hydration No./OG (r = 5.0 Å)	
	bog1	bog2	bog1	bog2
O2	0.96	0.98	5.28	5.37
O3	1.30	1.32	7.90	8.16
O4	1.18	1.22	7.95	8.26
O6	1.20	1.22	6.64	6.72
O1	0.31	0.31	3.20	3.21
O5	0.18	0.18	3.47	3.44
C1	*	*	3.60	3.64
C7	*	*	2.73	2.76

* Average hydration number for these atoms is negligible ($r \leq 3.0\,\text{Å} \approx 0$).

The reduction of water molecules for the O1 and O5 atoms is evident from the average hydration number calculated within a radius of 3.0 Å and 5.0 Å (Table 5-2). Integration of the areas within a 5.0 Å distance in the RDF's ensures that the O1 and O5 atoms are 2.1 times lower in the average hydration number than the hydroxyl oxygen atoms (O2-O4 and O6) (see Table 5-2) from both simulations. Verifying the data within the radius of 3.0 Å for the O1 and O5 atoms for both simulations shows that the average hydration number of 0.25 is 13 times lower in the average hydration number of 3.3 for the same atoms within 5.0 Å distance, whereas the difference is only around 6 times for the hydroxyl oxygen atoms between these distances. Published experimental values for the average hydration number per glucose head group are between 4 – 12 water molecules (La Mesa et al., 1993; Nilsson et al., 1996; He et al., 2000). The determined average hydration number per glucose head group within 5.0 Å of the OG head in both simulations is around 7 water molecules, which is consistent with the experimental values.

5.6 Diffusion coefficients (D)

As stated before the RHDO lattice type used in the simulations pose difficulties to update the image centering atoms for the calculation of diffusion coefficients through the dynamic trajectories. Since the micelle size is small and the shape is fluctuating around prolate ellipsoids with the small inertial ratios, the application of the proposed empirical correlation (Equation 4.15) in the isotropic concentration region can estimate the diffusion coefficients close to the experimental results (Focher et al., 1989; D'Aprano et al., 1996; Nilsson et al., 1996; He et al., 2000; Dixon et al., 2002; He and Niemeyer, 2003). Figure 5-9 shows the diffusion coefficients for the estimated (filled black symbols) and experimental (open symbols) measurements at various single concentrations, in addition to the measured D for the whole isotropic solution region (0 - 2.0 M) from the NMR experiments (open triangles) (Nilsson et al., 1996).

Other theoretical methods through which D can be estimated are also shown in the Figure 5-9 (filled grey symbols). There exists a clear difference in the D values calculated with the Einstein-Stokes (filled grey triangle) (Bogusz et al., 2001) and the Tyn and Gusek correlation (filled grey circle) (Tyn and Gusek, 1990) to the one estimated with the newly developed model (filled black symbols). The D value of $2.7\times10^{-11}\,m^2/s$ (filled black triangle in Figure 5-9) calculated for the OG micelle at 298.2 K is consistent with the NMR experiments performed by Nilsson et al. (open triangles). They stated in detail that the concentration

dependence of the diffusion coefficients becomes weaker at the concentration above 20 wt% (0.68 M). This region is indeed close to the concentration of the OG micelle employed in the present MD simulations (0.62 M). It is presumed that the micelle size will also get stabilized at this concentration region. Since the size and shape of the micelle changes above the *cmc* (0.025M), the utilization of the Einstein-Stokes relationship as expected at higher concentrations deviates significantly from the experimental results (filled grey square and triangle). The new correlation also seems to produce a reasonable D values for the concentration above 0.62 M for the whole isotropic solution region in close agreement with the experiments (Nilsson et al., 1996; D'Aprano et al., 1996).

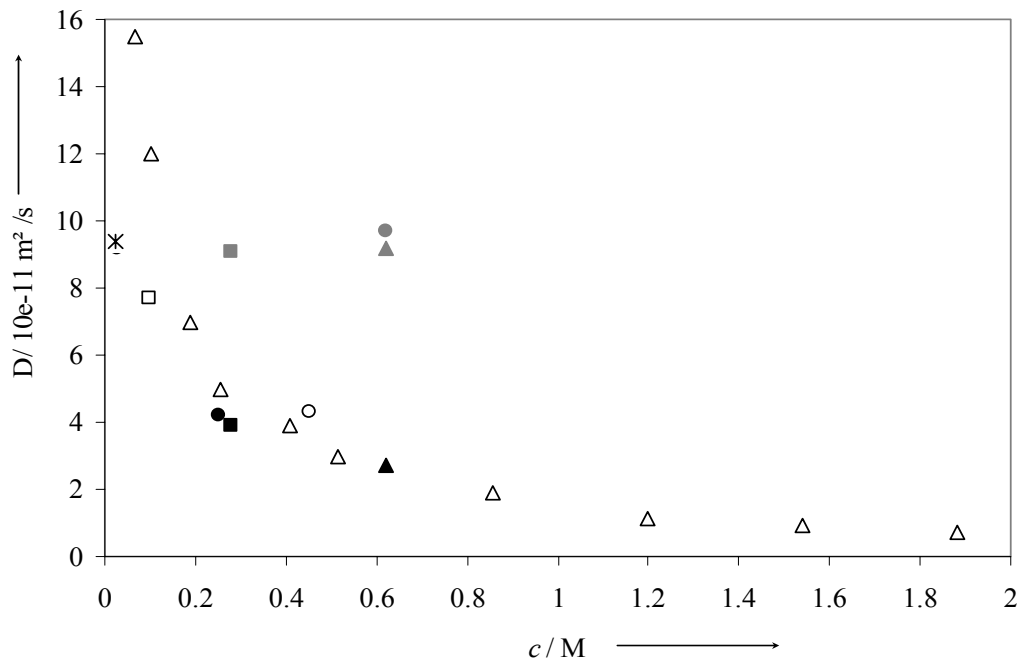

Figure 5-9: Estimated (filled symbols) and experimental (open symbols) diffusion coefficients for the OG micelle in the isotropic solution region with different correlations and experimental methods.

The empirical correlation proposed here was also tested for other MD simulation data (Bogusz et al., 2000) to estimate the diffusion coefficients. The D values of 3.9 and $4.2\times10^{-11}\ m^2/s$ was obtained for the 75 (R_g =17.6 Å) and 67 (R_g =16.9 Å) OG monomers with respect to the concentration and radius of gyration parameters stated therein (light gray in Table 5-3). The estimation calculated for their model seems to be in a very good agreement with the experimental data at that concentration (filled black square and circle in Figure 5-9).

Table 5-3: Estimated and experimental diffusion coefficients from different correlations and concentration ranges

	Correlations / Methods	Aggregation number / concentration	R_g (Å)	D (10^{-11} m^2/s)
Estimated D values from different correlations	Proposed correlation	92 (0.62 M)	19.8	2.7
	Einstein – Stokes	92 (0.62 M)	19.8	9.2[a]
	Tyn and Gusek	92 (0.62 M)	19.8	9.7[b]
	Einstein – Stokes	75 (0.28 M)	17.6	9.1[c]
	Applied to model in Bogusz et al., 2000.	75 (0.28 M)	17.6	3.9[d]
	Applied to model in Bogusz et al., 2000.	67 (0.25 M)	16.9	4.2[d]
Experimental work by others	DLS[f]	0.1 M	7.1	
	NMR[g]	0.45 M	4.3	
	QELS[h]	0.025 M	9.2	
	DLS[i]	0.025 M	9.4	
	NMR[e]	≈0.6 M[e] (0 - 60 wt%)	2.5 – 2.8	

[a] Calculated from the hydrodynamic radius through the Einstein-Stokes correlation (Bogusz et al., 2000; He et al., 2000).
[b] Calculated from Tyn and Gusek through R_g as a correlation parameter (Tyn and Gusek, 1990).
[c] Estimated by Bogusz et al. with the Einstein-Stokes correlation.
[d] Proposed correlation applied to the OG data of Bogusz et al. with their published R_g and concentration values (Bogusz et al., 2000).
[e] Diffusion coefficients measured for the whole isotropic solution region (open triangles in Figure 5-9) (Nilsson et al., 1996). Diffusion coefficient data taken at ≈0.6 M is around $2.5 - 2.8 \times 10^{-11}\, m^2/s$
[f] Data taken from He and B. Niemeyer, 2003:
[g] Data taken from Dixon et al., 2002.
[h] Data taken D'Aprano et al., 1996.
[i] Data taken Focher et al., 1989.

The estimated diffusion coefficients from the present simulations and the applied correlation with the structural parameters taken from the theoretical studies suggest that the proposed correlation is quite reasonable to estimate D in the

concentration region *cmc* < D_{OG} < 2.0 M. The theoretical determination of the diffusion coefficients to the other colloidal phases of the OG system is not straightforward because of its complex anisotropic shape transformations, which makes the calculation of the diffusion coefficients complicated. Moreover, the importance laid in this work was to estimate the diffusion coefficients for the OG micelles which will be very helpful in understanding the dynamics of the micelle transport (Dixon et al., 2002) and in the glycoconjugate research and processing (Garavito and Rosenbusch, 1986; Patthi et al., 1987; Lee and Lee, 2004).

5.7 Summary of the OG simulations

All-atom molecular dynamics simulations from different initial coordinates and velocities in explicit solvent have revealed the physico-chemical properties of the OG micelle. Two different molecular dynamics simulations of 11 nanoseconds and 6 nanoseconds were performed with the aggregate size of 92 OG molecules to characterize the structure and dynamic behaviour of a micelle at equilibrium conditions. Geometric packing properties determined from the simulations as well as the Small Angle Neutron Scattering experiment state that OG micelles are more likely to exist in a non-spherical shape, even at the concentration range near to the critical micelle concentration (0.025 M). Despite few large deviations in the principal moment of inertia ratios exist, the average micelle shape calculated from both simulations is a prolate ellipsoid. The deviations at these time scales are presumably due to the temporary shape change of a micelle. However, the size of the micelle and the accessible surface areas were more or less constant during the simulations with the micelle surface being rough and partially elongated. The radial distribution functions computed for the hydroxyl oxygen atoms show sharp peaks at a minimum van der Waals distance than the acetal oxygen, ring oxygen and anomeric carbon atoms. The hydroxyl oxygen atoms were pointing outwards at the hydrophilic/hydrophobic interface, make hydrogen bonds with the water molecules, and thus hydrate the micelle surface effectively. The diffusion coefficients were estimated using the modified empirical correlation. The results are in good agreement with the experiments in the micellar isotropic solution region.

6 Pea lectin simulations – results and discussion

The MD simulations were carried out with the pea lectin monomer consisting of α and β chains in an explicit solvent environment. Different lectin simulations were performed with the varying concentrations of OG, Ca^{2+} and Mn^{2+} ligand molecules freely distributed in a large tetragonal simulation cell. The structure, dynamics and the interaction properties of the pea lectin-ligand systems in the solution is outlined in the following sections.

6.1 Flexibility and dynamics of lectin structures

The flexible regions in the three dimensional structure of the globular proteins was customarily judged by the RMS deviations of the backbone and side chain atom coordinates in the MD simulations with reference to the crystal structure (Wriggers et al., 1998; Gerini et al., 2003). Despite, the simulations started from different initial conditions, concentrations of OG, number of ions and solvent molecules the global conformations of the lectin have fairly converged to the stable state as shown in the Figure 6-1.

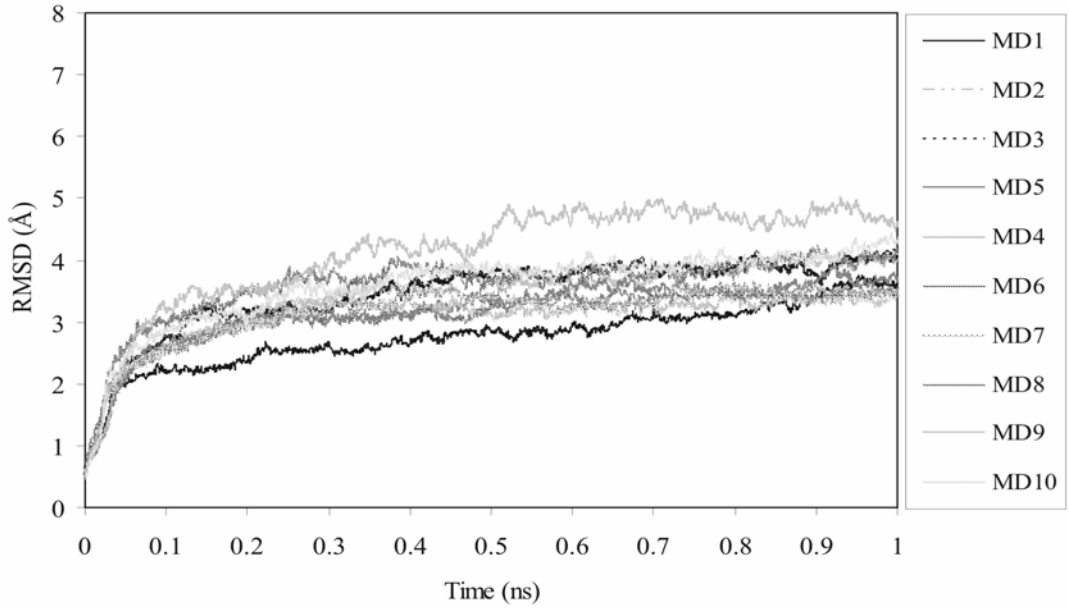

Figure 6-1: Time evolution of the RMS deviations of the pea lectin heavy atoms in the simulations.

Higher molecular fluctuations were observed in the MD9 simulation consisting of two OG monomers in the TETR cell. At above 0.5 ns the deviations in this system

was little higher than the other simulations but as it proceeds further the lectin relaxes in the solvent and slowly approached a decreasing plateau towards the end of the run (grey line in Figure 6-1). The exact reason for this temporary deviation was not known, but it is strongly suggested that due to the lack of enough divalent ions in this simulation might have caused the lectin to fluctuate more before stabilizing the structure. Nevertheless, noticed from the RMS deviations the overall conformation of the lectin in this simulation was relaxed similar to the other simulations at the end of the dynamic trajectory, albeit few larger fluctuations in the middle were observed. Kumar and Nussinov showed that the close range of electrostatic interactions tend to shift the equilibrium towards the native state and constrain the backbone flexibility (Kumar and Nussinov, 2002). However the presence of side chain atoms in the globular proteins seems to fluctuate more in response to the surrounding environment (Jang et al., 2004). The individual backbone conformations of the lectin from the equilibrium MD simulations are shown in the Figure 6-2 in a wire representation with the superimposed starting crystal structure.

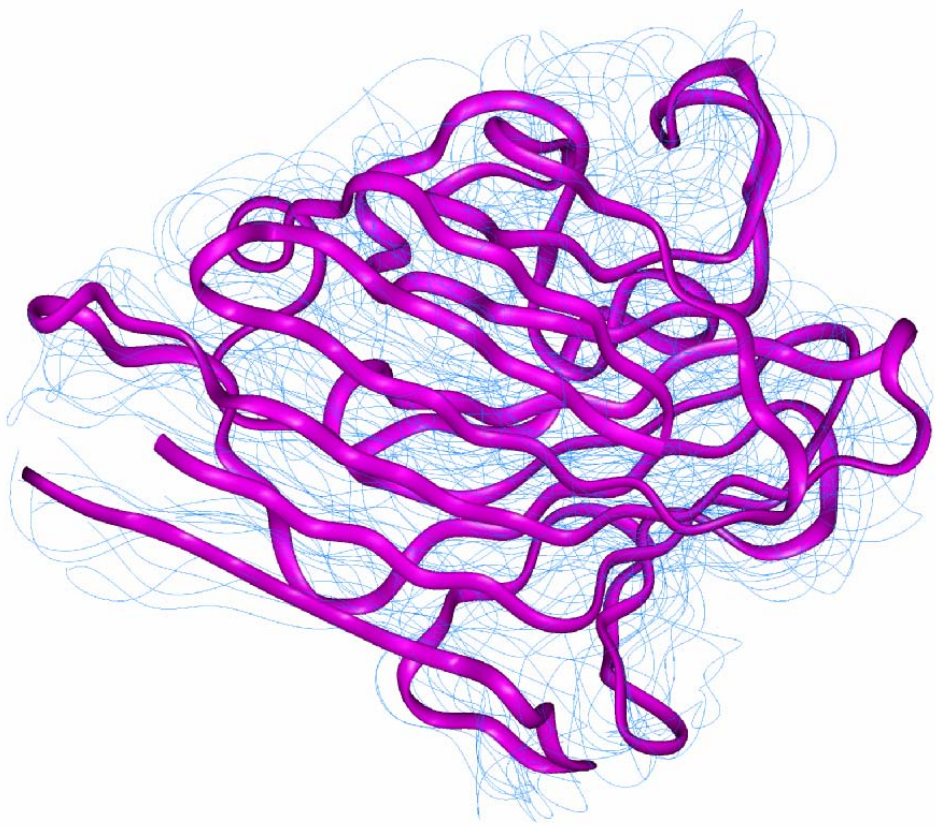

Figure 6-2: MD average flexible lectin conformers from all the simulations shown as wires (light blue) are compared with the superimposed initial crystal structure (violet ribbon). All other molecular components were blanked.

Analyzing the MD trajectories for the average residue RMSD in the solution shows the backbone atoms were more stabilized and fluctuate little in the solution. The lectin monomer has been separated into several molecular components in the order (pea lectin_monomer, monomer backbone, monomer side chains, β-chain, β-backbone, β-side chains, α-chain, α-backbone and α-side chains). The results were included in the Table 6-1. The same terminology is also followed in the other parts of this thesis wherever it is necessary. The RMSD values are agreeing well between the simulations except for the MD9 simulation. The averaged residue RMSD calculated for the MD9 simulation exhibit higher conformational changes in the monomer beta chain residues THR-57, LYS-93, GLN-114, TRP-128 and ASN-132 (Appendix C). These residues are mainly uncharged polar groups and have longer side chains which fluctuate in the surrounding medium.

Table 6-1: Root mean square deviations from the initial minimized crystal structure for the lectin molecular components averaged per residue

MD Simulations	Pea mono	Pea back	Pea side	β-chain	β-back bone	β-side chains	α-chain	α-back bone	α-side chains
MD1	2.13	1.85	2.42	2.20	1.91	2.50	1.81	1.56	2.05
MD2	2.49	2.11	2.86	2.60	2.19	3.00	2.03	1.82	2.30
MD3	2.76	2.42	3.11	2.81	2.44	3.14	2.47	2.21	2.71
MD4	2.39	2.11	2.75	2.55	2.21	2.92	1.76	1.43	2.10
MD5	2.41	2.12	2.67	2.51	2.21	2.81	1.98	1.75	2.17
MD6	2.43	2.12	2.76	2.52	2.21	2.91	2.14	1.95	2.24
MD7	2.32	2.11	2.61	2.48	2.13	2.75	1.65	1.41	1.92
MD8	2.79	2.49	3.10	2.97	2.64	3.32	2.05	1.90	2.19
MD9	3.15	2.79	3.49	3.33	2.95	3.68	2.49	2.23	2.76
MD10	2.64	2.27	2.97	2.80	2.42	3.16	1.99	1.67	2.25

In the Table 6-1 the RMSD are mainly due to the β-side chains and backbone atoms. The flexible side chains induce dynamics in the solution phase and thus exhibit higher deviations. The number of residues involved in the β chain is much higher than the α chain. These residues are constantly involved in the interactions

with the surrounding environment in the solution. Moreover, the size of the lectin calculated from the radius of gyration is not changed in the simulations (Table 6-2). The constant size has been attained in the early stage of the simulation and maintained steadily for the rest of the simulation (Appendix F). This indicates that the conformational shape change observed in the MD9 was only local (Sen and Nilsson, 1999).

Table 6-2: Size of the pea lectin averaged over the dynamic trajectories and the overall Temperature- or B-factor calculated from the mean square fluctuations data (Bryce et al., 2001; Guimaraes et l., 2004)

MD simulations	R_g	B-factor
MD1	17.0	32.4
MD2	17.2	36.0
MD3	17.2	32.4
MD4	17.1	24.7
MD5	17.2	33.0
MD6	17.0	27.9
MD7	17.3	34.8
MD8	17.4	30.1
MD9	17.6	53.0
MD10	17.3	36.6

The internal atomic motions of the pea lectin in the simulations were characterized by the RMS fluctuations averaged per residue around the average solution structure (Karplus and McCammon, 2002). The average RMSF value in each simulation is shown in the Figure 6-3 for the lectin heavy atoms, backbone and side chain atoms. The average data from the independent runs show that the lectin backbone in the simulations was closer to the average solution structure with smaller fluctuations than the lectin side chains (Figure 6-3). Also the fluctuations are symmetric in all the simulations with a little higher deviation in the MD9 simulation (also refer to Appendix D). As it was known already from the above RMS deviations that the structure in this simulation was more flexible and exhibit larger conformational fluctuations in the solution. However these fluctuations are quite acceptable for dealing with such large systems (Reyes et al., 2001; Jang et al., 2004).

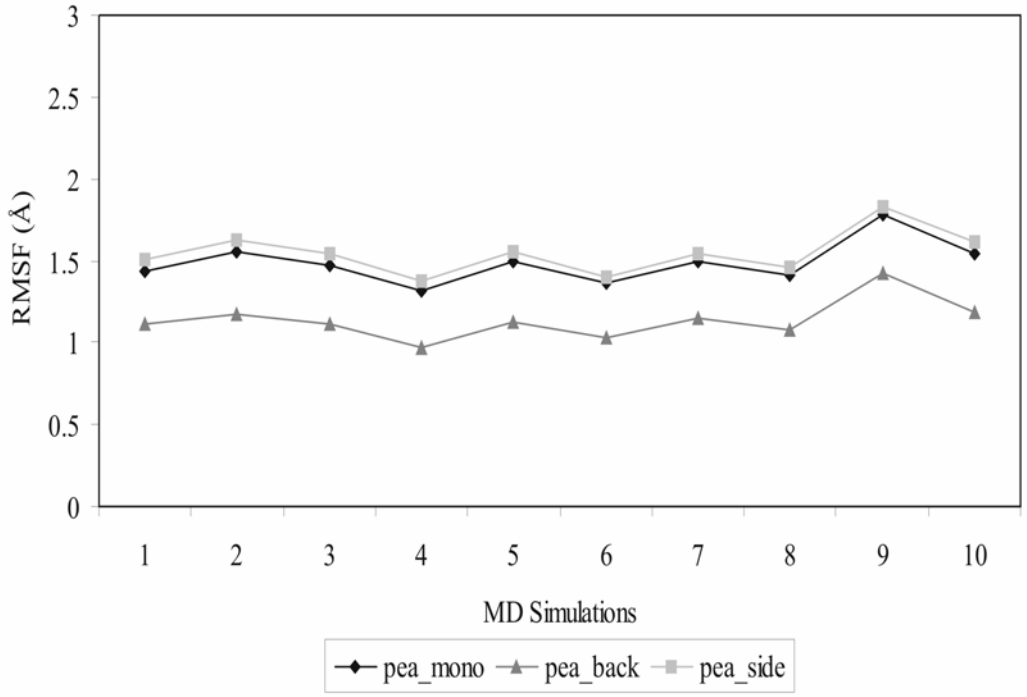

Figure 6-3: Average RMS fluctuations computed independently with respect to the solution average structure are compared between simulations.

Due to the thermal motion of the atoms in the globular proteins the molecular components were constantly moving in the solution. The B-factor is calculated to quantify the amino acid residues from the protein structure leading to larger dynamics. The B-factor determined from the MD simulations was calculated from the RMSF of the backbone atoms of the lectin (Sen and Nilsson, 1999; Wriggers et al., 1998) and was compared to the experimental results. The analysis of the B-factor will provide important information about the protein dynamics (Karplus and McCammon, 2002; Yuan, et al., 2005). The B-factors were also used in a variety of applications, such as prediction of the protein flexibility and thermal stability, analyzing the active sites and investigating the protein disordered regions. In Figure 6-4 the results of the B-factor from the crystallographic structure is compared with the MD simulations performed in the solution phase for the 20, 10, 5 and 2 OG monomers, respectively. The B-factor profiles for the remaining simulations are included in the Appendix E. The important conclusion from this result is that the larger dynamics observed in the x-ray crystal structure was also retained in the lectin simulations carried out in the solution phase. However, larger dynamics were observed in the solution because of the non-bonded (electrostatic and van der Waals) interaction forces from the surrounding environment and hydrogen bonding of the surface residues in the pea lectin with the OG and the

water molecules. These non-bonded interactions and hydrogen bonding networks are crucial for the stability and flexible movement of the specific residues in the lectin (Kumar and Nussinov, 2002). The overall B-factor determined from the crystallization experiments is 20.4, which is smaller than the values calculated from the average backbone RMS fluctuations in the MD simulations (Table 6-2). The averaged B-factor value from these simulations is 34.1 for the pea lectin.

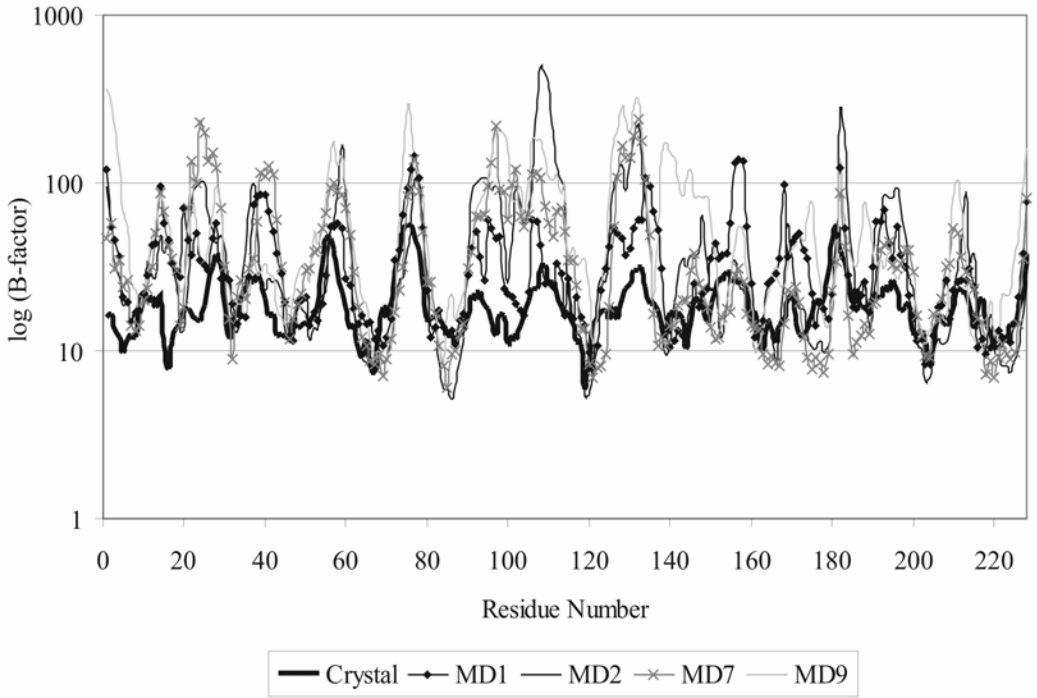

Figure 6-4: Lectin residue dynamics computed from the B-factors are compared between X-ray crystallographic data and the MD simulations performed with 20, 10, 5 and 2 OG monomers.

Interestingly, most of the peaks found in the experimental B-factor profile are well matched with the MD simulations except at few locations in the residues sequence from Pro-12 to Ile-19 and Lys-92 to Tyr-109 (Figure 6-4). The experimental B-factor is not well characterized at those regions because of lacking motion in the crystal structure in contrast to the free all-atom explicit solvent MD simulations. As it is known, proline is a cyclic residue which is most often involved in the turn regions of the β-pleated sheet similar to the region observed in the pea lectin. The latter region (Lys-92 to Tyr-109) is located at the surface of the globular protein in a shallow indent which provides place to the solvent or OG molecules to interact with it effectively. Besides that the 4-aminobutyl side chain of Lys-92 and the three carbon cyclic chain of Pro-93 induce further dynamics at these regions which is not captured correctly in the experimental measurements.

6.2 Surface characterization of the pea lectin monomer

The evolution of the conformational changes in the lectin surface near to the solvent environment plays a fundamental role for the hydration of the side chains of the surface hydrophilic amino acids in most of the globular proteins. As expected the simulations started from the minimized crystal structure increased surface areas at the early stage of the simulation because of the conformational relaxations in the solution. Once the lectin surface was stabilized, providing the constant access to the solvent molecules (Figure 6-5), an increase in the solvent accessible surface areas over time has not been observed in the simulations.

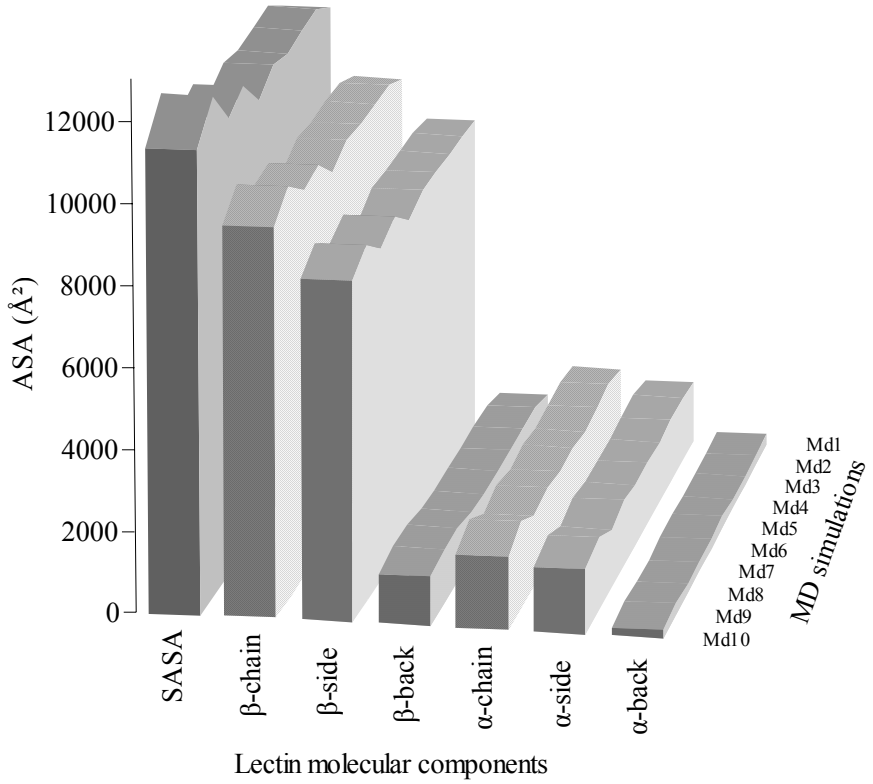

Figure 6-5: Average solvent accessible surface areas (ASA) for the total monomer as well as their individual molecular components of the pea lectin from all the simulations.

The α-chain of the pea lectin in the simulations reached the equilibrium state much quicker than the β-chain. Unlike the β-side chains, the β-backbone seems to be in

good correlation with the molecular components of the α-chain. The only factor which requires more equilibration than the other components in the lectin monomer are the β-side chains. After few hundred ps (~500 ps) the β-side chains in the solution have also reached an equilibrium condition. Due to the higher number of residues in the β-chain (approximately four times larger in size) the equilibration takes longer than for the α-chain entities (Sen and Nilsson, 1999). The ASA of these individual molecular components is shown in Figure 6-5. In spite of the increase in the ASA from the time averaged solution trajectories a good agreement with the results of the X-ray diffraction crystal structure was obtained. The average ASA from the ten simulations calculated with the probe radius of 1.6 Å is 11,200 Å² (Lee and Richards, 1971; Richards, 1977). This value is comparable to 9,200 Å² per pea lectin monomer reported by Pletnev et al. with the same size of the probe (Pletnev et al., 1997).

The relative contributions from each molecular component to the total pea lectin monomer are summarized in the Table 6-3. The results state that more than 80% of the ASA are contributed by the β-chain which is indeed from the β side chains. About only 16 % was contributed by the α-chain to the total monomer ASA.

Table 6-3: Percentage contribution of lectin chains to the solvent accessible surface areas (ASA)

MD simulations	Total ($\times 10^3$ A²)	β-chain (%)	β-side (%)	β-back (%)	α-chain (%)	α-side (%)	α-back (%)
MD1	10.8	83.6	74.2	9.4	16.4	14.0	2.4
MD2	11.2	83.2	72.3	10.9	16.8	14.6	2.2
MD3	11.1	83.7	73.1	10.7	16.3	14.2	2.1
MD4	11.0	84.3	73.8	10.5	15.7	14.0	1.6
MD5	11.2	83.6	73.1	10.4	16.4	14.0	2.5
MD6	10.7	83.7	73.4	10.3	16.3	14.2	2.2
MD7	11.5	83.6	73.3	10.3	16.4	14.5	1.8
MD8	11.2	84.7	72.8	11.9	15.3	13.2	2.1
MD9	12.2	82.5	71.4	11.1	17.5	15.1	2.4
MD10	11.3	84.2	73.5	10.7	15.8	14.0	1.8

In contrast, higher interactions of solvent molecules with the β-backbone (MD8 and MD9) and the α-side chains (MD9) were observed in the simulations. The β-

backbone in these structures was more dispersed into the solution and involved in the direct contact with the solvent molecules. This might be the result of more fluctuations observed in the MD9 simulation (Figure 6-1). Moreover, a very different result has been reported by Wriggers et al. Their MD simulation of calcium-sensing calmodulin protein exhibits a decrease in their ASA against the crystal structure. The authors examined the decrease in ASA's in the crystal and solution structure of calmodulin with the increasing probe radius from 1.4 Å to 20 Å (Wriggers et al., 1998). These studies are useful for the estimate of the likely encountered hydration sites in the globular proteins and to study the overall aggregate surface roughness of the proteins. These studies will not provide details of specific interactions between the protein-ligand complex and the hydrogen bonding/water bridge patterns involved in the active binding site of the protein. This can be evaluated accurately in terms of interaction energies between the molecular complexes (for e.g. protein-lipid, protein-carbohydrate and protein-DNA/RNA). Such studies will intensify the understanding of these complexes and allow quantifying the favourable interaction sites in these biological molecules.

6.3 Glycolipid-protein interaction energies

The specificity of binding/recognition of the OG monomer to the amino acid residues in the lectin was quantified by the interaction energy analysis between OG and the lectin monomer systems simulated in this study. The interaction energies calculated between these monomers are averaged over the last 400 ps trajectories (Figure 6-6). The non-bonded electrostatic (black) and van der Waals (dark grey) terms contributing to the total interaction energy were also shown in the Figure 6-6a, b and c, from the 20, 10 and 5 OG monomer simulations, respectively. It is clearly seen from these profiles that one or rarely two OG monomers in the simulations interact with the lectin monomer on a specific site location. The most specific interaction of MD1 comes from the OG-8 and OG-16 monomers, whereas for in the MD2 it is OG-6 and rarely OG-7. The OG-3 monomer interacts strongly in the MD7 simulation. Electrostatic interaction plays a very important role in all these simulations contributing significantly to the total interaction energy (light grey) (Jang et al., 2004). Moreover the van der Waals interactions have also been observed in the simulations which actually stabilize the lectin further (Lauterwein et al., 1979; Ostermeier and Michel. 1997).

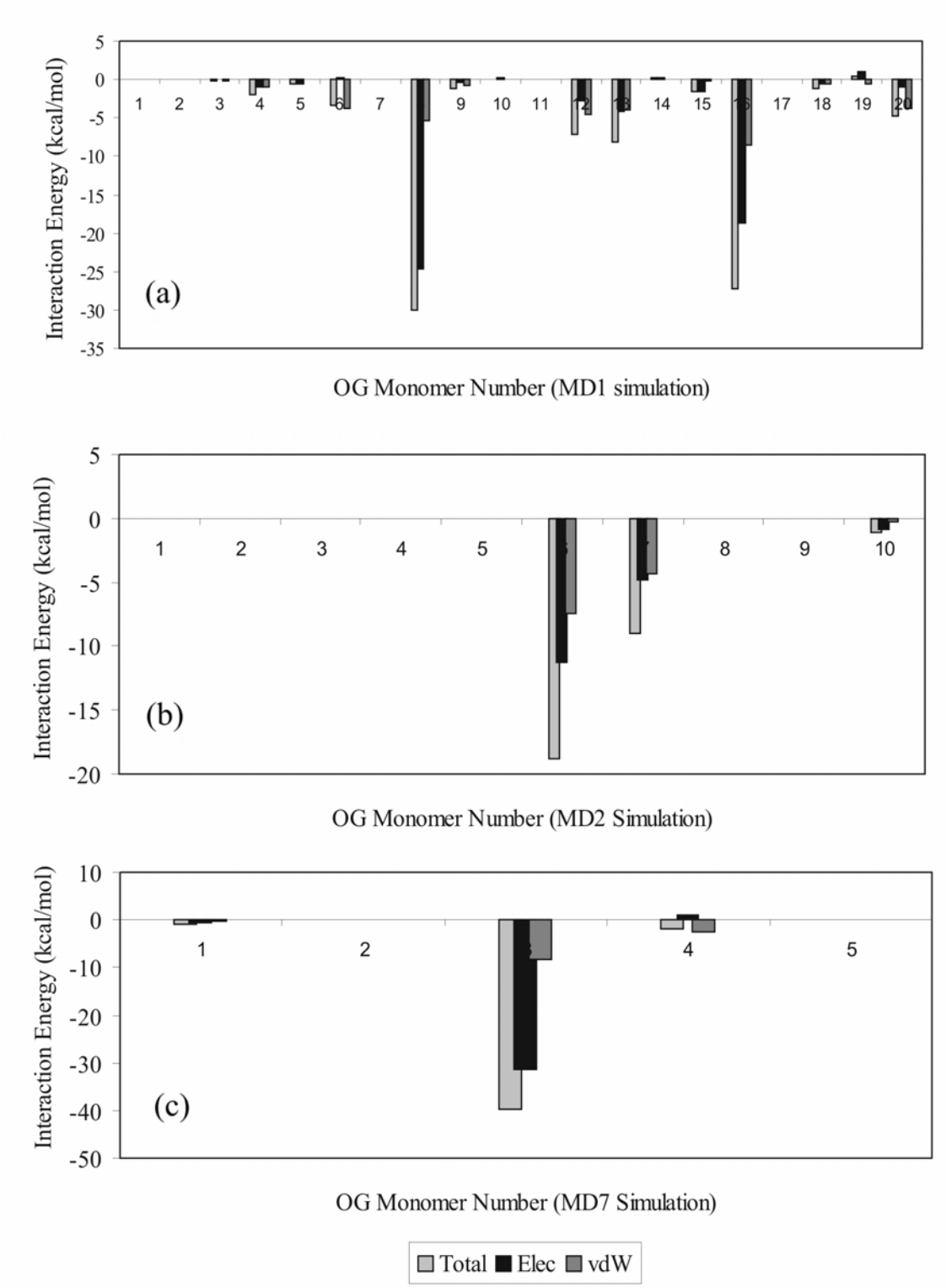

Figure 6-6: Total interactions energies (light gray line), electrostatic (black line) and van der Waals (dark grey line) energy contributions for each OG monomer with the pea lectin from MD1 (20 OG monomers), MD2 (10 OG monomers), and MD7 (5 OG monomers) simulations.

The detailed analysis of the OG binding on the specific site was carried out to check the involved forces which make some of the residues in the protein more favourable for the interactions. As observed from the stability analysis the addition of various concentrations of OG in the simulations was not disturbing the overall conformation of the lectin in the solution. It has to be kept in mind that the detergent concentration used in this work is very close to the critical micelle concentration. Higher concentrations of OG might have an influence on the structure of the lectin; this might be subject to the concern of another detailed work. Abundant research had been carried over decades on the usage of OG as a detergent for the crystallization and solubilization of proteins (Garavito et al., 1996; Ostermeier et al., 1997; le Maire et al., 2000). These studies conclude that the short chain glycolipids were more successful in solubilizing proteins than the longer chains because they fit around the hydrophobic regions without disturbing the hydrophilic inter-residues interactions (Eisele and Rosenbusch, 1989). Thus it is well known that the hydrocarbon chain in OG is small and the addition of these monomers will not hinder the flexibility of the lectin. The perspective is in fact in a good agreement with the interaction studies of Lauterwein et al. performed on the melittin protein with different detergents by a broad range of physicochemical methods. They concluded the interaction with the detergents were similar to that in a phospholipids bilayer environment (Lauterwein et al., 1979). With the past and the present knowledge it is assured that the interaction of OG with the pea lectin will preserve the lectin in a native state. In Figure 6-6a, the interaction of OG-16 with the lectin was stable throughout the trajectories analyzed. The detergent head group interacts strongly with the lectin residues Tyr-100, Leu-101, Ala-107, Glu-108 and Lye-145.

In addition, as shown in the Figure 6-7 the binding site of OG-16 from the MD1 simulation in the solution was very close to the X-ray diffraction crystal structure (Pletnev et al., 1997). The residues involved in the binding site of the glucose ligand in the crystal structure were marked in a yellow color in the Figure 6-7. Even though the ligand applied in the MD simulation is a glycolipid with an octyl hydrocarbon tail, the binding site of the glucose head group on the lectin surface was more or less similar to the glucose ligand determined from the crystal structure. In addition, similar to the observation in the crystal structure, the binding of the Ca^{2+} ligand ion on the lectin was also close to the OG-16 monomer in the solution simulation (red color in Figure 6-7).

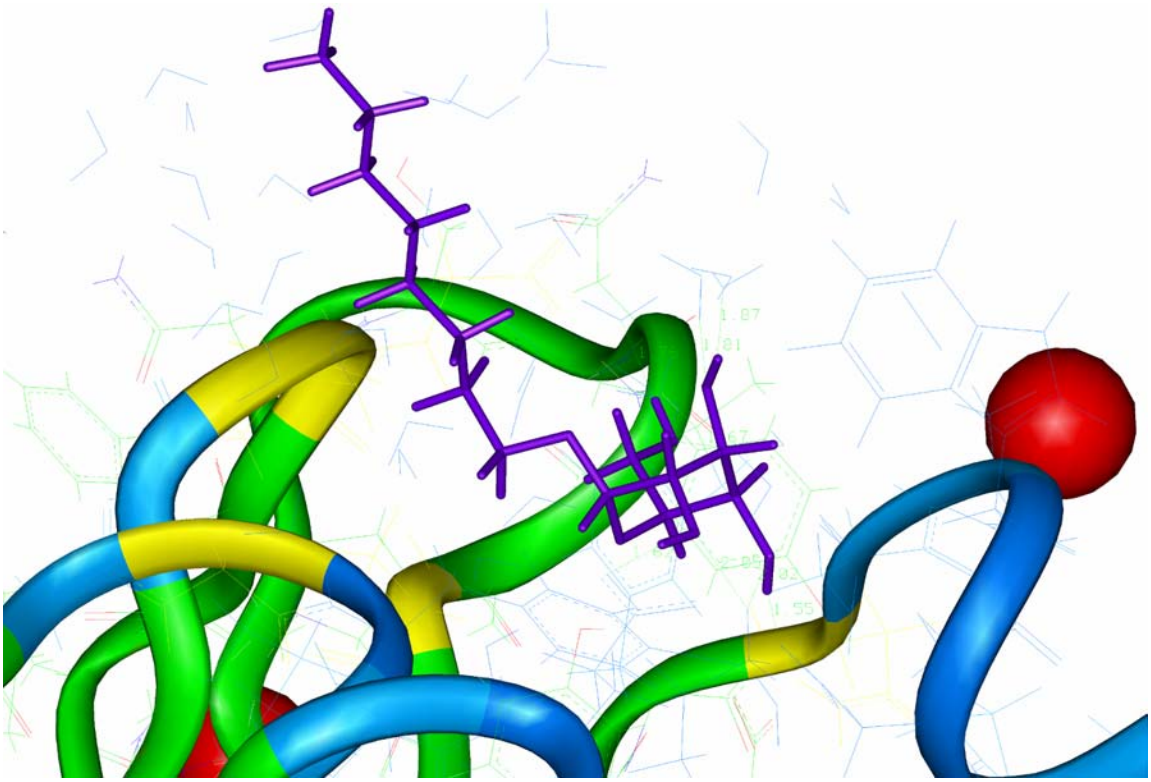

Figure 6-7: Close stereo view of the OG monomer (represented as violet sticks) at the lectin (ribbon) binding site. The OG monomer is stabilized by the Ca^{2+} divalent ion (red) next to the binding site. The yellow colours marked on the ribbon diagram of the lectin are the residues involved in the binding site reported from the X-ray diffraction study (Pletnev et al., 1997).

Different binding mechanism has also been observed in the MD2 simulation in comparison to the MD1 simulation (Figure 6-6b). The portion of the van der Waals energies involved in the interactions is higher than in the MD1 simulation where the electrostatic energies dominate. The two monomers (OG-6 and OG-7) in the simulation were lying completely on the lectin surface (Phe-6, Leu-7, Ile-8 and Leu-18 residues for OG-6 and Ile-8 and Lys-10 for OG-7) and shielded the hydrophobic residues with the OG monomers (le Maire et al., 2000). This must be the main cause to the increase of van der Waals energies for such bindings. The OG monomer tails interact with these hydrophobic residues (particularly isoleucine) very strongly on the lectin surface without releasing back into the solution. The interaction in MD7 simulation is similar to the interactions in MD1 with a single OG-3 monomer interacting with the residues Asp-23, Val-41, Gln-95 and Gly-97 (Figure 6-6c). The glucose head binds next to the charged aspartic acid and glutamine residue and the octyl tail with the valine residue.

Another interesting event is also observed in the MD1 simulation, i.e. "micellization" a very practical mechanism which commonly appears with the detergents used for the protein solubilization. The concentration of the OG in this simulation is just about the critical micelle concentration so it favours to form a micelle of a small hydrocarbon core. From the previous experience with the OG and from the study of other authors it is indicated that the OG aggregation number of 10 monomers or above were more stable in the solution for the formation of small sized micelle aggregates (Bogusz et al., 2000; Konidala et al., 2005). The randomly dispersed OG monomers around the lectin at the beginning of the simulation (Figure 6-8a) were more ordered by the formation of a small micelle (Figure 6-8b) of ~10 monomers.

As noticed from the dynamic trajectories and also from the interaction studies of Figure 6-6a, the OG-13 monomer was initially interacting with the charged residues of Lys-200 and Asp-201 because of favorable electrostatic energies, whereas the OG-8 monomer was interacting with the side chains of high hydrophobic residues Val-172, Leu-173 and Val-196 at the surface. Keeping this OG-8 monomer tightly with the strong hydrophobic interactions at the protein surface, it starts slowly to drag the surrounding OG monomers in order to avoid solvent molecules interacting with the lectin hydrophobic site. Thus the lectin surface residues provide a base for the aggregation of OG monomers. The so formed small aggregate interacts with the other OG hydrocarbon tails and start to build the micelle with the strong hydrophobic core centered on the surface of the lectin hydrophobic site (Figure 6-8b).

Due to this hydrophobic effect the OG-13 and the nearby OG monomers in the solution was influenced by the screening effect of the hydrophobic forces from the newly created micelle. During this process the hydrophobic side chains of the lectin hold the central part (most probably the acetal oxygen of the glucose head group) of the OG-8 monomer close to it with the OG head group pushed little inside the proton cloud of Thr-192 and Asp-195 residues increasing the electrostatic interaction with the lectin and the tail lying in the micellar core (OG-8 monomer marked blue color in the Figure 6-8b). Thus in this case the OG-8 monomer acts as a bridge between lectin and the micelle aggregate.

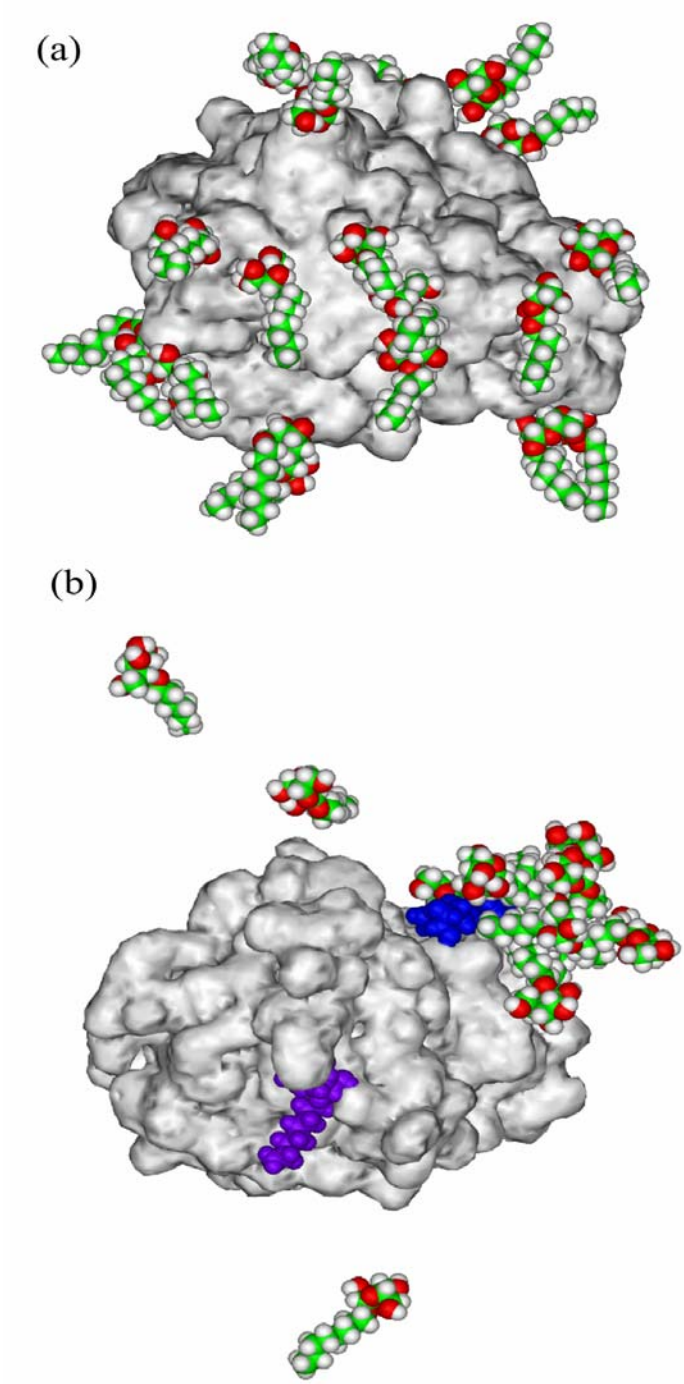

Figure 6-8: The snaps outline the system with 20 OG (van der Waals spheres) monomers. The evenly distributed OG monomers at the (a) start of the simulation have formed a micelle aggregate (b) of about 10 OG monomers at the surface of the lectin (grey solid) with the OG-8 (blue color) acting as a bridge between these two. Also the OG-16 monomer (also refer to Figure 6-6) at the specific binding site interacting selectively with its sugar head group on the lectin surface was shown in violet color.

The monomer involved in the aggregation process was analyzed by the distance matrix averaged over the center of geometry of the hydrocarbon tails in the OG monomers. The distance was calculated for the OG monomers based on the center of geometry of the OG-8 monomer before and after the formation of the micelle. This analysis gives the trace of distance traveled by the OG monomers in the solution and compares it with the initial distances shown in the Table 6-4. The process of spontaneous micelle formation, which usually takes a longer simulation time, in the range of several nanoseconds, was observed to be very quick in this work. It seems that the presence of the lectin hydrophobic side chains at the surface, however, to some extent facilitates the formation of the micelle. In addition to the individual lipid monomers interacting with the proteins (Eisele and Rosenbusch, 1989; Garavito et al., 1996), the formation of the micelle aggregate will also contribute to the solubilization process of the proteins investigated in the laboratory and technical processes to obtain membrane proteins.

Table 6-4: Initial and the final (averaged) distance from the center of geometry of OG-8 tail with respect to the other OG monomer tails that participate in the micelle formation in the MD1 simulation

OG Monomer No.	Initial Distance (Å)	Final* Distance (Å)
OG-3	11.9	11.3
OG-8	3.60	3.80
OG-9	50.5	10.9
OG-10	16.7	11.9
OG-12	18.2	10.9
OG-13	50.9	19.9
OG-15	27.5	8.80
OG-18	30.6	11.1
OG-20	13.3	7.20

* The final distance averaged from the last 400 ps dynamic trajectories

The Figure 6-9a, b, and c, shows the interaction energies for the systems containing either Ca^{2+} or Mn^{2+} ions and neither of these ions (only OG monomers) from the MD4, MD5 and MD6 simulations.

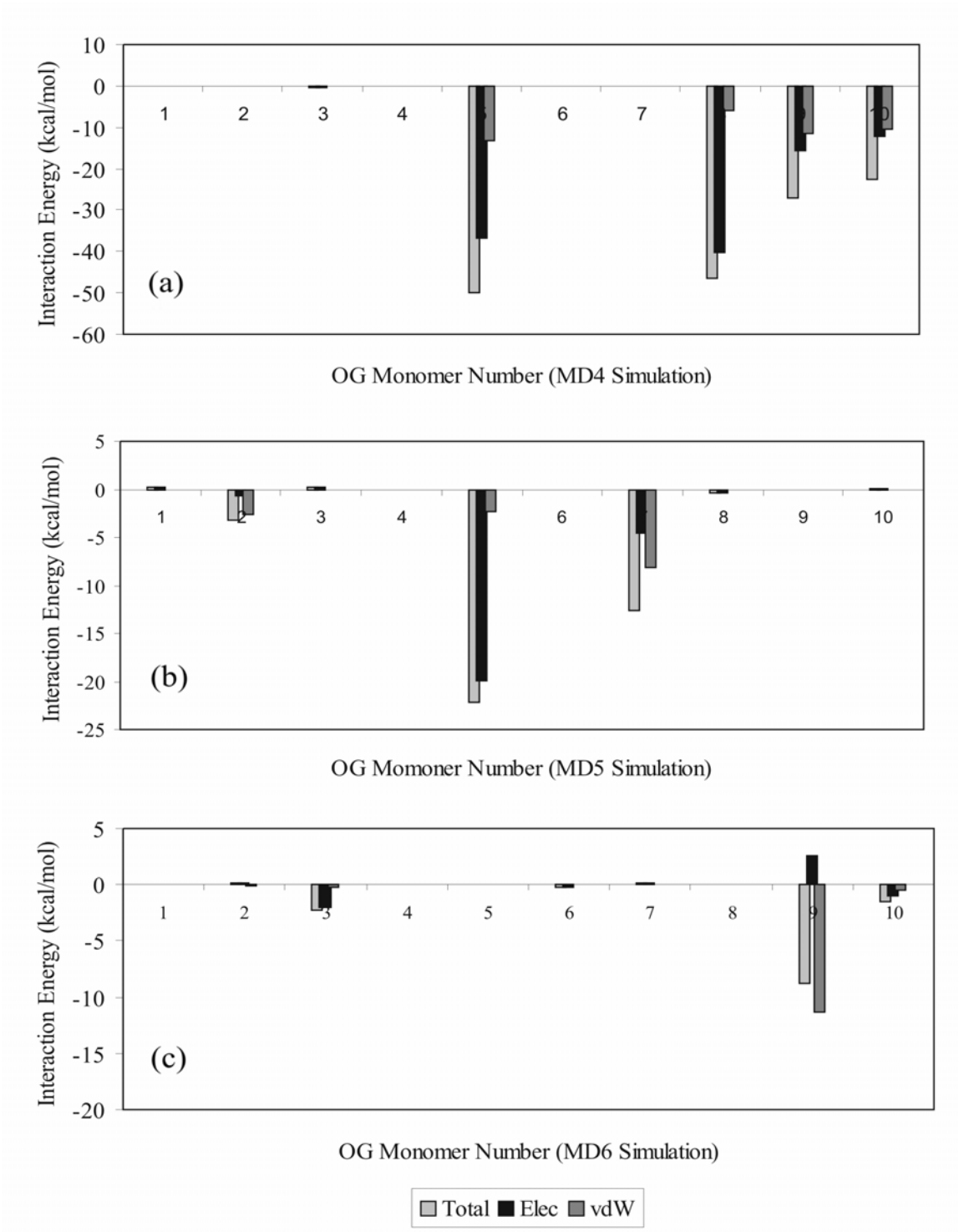

Figure 6-9: Total interactions energies (light gray line), electrostatic (black line) and van der Waals (dark grey line) energy contributions for each OG monomer with the pea lectin from MD4 (10 OG and only Ca^{2+} divalent ions), MD5 (10 OG and only Mn^{2+} divalent ions), and MD6 (only 10 OG monomers, no ions) simulations.

It is clear that the binding of OG monomers to the lectin has been enhanced in the system containing Ca^{2+} ions and OG (Figure 6-9a). Several bindings of OG monomers (OG-5, OG-8, OG-9 and OG-10) are apparent in this simulation with significant electrostatic interactions to the lectin residues. An increase in the magnitude of interaction strength of OG-8 has been observed in addition to the enhanced binding of OG-5 monomer (see ordinate scale) contributing with a favourable electrostatic interaction to the Gly-98, Gly-99, Leu-101, Thr-215 and Gly-216 residues (Figure 6-9a). Evidently, the presence of only Ca^{2+} ions in the simulation increases the binding of OG monomers and also indicates the importance of this divalent ion for the binding of ligand molecules with the lectin. The binding was also near to the crystal structure binding site. In an excellent review article by Clapham 1995, he stated that multiple bindings can be found within the same protein, which acts as a buffer to simply bind the Ca^{2+} ions as the concentration is increased. However, in this specific case with the OG ligands there is no clear evidence to support such behaviour. This point is not yet proved or analyzed by others so it is speculative at this moment.

A relatively smaller number of binding interactions was observed for OG monomers with only Mn^{2+} ions (Figure 6-9b). The OG-5 monomer is the only monomer which shows a larger magnitude of binding to the residues of Glu-108 and Trp-128. Finally, the MD6 simulation with only OG monomers (all ions were removed in this simulation) clearly shows no interaction at all (Figure 6-9c). In contrary to the other systems, the simulation without ions shows a positive electrostatic energy and a high negative van der Waals interaction energy for the OG-9 monomer. The OG-9 hydrocarbon tail interacts with the Phe-11, Ser-12 and Pro-13 residues and was enhanced by the van der Waals forces. This is the only monomer which shows unfavourable electrostatic interaction and a large van der Waals interaction with the lectin. The interaction energies of other OG-monomers are trivial or almost zero with the lectin monomer. From these results one can conclude that the carbohydrate binding to the pea lectin are strongly dependent on the divalent ions. Without these ions the binding of the ligand molecules on the lectin active site would not be possible.

6.4 Hydrogen bonding and water bridging between ligand-lectin complexes

The orientation of water molecules and the OG monomers next to the lectin monomer are influenced by the constantly changing intermolecular attractions of the proton donor and the electronegative acceptor atoms (Grigera, 2002; Jang et al., 2004). The hydrogen bond calculations within the solvent molecules and with the other molecular segments (Table 3-4) produced enormous data because of the large amount of the water molecules. More than 90% of the atoms in the model belong to the solvent, so only the hydrogen bonding and water bridges formed between the pea lectin and OG monomers with a lifetime of above 50 ps from the solution trajectories of 400 ps were listed in the Table 6-5 for all the simulations except the MD9 and MD10.

The high hydrogen bond occupancy observed from the hydroxyl atoms of OG-13 with the Asp-201 residue was in the early stage of the simulation, it was disrupted after the formation of the micelle aggregate. As stated above the hydrocarbon micellar core screens the surrounding OG molecules to diffuse into their hydrocarbon micellar core. After the aggregate has been formed the OG-8 monomer took control of the hydrogen bonding mechanism from the micelle side with the strong attraction to the polar Asp-195 residue. The OG-16 monomer hydrogen bonds with the three different charged residues (Glu-108, Asp-129 and Lys-145) besides a very stable water bridge with the Tyr-100. The presence of this long-lived water bridge additionally stabilizes the protein-lipid complex (Sen and Nilsson, 1999; Bryce et al., 2001).

The hydrogen bond fluctuations in this monomer seem to be higher than the other monomers which indicate that the binding site is more flexible and dynamic in nature. Continues sharing of intermolecular attractions between the hydroxyl atoms of the ligand and the charged residues of the lectin on the binding site might be the cause for the low average hydrogen bond life time observed with this monomer. Also the longer side chain of Lys-145 residue at the active site moves quickly in the solution which leads to disruption and reformation of the hydrogen bonds (Sen and Nilsson, 1999). In spite of these fluctuations the ligand monomer is thermodynamically favourable at the shallow region of the carbohydrate binding site (Figure 6-7 and violet color in Figure 6-8b).

Table 6-5: Hydrogen bonding and water bridges in the simulations formed between pea lectin and OG

MD Simulations	Lectin atoms	Ligand atoms	Average life time (ps)
MD1			
H-Bond	LEC:GLU-108:OE2	HO2:16-OG:LIG	90.0
	LEC:ASP-129:OD1	HO4:16-OG:LIG	70.0
	LEC:LYS-145:HZ2	O2:16-OG:LIG	65.0
	LEC:ASP-195:OD1	HO4:8-OG:LIG	150.0
	LEC:ASP-195:OD2	HO4:8-OG:LIG	97.5
	LEC:ASP-201:OD1	HO4:13-OG:LIG	175.0
	LEC:ASP-201:OD2	HO3:13-OG:LIG	135.0
Water Bridge	LEC:TYR-100:O	O4:16-OG:LIG	220.0
MD2			
H-Bond	LEC:PHE-6:O	HO3:6-OG:LIG	41.0
	LEC:ILE-8:HN	O6:6-OG:LIG	97.5
	LEC:ILE-8:O	HO6:6-OG:LIG	67.5
Water Bridge	LEC:ILE-8:O	HO6:6-OG:LIG	50.0
MD3			
H-Bond	LEC:SER-190:OG	HO2:8-OG:LIG	25.0
	LEC:SER-226:HG1	O4:4-OG:LIG	35.0
Water Bridge	LEC:THR-65:OG1	HO6:4-OG:LIG	30.0
MD4			
H-Bond	LEC:ASN-17:HN	O2:10-OG:LIG	93.7
	LEC:LEU-18 :HN	O2:10-OG:LIG	80.0
	LEC:THR-96:OG1	HO2:5-OG:LIG	60.0
	LEC:GLY-97:HN	O2:5-OG:LIG	60.0
	LEC:GLY-98:HN	O2:5-OG:LIG	90.0
	LEC:LYS-111:HZ1	O2:8-OG:LIG	70.0
	LEC:THR-113:OG1	HO6:8-OG:LIG	230.0
	LEC:THR-215:O	HO6:5-OG:LIG	235.0

Table 6-5 continued

MD Simulations	Lectin atoms	Ligand atoms	Average life time (ps)
MD4			
Water Bridge	LEC:ILE-8:O	HO4:10-OG:LIG	60.0
	LEC:THR-40:O	O4:5-OG:LIG	85.0
MD5			
H-Bond	LEC:GLU-108:OE2	HO2:5-OG:LIG	60.0
MD6			
H-Bond	LEC:LYS-30:HN	O2:9-OG:LIG	60.0
MD7			
H-Bond	LEC:ASP-23:OD1	HO2:3-OG:LIG	280.0
	LEC:ASP-23:OD1	HO3:3-OG:LIG	80.0
	LEC:GLY-42:HN	O2:3-OG:LIG	82.5
MD8			
Water Bridge	LEC:ASN-78:HD21	HO2:4-OG:LIG	80.0
	LEC:TYR-219:O	HO2:4-OG:LIG	225.0

In the MD2 simulation only OG-6 was involved in the hydrogen bond and also the same monomer was occupied in the water bridge with the Phe-6 and Ile-8 residues. As seen the hydroxyl atom of OG-6 hydrogen bond with the Ile-8 residue directly and once it was broken at some instants, the interactions between these two atoms are still similar but the interaction was stabilized through a water bridge. Relatively weak hydrogen bonding pattern has been observed in the MD3 simulation. Interestingly, a very strong hydrogen bonding and couple of water bridges are apparent in the simulation with only Ca^{2+} ions (no Mn^{2+}) (MD4). Out of three monomers OG-5, OG-8 and OG-10 involved in the hydrogen bonding. The first two contribute largely to the hydrogen bonds. As seen in the Table 6-5 the HO6 donor atoms of the ligand are forming hydrogen bonds with the O and the OG1 acceptor atoms of Thr-113 and Thr-215. These results state that the presence of Ca^{2+} ions improves the hydrogen bonding pattern between OG and the lectin monomer. A reduction in the hydrogen bonding pattern was seen in the MD5 simulation with only Mn^{2+} (without Ca^{2+}) ions. The HO2 of OG-5 formed hydrogen bond with the acidic Glu-108 residue. The average lifetime of the hydrogen bond has been lower in this simulation without the presence of Ca^{2+}

ions. Analyzing the results of MD6 indicate that the strength of the hydrogen bonding is decreased to a great extent in this simulation with only OG monomers (without Ca^{2+} and Mn^{2+} ions). The negatively charged Lys-30 residue forms a hydrogen bond with the oxygen acceptor atom of OG-9. The more flexible side chain of this residue moves freely in the solution and creates polar contacts at some instants with the ligand atom. The hydrogen bonding in the MD7 simulation shows that the OG-3 is interacting with OD1 and HN atoms of ASP-23 and GLY-42 residues. Finally, MD8 simulation with the same number of OG monomers as in MD7 shows no hydrogen bonding pattern (with the average lifetime cutoff of above 50 ps) but only the strong long-lived water bridges between the hydroxyl atoms of OG-4 and ASN-78 and TYR-219 residues of the lectin. In spite of the fact that ligand molecules used in the simulations were not constrained and are allowed to move freely in the solution, the interaction and the hydrogen bonding patterns calculated from these simulations were qualitatively similar to the other studies (Bryce et al., 2001).

6.5 Diffusion coefficients of solvent and ligand molecules

From the structural analysis of the lectin it is evident that the properties of the water and the ligand molecules at the binding site are distorted by the lectin residues and their flexible conformations were dictated largely by the surrounding environment (Bruce et al., 2002). The motions connected with the OG monomer are of prime importance for studying the dynamic properties of detergents used in the solubilization process. Translational diffusion coefficients (D) were calculated from the Einstein's mean square displacement function for the water, OG monomers and the ions in the periodic box (Table 6-6). Since the concentrations of OG in the simulations are very close to the critical micelle concentration, it is fully acceptable to use the mean square displacement function to calculate the diffusion coefficients. As shown in the Table 6-6 the diffusion coefficient of water molecules from different simulations are on average $5.4 \times 10^{-9}\ m^2/s$ except in the MD6 simulation without any ions. It should be noted that in general the performance of the TIP3P water model by itself deviates from the experimental viscosity and diffusion coefficients by a factor of 2.8 and 2.2 (Essmann et al., 1995; Feller et al., 1996). The reported average values for the viscosity and diffusion coefficient from the TIP3P with the electrostatic Ewald summation method are 0.35 cP and $5.2 \times 10^{-9}\ m^2/s$ (Feller et al., 1996). The calculated D for the water molecules in the present simulations are indeed very well overlapping

with these works performed previously with the TIP3P water model. Interestingly, the diffusion coefficient of water in the MD6 simulation is much lower than the other MD simulations. It is not proved, however it is suspected that the presence of OG monomers without divalent ions in the protein simulations seems to lower the water diffusion transport behaviour.

Table 6-6: Diffusion coefficients calculated from the mean square displacement equation for the solvent molecules, OG monomers and the ions

MD simulations	D_{water} $10^{-9}\ m^2/s$	D_{mon}[a] $10^{-9}\ m^2/s$	D_{avg}[b] (D_{hd})[c] $10^{-9}\ m^2/s$	$D_{Ca^{2+}}$ $10^{-9}\ m^2/s$	$D_{Mn^{2+}}$ $10^{-9}\ m^2/s$	D_{Cl^-} $10^{-9}\ m^2/s$
MD1	5.4	1.6	0.7 (0.7)	1.3	1.4	3.0
MD2	5.4	1.2	1.2 (1.2)	1.2	1.1	3.2
MD3	5.4	0.9	0.9 (1.1)	1.1	1.4	3.3
MD4	5.4	2.5	1.0 (1.0)	1.1	-	3.4
MD5	5.5	1.1	1.1 (1.0)	-	1.5	3.2
MD6	4.9	0.9	0.9 (0.9)	-	-	-
MD7	5.4	1.1	0.7 (0.7)	0.7	1.0	3.5
MD8	5.4	1.1	0.8 (0.7)	1.9	1.2	3.3
MD9	5.6	0.7	0.7 (0.6)	2.3	1.7	2.5
MD10	5.4	1.3	0.9 (0.8)	0.9	1.7	5.8

[a] free monomer in the bulk solution which is far away from the lectin monomer

[b] average monomer diffusion coefficient of OG monomers in the solution

[c] average OG head group diffusion

Based on the distance from the center of the lectin, the OG monomer lying far away from the lectin surface was selected in each simulation and their free motion in the bulk solvent was calculated. It is shown that the individual monomer diffusion (D_{mon}) is independent of their positions visited in the simulations performed in a large solvent periodic box (Table 6-6). Moreover their average diffusion with the other OG monomers presented in the fourth column of Table 6-6 gives some meaningful results (D_{avg}). The diffusion coefficients calculated for the OG head group listed in Table 6-6 in the parentheses were also synchronizing

with their average monomer diffusion. The average diffusion of the monomers in the MD1 simulation was much lower than their individual monomers because 50% of the monomers in the solution are in the micelle aggregate and thus hinders the motion of OG. The slow motion of these monomers summing up with the other monomers having higher diffusion causes the averaging diffusion of OG in MD1 simulation to a lower value. To further investigate the diffusion constants in this simulation the monomers in the aggregate and in the free solution were calculated individually to see the influence of their D values. As evident from the Table 6-7 the D values are affected by the monomers in the micelle aggregate. The D values are lowered by approximately 2-3 times in the aggregate than in the free monomer state. The OG monomer on the binding site was reduced to more than 5 times the value of the free monomers in the solution.

Table 6-7: Comparison of the diffusion coefficient of OG monomers in the free solution, micelle aggregate and at the binding site

Free OG monomers	D_{free} $10^{-9}\ m^2/s$	Aggregated OG monomers	$D_{aggregate}$ $10^{-9}\ m^2/s$
OG-1	1.6	OG-3	0.8
OG-2	1.4	OG-6	0.5
OG-4	0.8	OG-8	0.7
OG-5	1.1	OG-9	1.0
OG-7	1.6	OG-10	0.9
OG-11	1.1	OG-12	0.5
OG-14	1.5	OG-13	0.6
OG-17	1.2	OG-15	0.8
OG-19	1.0	OG-18	0.6
		OG-20	0.5
		OG on the binding site	
		OG-16	0.3

The diffusion coefficients of the divalent Ca^{2+}, Mn^{2+} ions and the monovalent Cl^- ion in the simulations are also very well reproduced except the simulations MD9 and MD10 (Table 6-6). The Ca^{2+} diffusion coefficient reported in the literature is 0.8-1.0 $\times 10^{-9}\ m^2/s$ (Clapham, 1995). This value is comparable to the MD simulations (Table 6-6) carried out in this work. Similarly Mn^{2+} ions diffusion coefficients were also stable in the simulations and can be compared to the dynamics of calcium ions. The calculated D values for the chloride ions are a little higher than the experimental value of 2.03 $\times 10^{-9}\ m^2/s$ (Wriggers et al., 1998). The deviations in the D values for the ions in the MD 9 and MD10 simulations can be attributed to the lack of proper ion-ion dipole interactions. Also the calcium ions have a low affinity towards water which might cause them to move around randomly in the bulk solution contrary to the manganese ions (Clapham, 1995). As stated above the water model by itself overshoots the experimental D values, although the D values calculated for the OG monomers in this work are at very low concentrations (Table 3-4). They might need a correction factor to compare the results with the experiments (Dixon et al., 2002; He and Niemeyer, 2003).

Improvement in the performance of the dynamic properties of the OG can be achieved from the new TIP4P-Ew water model with the application of Particle-Mesh Ewald summation method (Horn et al., 2004). The addition of a new fictitious site at the center on the bisector of the H-O-H angle improves the dynamic properties of the biomolecular systems over the previous three-site water models. It is currently being tested with these water potentials along with the charmm carbohydrate parameters (MacKerell et al., 1998) for the investigation of glycolipid properties. The only limitation to this model is the computation time which is longer than the one employed in this work because of an extra atom in the model (Horn et al., 2004). The preliminary glycolipid properties calculated from the TIP4P-Ew model for the diffusion coefficient of water are in excellent agreement with the experimental values. The D values of 2.4 and 2.5 $\times 10^{-9}\ m^2/s$ have been calculated with these potentials for the solvent molecules. The experimental value for the self-diffusion coefficient of water reported in the literature is $2.3 \times 10^{-9}\ m^2/s$ (Sen and Nilsson. 1999; Mark and Nilsson, 2002). In addition, the model has been also tuned to work efficiently with the protein force field parameters to reproduce the interfacial thermodynamic and kinetic properties.

6.6 Summary of pea lectin simulation

Pea (*pisum sativum*) is a legume lectin, recognized as glycoprotein of non-immunoglobulin nature that is capable of the specific recognition of glucose derivatives without altering their sugar covalent structure. Ten different molecular dynamic simulations were performed to investigate the structure and interaction properties of pea lectin with various concentrations of *n*-octyl-β-D-glucopyranoside (OG) monomers freely distributed inside the periodic solvent cell. The dynamical properties of the ligands (OG, Ca^{2+}, Mn^{2+} and Cl^{-}) and the solvent molecules were also reported.

The structural conformational flexibility of the lectin was conserved in all the simulations. Higher fluctuations were observed, as expected, from the side chains of the lectin residues compared to the backbone heavy atoms. The interaction of the side chains with the water molecules induces fluctuations in the lectin monomer. The dynamic regions in the lectin monomer were determined from the Temperature- or B-factor analysis. The important amino acid residues contributing largely to the protein dynamics were determined. The results were also compared to the crystallographic X-ray diffraction B-factor profiles.

The interaction energy analysis concludes that the lectin might have different binding sites to the ligand molecules in the solution. The globular structure of the lectin is best interpreted as adaptive structures which restructure according to the thermodynamic environment conditions. The simulation results state that the binding site was not the same in each simulation albeit there are one or two OG bindings observed at a time which are electrostatically favourable for the OG molecule to interact with those lectin surface amino acid residues. Several strong binding of OG to the pea lectin are apparent in the system containing only Ca^{2+} divalent ions. Interestingly, there was no binding observed in the simulation without any divalent ions. Furthermore, the lectin-ligand complex is stabilized by multiple hydrogen bonding and at least with one water bridge in the aqueous solution. The self-diffusion coefficients of water and the ligands (OG monomers and the charged ions) are compared with the previously published experimental and theoretical data. The diffusion coefficients of water molecules were in good agreement with the TIP3P water model results. The ligand diffusion coefficients were also consistent with the available data reported in the literature.

7 Conclusions

Molecular dynamics simulations are applied extensively in the present study to characterize the glycoconjugate structure, dynamic and interaction properties. The molecular systems which comprise of glycolipids and carbohydrate binding protein are treated explicitly in the aqueous solvent environemnt. Application of these computational methods have already proved its indispensable role in many parts of the scientific perspectives and till date it remains an ultimate tool to explore the time dependent properties of the system, particularly the jiggling of biological macromolecules.

Investigation of structural and dynamical properties of the *n*-octyl-β-D-glucopyranoside (OG) detergents in the aqueous solution using MD simulations were presented. In addition to its monomer conformation several factors come into play when dealing with these self-assembled molecules in the solution at and above *cmc*. The build-in dual character in their conformation towards water is the specialty of such detergents. The critical packing parameter determined from the MD simulations state that due to the geometric constraints of the glucose head group area and hydrocarbon chain length, the micelle formed is not likely to have a spherical shape, even at the concentrations close to the *cmc*. A packing factor of above 0.50 was calculated from the simulations and experiment.

The dynamic shape of the micelle was analyzed from the principle moment of inertia ratios. Through these inertia ratios accurate shape transformations of the micelle in the solution were studied, from which it was discovered that the OG micelle was most often in a prolate ellipsoidal form with two similar and one smaller moment of inertia ratios. An interesting behaviour has also been observed from the 11 ns simulation. The micelle tends to generate in a small bilayer or cylindrical rod form at some infrequent intervals where large deviations in the two principal moment of inertia ratios occurred, as compared to the third lowest ratio. In order to characterize such dynamic events (i.e. micelle shape fluctuations which arise due to the governing thermodynamic conditions, geometric constraints, different initial coordinates and velocities) completely, the MD simulations have to be carried out for very long time scales and with many different initial coordinates, which could be computationally taxing.

Glycolipid simulations initiated from the different coordinates and velocities conclude that the micelle size (radius of gyration, R_g) and the accessible surface

areas (ASA) were remained constant except at one instant (around 9 ns) in the bog1 simulation where large deviations in the principal moment of inertia ratios of up to 2.0 were observed. Nevertheless the OG micelle studied is small and short ranged; no ordering of the micelle to bilayer took place in the simulated concentration region. The R_g calculated from the simulations was consistent with some experimental and theoretical results. Analyzing the micelle aggregate surface states that major contribution to the total ASA of the OG micelle comes from the glucose head group in contrast to the hydrocarbon chain due to the roughness of the surface and also the elongation of the micelle in the solution. The radial distribution functions constructed for different atoms of the monomer confirm that the hydroxyl oxygen atoms at the hydrophilic/hydrophobic interface are more strongly hydrated than the acetal oxygen, ring oxygen, and anomeric carbon atoms, which face outward towards the aqueous environment. These outer atoms were primarily involved in the hydrogen bonding with the solvent molecules. Diffusion coefficients for the OG micelle were estimated through the modified empirical correlation. The estimated values calculated with the structural parameters obtained from the MD simulations are consistent with the experimental results. The applied correlation to the structural parameters of other works from the literature concludes that the proposed correlation is quite reasonable to apply in the isotropic concentration region $cmc < D_{OG} < 2.0$ M.

With the established knowledge of OG simulations in the initial phase, further work has been carried out to explore the properties of the lectin system accompanied with OG monomers. The pea lectin (*pisum sativum*) is a carbohydrate binding protein studied with various combinations of ligands in an explicit solvent environment. Besides, interaction energy analysis of the pea lectin-OG systems, the structure and dynamics of the lectin and the individual ligands were also investigated. This will facilitate the understanding of the role of the glycolipids with in lectin macromolecular systems and will be also very useful for tracing the complex interactions involved in the solubilization of proteins with the biological detergents. Because there was no prior results reported on the specific interaction or binding analysis of pea lectin, the present investigations addressed these issues from the MD simulations in a more realistic way in this manner: fully solvated lectin monomer applied with the periodic boundary conditions, different combinations of OG monomers, Ca^{2+}, Mn^{2+} and Cl^- ions, none of the molecular components in the cell were constrained during the simulations, started from different initial coordinates and random velocities, as well as higher non-bonded cutoffs with the Particle-Mesh Ewald summation method for treating the electrostatic interactions. The approached followed in

solving such systems seems to provide more information and clues about the biomolecule structure and function than the single long simulation performed with either unconstrained or partly constrained molecular entities.

The conformational flexibility of the pea lectin over time in the aqueous solution was conserved in all the simulations. The RMS deviation in one simulation with two OG monomers and divalent ions were higher in the middle of the trajectory, albeit the deviations in this system was decreased after few hundred picoseconds once the lectin structure in the solution was fully relaxed. Despite, of the independent MD runs started from the different initial conditions the global conformations of the lectin have fairly converged to the stable state in all of the simulations. The size of the pea lectin determined from the radius of gyration reached constant at the early stage and remains steady throughout the simulations. The RMS fluctuations per residue in the lectin backbone and the side chains with reference to the average solution structure from the trajectories were also maintained in the simulations with the latter contributing largely to the overall structure. The dynamics in the lectin residues were calculated from the B-factor distribution and compared to the crystal structure derived from the X-ray diffraction studies. Most of the peaks observed in the crystal structure results were also retained in the simulations which indicate the quality of the models used in the simulations. Few regions in the MD simulations show higher peaks than the crystal structure B-factor results. The cyclic residues and the shallow regions of the lectin at the surface leads to larger dynamics in the aqueous solution which is not well characterized it the crystal structure analysis.

Different binding sites were observed in the simulations through the interaction energy analysis. Each OG monomer interaction with the amino acid residues were calculated from the dynamic trajectories and found that the binding was facilitated by the favourable electrostatic interactions between the lectin and OG monomer. These forces are the primary driving force (e.g. electrostatic, van der Waals and hydrogen bond interactions) for the stable ligand-lectin complex formed in the solution. This proves that the lock and key mechanism for the pea lectin with the OG monomer was not appropriate. The local conformation of the lectin might flexibly adapt to the surrounding environment and accommodates the visiting ligand molecule. The binding process in such systems might work close to the induced-fit principle. To explore these interaction effects elaborately in the aqueous solution for such large lectin complexes with the detergents can be visualized in terms of free energy calculations. The full free energy calculation for these systems would be a major undertaking, often with prior experience and

knowledge about the system under investigation. Interestingly, the simulation with 20 OG monomers shows the binding site very close to that of the crystal structure site reported from the X-ray diffraction experiment. The OG-16 monomer head group interacts strongly with the lectin residues Tyr-100, Leu-101, Ala-107, Glu-108 and Lye-145. The Ca^{2+} divalent ion was also found near to the carbohydrate binding site.

Multiple enhanced bindings of OG to the lectin monomer were noticed in the simulation with only Ca^{2+} divalent ions. This indicates that the presence of Ca^{2+} ions (without Mn^{2+} ions) in the solution simply binds to the lectin. As evident from the simulations with both divalent ions, it seems the inclusion of Mn^{2+} divalent ions compensates the effect of multiple bindings of OG to the lectin. Intriguingly, the MD simulation without any charged ions in the solution reflects no binding of the OG onto the lectin monomer. This substantiates the importance of divalent ions for the ligand-lectin binding processes. Since they are the essential part of most leguminous lectins it is required for the biological activity of the lectins. These results also support the performance of charmm parameters employed in the simulations to analyze such specific effects.

The inherent nature of detergents self-assembling to molecular aggregates (small micelle) was apparent in the simulation with 20 OG monomers. About 9-10 monomers in the solution are joined together at the lectin surface and started to build a micelle with the hydrocarbon chains facing inside the micelle along with the lectin hydrophobic side chain residues, thus these OG tails are avoiding the rest of the aqueous polar environment. The OG-8 monomer head group was pushed into the proton cloud of Thr-192 and Asp-195 lectin residues with the increase in the electrostatic interaction energy and at the same time holding the tail tightly with the high hydrophobic side chain residues Val-172, Leu-173 and Val-196 of the lectin. The center of geometry of the OG monomer tails involved in the micelle are analyzed based on their distances with respect to the OG-8 monomer (foundation for the micelle building process) before and after the formation of the micelle from the dynamic trajectories. It was found that the distances of these selected monomers were decreased after the micelle was formed. Moreover, the decreased distances between OG-8 and the monomers in the aggregate were stable after the formation of the micelle.

The surface of the lectin was characterized by the solvent accessible surface areas calculated for the different molecular components in the lectin. A probe sphere of a radius equivalent to the water molecule was rolled over the lectin's molecular

surface. These results indicated that the surface is loose and has plenty of cavities where it can interact effectively with the solvent molecules. The major contribution to this comes from the β-side chains where most of these residues are in direct contact with the water. The lectin residue staying next to the water molecules participates in hydrogen bonding/water bridges and thus stabilizes the overall structure of the lectin monomer. Improved hydrogen bonding pattern and water bridges between the lectin and the OG were also identified from the simulations performed in a large solvent periodic box.

Finally, the diffusion coefficients calculated for the water molecules are in good agreement with the standard diffusion coefficient values reported for the TIP3P water model. The average D value of $5.4\times10^{-9}\,\mathrm{m}^2/\mathrm{s}$ was calculated for the solvent molecules in the present simulations which is comparable to the value of $5.2\times10^{-9}\,\mathrm{m}^2/\mathrm{s}$ reported using this water model. In addition, the diffusion coefficients for the individual OG monomers in the solution are calculated and averaged over the total number of OG monomers in each simulation, respectively. Comparing the average D values for the OG monomers from each independent simulation show good correlation in their average D values for the OG monomers. The diffusion coefficients calculated for the charged ions were also consistent with the published results. The pea lectin and OG properties reported in this work concludes that the molecular models were highly realistic and the force field parameters applied were appropriate to explain the phenomenological behaviour observed from the experiments for these macromolecules.

Bibliography

Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Oxford University Press, Oxford, UK.

Allinger, N. L., R. A. Kok, and M. R. Imam. 1988. Hydrogen bonding in MM2. J. Comp. Chem. 9:591-595.

Aqvist, J., and S. L. Mowbray. 1995. Sugar recognition by a glucose/galactose receptor. Evaluation of binding energetics from molecular dynamics simulations. J. Biol. Chem. 270:9978-9981.

Bandyopadhyay, S., M. L. Klein, G. J. Martyna, and M. Tarek. 1998. Molecular dynamics studies of the hexagonal mesophase of sodium dodecylsulphate in aqueous solution. Mol. Phys. 95:377-384.

Beglov, D., and B. Roux. 1994. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. J. Chem. Phys. 100:9050-9063.

Birdwell, C. R., and J. H. Strauss, 1973. Agglutination of sindbis virus and of cells infected with sindbis virus by plant lectins. Journal of Virology. 11:502-507.

Bogusz, S., R. M. Venable, and R. W. Pastor. 2000. Molecular dynamics simulations of octyl glucoside micelles: Structural properties. J. Phys. Chem. B. 104:5462-5470.

Bogusz, S., R. M. Venable, and R. W. Pastor. 2001. Molecular dynamics simulations of octyl glucoside micelles: Dynamic properties. J. Phys. Chem. B. 105:8312-8321.

Boyd, W. C., and E. Shapleigh. 1954. Separation of individuals of any blood group into secreators and non-secretors by use of plant agglutinin (lectin). Blood. 9:1195-1198.

Boyd, W.C., and E. Shapleigh. 1954. Specific precipitating activity of plant agglutinins (lectins). Science. 119:419-427.

Bradbrook, G. M., J. R. Forshaw, and S. Perez. 2000. Structure/thermodynamics relationships of lectin-saccharide complexes. The erythrina corallodendron case. Eur. J. Biochem. 267:4545-4555.

Bridges, C. D., and S. L. Fong. 1979. Different receptors for distribution of peanut and ricin agglutinins between inner and outer segments of rod cells. Nature. 282:513-515.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4:187-217.

Bruce, C. D., M. L. Berkowitz, L. Perera, and M. D. E. Forbes. 2002. Molecular dynamics simulation of sodium dodecyl sulfate micelle in water: Micellar structural characteristics and counterion distribution. J. Phys. Chem. B. 106:3788-3793.

Bruce, C. D., S. Senapati, M. L. Berkowitz, L. Perera, and M. D. E. Forbes. 2002. Molecular dynamics simulations of sodium dodecyl sulfate micelle in water: The behavior of water. J. Phys. Chem. B. 106:10902-10907.

Bryce, R. A., I. H. Hillier, and J. H. Naismith. 2001. Carbohydrate-protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. Biophys. J. 81:1373-1388.

Cheng A., M. Merz Jr. 1999. Application of a multiple time step algorithm to biomolecular systems. J. Phys. Chem. B. 103:5396-5405.

Chong, L. T., Y. Duan, L. Wang, I. Massova, and P. A. Kollman. 1999. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. Proc. Natl. Acad. Sci. U. S. A. 96:14330-14335.

Clapham, D. E. 1995. Calcium signaling. Cell. 80:259-268.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. J. Am. Chem. Soc. 117: 5179-5197.

Cortes, A. B., and M. Valiente. 2003. The effect of a minimum amount of octyl-β-D-glucoside on micellar, nematic, and hexagonal phases of the CTAB/glycerol/water system. Colloid. Polym. Sci. 281:319-324.

Cunningham, B. A., J. L. Wang, M. J. Waxdal, and G. M. Edelman. 1975. The covalent and three dimensional structure of concanavalin A II. Amino acid sequence of cyanogen bromide fragment F_3. J. Biol. Chem. 250:1503-1512.

D'Aprano, A., R. Giordano, M. P. Jannelli, S. Magazu, G. Maisano, and B. Sesta. 1996. QELS and SANS studies of octyl-β-glucoside micellar solutions. J. Mol. Struct. 383:177-182.

Danielsson, I. 1963. Thermodynamic aspects of micelle formation. Finska Kemists. Medd. 72:90-98.

Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. J. Chem. Phys. 98:10089–10092.

Dauber-Osguthorpe, P., V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler. 1988. Structure and energetics of ligand binding to proteins: E. coli dihydrofolate reductase- trimethoprim, a drug-receptor system. Proteins: Structure, Function and Genetics. 4:31-47.

Debray, H., D. Decout, G. Strecker, G. Spik, and J. Montreuil. 1981. Specificity of twelve lectin towards oligosaccharides and glycopeptides related to N-glycosylproteins. Eur. J. Biochem. 117:41-55.

Derewenda, Z., J. Yariv, J. R. Helliwell, A. J. Kalb, E. J. Dodson, M. Z. Papiz, T. Wan, and J. Campbell. 1989. The structure of the saccharide-binding site of concanavalin A. EMBO J. 8:2189-2193.

Dixon, A. M., R. M. Venable, R. W. Pastor, and T. E. Bull. 2002. Micelle-bound conformation of a hairpin-forming peptide: Combined NMR and molecular dynamics study. Biopolymers. 65:284-298.

Duan, Y., and P. A. Kollman. 2001. Computational protein folding: From lattice to all-atom. IBM Syst. J. 40:297-309.

Durchschlag, H., and P. Zipper. 1997. Prediction of hydrodynamic parameters of biopolymers from small angle scattering data. J. Appl. Cryst. 30 :1112-1124.

Durell, S. R., B. R. Brooks, and A. Ben-Naim. 1994. Solvent-induced forces between two hydrophilic groups. J. Phys. Chem. 98:2198-2202.

Einspahr, H., E. H. Parks, K. Suguna, E. Subramanian, and F. L. Suddath. 1986. The crystal structure of pea lectin at 3.0-Å resolution. J. Biol. Chem. 261:16518-16527.

Eisele, J.-L., and J. P. Rosenbusch. 1989. Crystallization of porin using short chain phospholipids. J. Mol. Biol. 206:209-212.

Elemans, J. A. A. W., A. E. Rowan, and R. J. M. Nolte. 2003. Mastering molecular matter. Supramolecular architectures by hierarchical self-assembly. J. Mat. Chem. 13:2661-2670.

Esselink, K., P. A. J. Hilbers, N. M. van Os, B. Smit, and S. Karaborni. 1994. Molecular dynamics simulations of model oil/water/surfactant systems. Colloids Surf. A. 91:155-167.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. J. Pedersen. 1995. A smooth particle mesh ewald method. J. Chem. Phys. 103:8577-8593.

Faraldo-Gomez, J. D., G. R. Smith and M. S. P. Sansom. 2003. Molecular dynamics simulations of the bacterial outer membrane protein FhuA: a comparative study of the ferrichrome-free and bound states. Biophys. J. 85:1406-1420.

Feller, S. E., R. W. Pastor, A. Rojnuckarin, S. Bogusz, and B. R. Brooks. 1996. Effect of electrostatic force truncation on interfacial and transport properties of water. J. Phys. Chem. 100:17011-17020.

Focher, B., G. Savelli, G. Torri, G. Vecchio, D. C. McKenzie, D. F. Nicoli, and C.A. Bunton. 1989. Micelles of 1-alkyl glucoside and maltoside: anomeric effects on structure and induced chirality. Chem. Phys. Lett. 158:491-494.

Fogolari, F., A. Brigo, and H. M. 2003. Protocol for MM/PBSA molecular dynamics simulations of proteins. Biophys. J. 85:159-166.

Gabius, H. -J. 2001. Probing the cons and pros of lectin-induced immunomodulation: case studies for the mistletoe lectin and galectin-1. Biochimie. 83:659-666.

Gabius, H.-J. 2000. Biological information transfer beyond the genetic code: the sugar code. Naturwissenschaften 87:108-121.

Gao, X., and T. C. Wong. 2001. Molecular dynamics simulation of adrenocorticotropin (1-10) peptide in a solvated deodecylphosphocholine micelle. Biopolymers. 58:643-659.

Garavito, R. M., and J. P. Rosenbusch. 1986. Isolation and crystallization of bacterial porin. Meth. Enzymol. 125:318-328.

Garavito, R. M., D. Picot, and P. J. Loll. 1996. Strategies for crystallizing membrane proteins. J. Bioenerg. Biomembr. 28:13-26.

Gerini, M. F., D. Roccatano, E. Baciocchi, and A. Di Nola. 2003. Molecular dynamics simulations of lignin peroxidase in solution. Biophys. J. 84:3883-3893.

Gould, R. J., B. H. Ginsberg, and A. A. Spector. 1981. Effect of octyl β-glucoside on insulin binding to solubilized membrane receptors. Biochemistry. 20:6776-6781.

Goyal, P. S., and V. K. Aswal. 2001. Micellar structure and inter-micelle interactions in micellar solutions: Results of small angle neutron scattering studies. Current Science. 80:972-979.

Grigera, J. R. 2002. Molecular dynamics simulation of ligand-receptor studies. Carbohydrates interaction in aqueous solutions. Curr. Pharm. Design. 8:1579-1604.

Guimaraes, C. R. W., G. Barreiro, C. A. F. de Oliveira and R. B. de Alencastro. 2004. On the application of simple explicit water models to the simulations of biomolecules. Braz. J. Phys. 34:126-136.

Gupta, D., T. K. Dam, S. Oscarson, and C. F. Brewer. 1997. Thermodynamics of lectin-carbohydrate interactions. Binding of the core trimannoside of asparagine-linked carbohydrates and deoxy analogs to concanavlin A. J. Biol. Chem. 272:6388-6392.

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

Hagler, A. T., and C. S. Ewig. 1994. On the use of quantum energy surfaces in the derivation of molecular force fields. Comp. Phys. Comm. 84:131-155.

Haile, J. M. 1992. Molecular Dynamics Simulations: Elementary Methods. John Wiley & Sons, Inc. New York, 1992.

Halgren, T. A. 1996. Merck Molecular Force Field. III. Molecular geometrics and vibrational frequencies for MMFF94. J. Comp. Chem. 17:553-586.

Halgren, T. A. 1996. Merck Molecular Force Field. V. Extension of MMFF94 using experimental data, additional computational data and empirical rules. J. Comp. Chem. 17:616-641.

Halgren, T. A., and R. B. Nachbar. 1996. Merck Molecular Force Field. IV. Conformational energies and geometries. J. Comp. Chem. 17:587-615.

Hamelryck, T. W., R. Loris, J. Bouckaert, and L. Wyns. 1998. Properties and structure of the legume lectin family. Trends Glycosci. Glycobiol. 10:349-404.

Hayryan, S., C.-K. Hu, J. Skrivanek, E. Hayryan, and I. Pokorny. 2005. A new analytical method for computing solvent-accessible surface area of macromolecules and its gradients. J. Comput. Chem. 26:334-343.

He, L.-Z., and B. Niemeyer. 2003. A novel correlation for protein diffusion coefficients based on molecular weight and radius of gyration. Biotechnol. Porgr. 19:544-548.

He, L.-Z., S. Andre, H.-C. Siebert, H. Helmholz, B. Niemeyer, and H.-J. Gabius. 2003. Detection of ligand- and solvent-induced shape alterations of cell-growth-

regulatory human lectin galectin-1 in solution by small angle neutron and x-ray scattering. Biophys. J. 85:511-524.

He, L.-Z., V. Garamus, B. Niemeyer, H. Helmholz, and R. Willumeit. 2000. Determination of micelle structure of octyl-β-glucoside in aqueous solution by small angle neutron scattering and geometric analysis. J. Mol. Liq. 89:239-249.

He, L.-Z., V. Garamus, S. S. Funari, M. Malfois, R. Willumeit, and B. Niemeyer. 2002. Comparison of small-angle scattering methods for the structural analysis of octyl-β-maltopyranoside micelles. J. Phys. Chem. B. 106:7596-7604.

Heermann, D. W. 1986. Computer simulation methods: in Theoretical Physics. Springer-Verlag, Heidelberg, Germany.

Hermans, J., H. J. C. Berendsen, W. F. van Gunsteren, and J. P. M Postma. 1984. A consistent empirical potential for water-protein interactions. Biopolymers 23:1513-1518.

Hess B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. J. Comp. Chem. 18:1463-1472.

Higgins, T. J. V., P. M. Chandler, G. Zurawski, S. C. Button, and D. Spencer. 1983. The biosynthesis and primary structure of pea seed lectin. J. Biol. Chem. 258:9544-9549.

Hoover, W.G. 1985. Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A. 31:1695-1697.

Horn, H. W., W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon. 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J. Chem. Phys. 120:9665-9678.

Israelachvili, J. N. 1998. Intermolecular and Surface Forces, 2nd edition, Academic Press, London.

Jang, H., P. S. Crozier, M. J. Stevens, and T. B. Woolf. 2004. How environment supports a state: molecular dynamics simulations of two states in bacteriorhodopsin suggest lipid and water compensation. Biophys. J. 87:129-145.

Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS potential functions for proteins. energy minimization for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110:1657-1666.

Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. 118:11225-11236.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79: 926–935.

Karaborni, S., N. M. van Os, K. Esselink, and P. A. J. Hilbers. 1993. Molecular dynamics simulations of oil solubilization in surfactant solutions. Langmuir. 9:1175-1178.

Karplus, M., and J. A. McCammon. 2002. Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. 9:646-652.

Kawai, S., and I. Takeuchi. 1976. Concanavalin A-induced agglutination and binding of con A to the differentiating cells of dictyostelium dscoideum. Dev. Growth Differ. 18:311-317.

Kiwada, H., H. Nimura, Y. Fujisaki, S. Yamada, and Y. Kato. 1985. Application of synthetic alkyl glycoside vesicles as drug carriers. I. Preparation and physical properties. Chem. Pharm. Bull. 33:753-759.

Konidala, P., L.-Z. He, and B. Niemeyer. 2005. Molecular dynamics characterization of n-octyl-β-D-glucopyranoside micelle structure in aqueous solution. J. Mol. Graphics Modell. In press.

Krammer, A., P. D. Kirchhoff, X. Jiang, C. M. Venkatachalam, and M. Waldman. 2005. Ligscore: a novel scoring function for predicting binding affinities. J. Mol. Graphics Modell. 23:395-407.

Kraütler V., W. F. Van Gunsteren, and P. H. Hüneberger. 2001. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J. Comp. Chem. 22:501-508.

Kuhn, H., and H. Rehage. 1997. The molecular structure of sodium octanoate micelles studied by molecular dynamics computer experiments. Ber. Bunsenges. Phys. Chem. 101:1485-1492.

Kuhn, H., and H. Rehage. 1997. The molecular structure of sodium octanoate micelles studies by molecular dynamics computer experiments. Ber. Bunsenges. Phys. Chem. 101:1485-1492.

Kuhn, H., B. Breitzke, and H. Rehage. 1998. The phenomenon of water penetration into sodium octanoate micelles studied by molecular dynamics computer simulation. Colloid. Polym. Sci. 276:824-832.

Kuhn, H., B. Breitzke, and H. Rehage. 2002. A molecular modelling study of pentanol solubilized in a sodium octanoate micelle. J. Colloid Interface Sci. 249:152-161.

Kumar, S., and R. Nussinov. 2002. Close-range electrostatic interactions in proteins. ChemBioChem. 3:604-617.

Kuttel, M., J. W. Brady, and K. J. Naidoo. 2002. Carbohydrate solution simulations: Producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. J. Comput. Chem. 23:1236-1243.

La Mesa, C., A. Bonincontro, and B. Sesta. 1993. Solution properties of octyl β-D-glucoside. Part 1: Aggregate size, shape and hydration. Colloid. Polym. Sci. 271:1165-1171.

Laaksonen, L., and J. B. Rosenholm. 1993. Molecular dynamics simulations of the water/octanoate interface in the presence of micelles. Chem. Phys. Lett. 216:429-434.

Laczko-Hollosi, I., M. Hollosi, V. M. Lee, and H. H. Mantsch. 1992. Conformational change of a synthetic amyloid analogue des[Ala21,30]A42 upon binding to octyl glucoside micelles. Eur. Biophys. J. 21:345-348.

Laidler, K. J., and J. H. Meiser. 1982. Physical Chemistry. The Benjamin/ Cummings Publishing Company, Inc. California, USA.

Lauterwein, J., C. Bösch, L. R. Brown, and K. Wüthrich. 1979. Physicochemical studies of the protein-lipid interactions in melittin-containing micelles. Biochim. Biophys. Acta. 556:244–264.

Lazaridis, T. 2002. Binding affinity and specificity from computational studies. Curr. Org. Chem. 6:1319-1332.

le Maire, M., P. Champeil, and J. V. Moller. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. Biochim. Biophys. Acta. 1508:86-111.

Leach, A. R. 2001. Molecular Modelling: Principles and Applications. 2nd edition, Pearson Education Limited, UK.

Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55:379-400.

Lee, W. C., and K. H. Lee. 2004. Applications of affinity chromatography in proteomics. Anal. Biochem. 324:1-10.

Li, Z., and T. Lazaridis. 2005. The effect of water displacement on binding thermodynamics: concanavalin A. J. Phys. Chem. B. 109:662-670.

Lii, J-H., and N. L. Allinger. 1991. The MM3 force field for amides, polypeptides and proteins. J. Comp. Chem. 12:186-199.

Lii, J-H., and N. L. Allinger. 1998. Directional hydrogen bonding in the MM3 force field. II. J. Comp. Chem. 19:1001-1016.

Lii, J-H., S. Gallion, C. Bender, H. Wikstrom, N. L. Allinger, K. M. Flurchick, and M. M. Teeter. 1989. Molecular Mechanics (MM2): Calculations on peptides and on the protein crambin using the cyber 205. J. Comp. Chem. 10:503-513.

Lorber, B., J. B. Bishop, and L. J. DeLucas. 1990. Purification of octyl beta-D-glucopyranoside and re-estimation of its micellar size. Biochim. Biophys. Acta. 1023:254-265.

Ma, B., S. Kumar, C.-J. Tsai, H. Wolfson, N. Sinha, and R. Nussinov. 2001. Protein-ligand interactions: induced fit. In Robertson, S. Editor. Encyclopedia of Life Sciences, London. 1-8.

MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics. Studies of proteins. J. Phys. Chem. B. 102:3586-3616.

Mark, P., and L. Nilsson. 2002. A molecular dynamics study of tryptophan in water. J. Phys. Chem. B. 106:9440-9445.

Marrink, S. J., D. P. Tieleman, and A. E. Mark. 2000. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. J. Phys. Chem. B. 104:12165-12173.

Martyna G. J., M. E. Tuckerman, D. J. Tobias, and M. L. Klein. 1996. Explicit reversible intergators for extended systems dynamics. Mol. Phys. 87:1117-1157.

Matteoli, E., and G. A. Mansoori. 1995. A simple expression for radial distribution functions of pure fluids and mixtures. J. Chem. Phys. 103:4672-4677.

McCammon, J. A., B. R. Gelin, and M. Karplus. 1977. Dynamics of folded proteins. Nature. 267:585-590.

McQuarrie, D. A. 1983. Quantum Chemistry. Oxford University Press, Oxford, UK.

Meehan, E. J., J. McDuffie, H. Einspahr, C. E. Bugg, and F. L. Suddath. 1982. The crystal structure of pea lectin at 6-Å resolution. J. Biol. Chem. 257:13278-13282.

Melchonna, S., G. Ciccotti, and B. L. Holian. 1993. Hoover NPT dynamics for systems varying in shape and size. Mol. Phys. 78:533-544.

Miyamoto S., and P. A. Kollman. 1992. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water molecules. J. Comp. Chem. 18:1463-1472.

Momany, F. A., R. Rone. 1992. Validation of the general purpose QUANTA 3.2 / CHARMm force field. J. Comp. Chem. 13:888-900.

Moura, A. F., and L. C. G. Freitas. 2004. Molecular dynamics simulation of the sodium octanoate micelle in aqueous solution: Comparison of force field parameters and molecular topology effects on the micellar structure. Braz. J. Phys. 34:64-72.

Naidoo, K. J., and J. W. Brady. 1997. The application of simulated annealing to the conformational analysis of disaccharides. Chem. Phys. 224:263-273.

Nelson, P. H., G. C. Rutledge, and T. A. Hatton. 1997. On the size and shape of self-assembled micelles. J. Chem. Phys. 107:10777-10781.

Neumann, D., O. Kohlbacher, H.-P. Lenhof, and C.-M. Lehr. 2002. Lectin-sugar interaction: calculated versus experimental binding energies. Eur. J. Biochem. 269:1518-1524.

Nieto-Draghi, C., J. B. Avalos, and B. Rousseau. 2003. Dynamic and structural behavior of different rigid nonpolarizable models of water. J. Chem. Phys. 118:7954-7964.

Nilsson, F., O. Soederman, and I. Johansson. 1996. Physical-chemical properties of the *n*-octyl β-D-glucoside/water system. A phase diagram, self-diffusion NMR, and SAXS study. Langmuir. 12:902-908.

Ostermeier, C., and H. Michel. 1997. Crystallization of membrane proteins. Curr. Opin. Struct. Biol. 7:697-701.

Ott, K-H., and B. Meyer. 1996. Parametrization of GROMOS force field for oligosaccharides and assessment of efficiency of molecular dynamics simulations. J. Comp. Chem. 17:1068-1084.

Palma, R., M. E. Himmel, G. Liang, and J. W. Brady. 2001. Molecular mechanics studies of cellulases, in: Glycosyl hydrolases in biomass conversion, M. E. Himmel, editor. American Chemical Society, Washington DC.

Patist, A., J. R. Kanicky, P. K. Shukla, and D. O. Shah. 2002. Importance of micellar kinetics in relation to technological processes. J. Colloid Interface Sci. 245:1- 15.

Patthi, S., M. Akong, and G. Velicelebi. 1987. Hydrodynamic characterization of vasoactive intestinal peptide receptors extracted from rat lung membranes in triton X-100 and *n*-octyl- β-D-glucopyranoside. J. Biol. Chem. 262:15740-15745.

Pearlman, D. A., D. A. Case, J. C. Caldwell, G. L. Seibel, U. C. Singh, P. Weiner, and P. A. Kollman. 1991. AMBER 4.0, University of California, San Francisco.

Perram, J. W., H. G. Petersen, and S. W. de Leeuw. 1988. An algorithm for the simulation of condensed matter which grows as the 3/2 power of the number of particles. Mol. Phys. 65:875–893.

Petersen. H. G. 1995. Accuracy and efficiency of the particle mesh Ewald method. J. Chem. Phys. 103:3668–3679.

Pletnev, V. Z., S. N. Ruzheinikov, I. N. Tsygannik, I. Y. Mikhailova, W. Duax, D. Gosh, and W. Pangborn. 1997. The structure of Pea lectin-D-glucopyranose complex at a 1.9 Ǻ resolution. Russ. J. Bioorg. Chem. 23:436-445.

Poveda, A., and J. Jimenez-Barbero. 1998. NMR studies of carbohydrate-protein interactions in solution. Chem. Soc. Rev. 27:133-143.

Rahman, A., and F. H. Stillinger. 1971. Molecular dynamics study of liquid water. J. Chem. Phys. 55:3336-3359.

Reiling, S., M. Schlenkrich, and J. Brickmann. 1996. Forcefield parameters for carbohydrates. J. Comput. Chem. 17:450-468.

Reyes, C. M., R. Nifosi, A. D. Frankel, and P. A. Kollman. 2001. Molecular dynamics and binding specificity analysis of the bovine immunodeficiency virus BIV Tat-TAR complex. Biophys. J. 80:2833-2842.

Richards, F. M. 1977. Areas, volumes, packing, and protein structure. Annu. Rev. Biophys. Bioeng. 6:151-176.

Rin, J. M., K. D. Hardman, H. Einspahr, F. L. Suddath, and J. P. Carver. 1993. X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 Ǻ resolution. J. Biol. Chem. 268:10126-10132.

Ryckaert J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys. 23:327-341.

Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of N-alkanes. J. Comput. Phys. 23:327-341.

Saiz, L., and M. L. Klein. 2002. Computer simulation studies of model biological membranes. Acc. Chem. Res. 35:482-489.

Schwarz, F. P., K. D. Puri, R. G. Bhat, and A. Surolia. 1993. Thermodynamics of monosaccharide binding to concanavalin A, pea (Pisum sativum) lectin, and Lentil (Lens culinaris) lectin. J. Biol. Chem. 268:7668-7677.

Schwarz, F. P., S. Misquith, and A. Surolia. 1996. Effect of substituent on the thermodynamics of D-glucopyranoside binding to concanavalin A, pea (Pisum sativum) and lentil (Lens culinaris) lectin. Biochem. J. 316:123-129.

Sen, S., and L. Nilsson. 1999. Free energy calculations and molecular dynamics simulations of wild-type and variants of the DNA-EcoRI complex. Biophys. J. 77:1801-1810.

Sen, S., and L. Nilsson. 1999. Structure, interaction, dynamics and solvent effects on the DNA-EcoRI complex in aqueous solution from molecular dynamic simulation. Biophys. J. 77:1782-1800.

Sharon, N., and H. Lis, 1972. Lectins: cell-agglutinating and sugar-specific proteins. Science. 177:949–959.

Sharon, N., and H. Lis. 1990. Legume lectins. FASEB J. 4:3198-3208.

Sharon, N., and H. Lis. 2004. History of lectin: from hemagglutinins to biological recognition molecules. Glycobiology. 14:53-62.

Shinoda, K., T. Yamanaka, and K. Kinoshita. 1959. Surface chemical properties in aqueous solutions of non-ionic surfactants: octyl glycol ether, α-octyl glyceryl ether and octyl glucoside. J. Phys. Chem. 63:648-650.

Smit, B., P. A. J. Hilbers, K. Esselink, L. A. M. Rupert, N. M. van Os, and A. G. Schlijper. 1990. Computer simulations of a water/oil interface in the presence of micelles. Nature. 348:624-625.

Straathof, A. J. J., H. van Bekkum, and A. P. G. Kieboom. 1988. Efficient preparation of octyl α-D-glucopyranoside monohydrate: A recirculation procedure involving water removal by product crystallization. Starch/Stärke. 40:229-234.

Straathof, A. J. J., H. van Bekkum, and A. P. G. Kieboom. 1988. Solid state and solution properties of octyl D-glucopyranosides. Starch/Stärke. 40:438-440.

Stubbs, G. W., H. G. Smith, Jr. and B. J. Litman. 1976. Alkyl glucosides as effective solubilizing agents for bovine rhodopsin. A comparison with several commonly used detergents. Biochim. Biophys. Acta. 425:46-56.

Sun, H. 1998. COMPASS: An ab initio force field optimized for condensed-phase applications, overview with details on alkane and benzene compounds. J. Phys. Chem. B. 102:7338-7364.

Tanford, C. 1980. The Hydrophobic Effect: Formation of Micelles and Biological Membranes. 2nd edition, John Wiley & Sons, New York.

Tavernelli, I., S. Cotesta, and E. E. Di lorio. 2003. Protein dynamics, thermal stability, and free-energy landscapes: a molecular dynamics investigation. Biophys. J. 85:2641-2649.

Thiesen, P. H., H. Rosenfeld, P. Konidala, V. M. Garamus, L.-Z. He. A. Prange, and B. Niemeyer. 2006. Glycolipids from a colloid point of view. J. Biotechnol. revised.

Tieleman, D. P., D. van der Spoel, and H. J. C. Berendsen. 2000. Molecular dynamics simulations of dodecylphosphocholine micelles at three different

aggregate sizes: Micellar structure and chain relaxation. J. Phys. Chem. B. 104:6380-6388.

Tobias D.J., and C. L. Brooks. 1988. Molecular dynamics with internal coordinate constraints. J. Chem. Phys. 76:3271-3276.

Trowbridge, I. S. 1974. Isolation and chemical characterization of a mitogenic lectin from pisum sativum. J. Biol. Chem. 249:6004-6012.

Tuckerman M. E., B. J. Berne, and G. J. Martyna. 1992. Reversible multiple time scale molecular dynamics. J. Chem. Phys. 97:1990-2001.

Tuckerman M.E. 2000. Understanding modern molecular dynamics: techniques and applications. J. Phys. Chem. B. 104:159-178.

Tyn, M. T., and Gusek, T. D. 1990. Prediction of diffusion coefficients of proteins. Biotechnol. Bioeng. 35:327-338.

Wang, J. L., B. A. Cunningham, M. J. Waxdal, and G. M. Edelman. 1975. The covalent and three dimensional structure of concanavalin A I. Amino acid sequence of cyanogen bromide fragments F_1 and F_2. J. Biol. Chem. 250:1490-1502.

Watanabe, K., M. Ferrario, and M. L. Klein. 1988. Molecular dynamics study of sodium octanoate micelle in aqueous solution. J. Phys. Chem. 92:819-821.

Woo, H.-J., and B. Roux. 2005. Calculation of absolute protein-ligand binding free energy from computer simulations. Proc. Natl. Acad. Sci. U. S. A. 102:6825-6830.

Wriggers, W., E. Mehler, F. Pitici, H. Weinstein, and K. Schulten. 1998. Structure and dynamics of calmodulin in solution. Biophys. J. 74:1622-1639.

Wymore, T., and T. C. Wong. 1999. Molecular dynamics study of substance P peptides partitioned in a sodium dodecylsulfate micelle. Biophys. J. 76:1213-1227.

Yan, S., M. Wu, D. J. Patel, N. E. Geacintov, and S. Broyde. 2003. Simulating structural and thermodynamic properties of carcinogen-damaged DNA. Biophys. J. 84:2137-2148.

Yuan, Z., T. L. Bailey, and R. D. Teasdale. 2005. Prediction of protein B-factor profiles. Proteins: Struct. Funct. Bioinf. 58:905-912.

Notation

Roman notation

a, b, c	Primary cell length geometry of the crystal lattice, [Å]
a_h	Area of the glucose head group, [Å²]
a_i	Acceleration of atom i, [m/s^2]
a^*_α	Conjugate reciprocal vectors
A	Surface area of wall acting on the system, [Å²]
B_i	Temperature or B-factor, [Å²]
C_{cmc}	Concentration of OG at *cmc*, [M]
CP_f	Critical geometry packing parameter
C_{MD}	Concentration of the solvated OG, [M]
C_V	Heat capacity of the system
D	Diffusion coefficient, [m²/s]
D_{dip}	Dipole moment, [C Å]
e	Elementary charge, [1.60219×10^{-19} C]
e_s	Eccentricity in equation 4.8
E	Total energy of the system, [kcal/mol]
E_e	Electronic energy *wrt* fixed nuclear positions, [kcal/mol]
E_P	Potential energy of the system, [kcal/mol]
E_T	Total energy of a molecule, [kcal/mol]
E_{Ke}	Kinetic energy of an electron, [kcal/mol]
E_{KN}	Kinetic energy of nuclei, [kcal/mol]
$E_{External}$	Non-bond energy terms in the potential energy surface, [kcal/mol]
$E_{Internal}$	Bonded energy terms in the potential energy surface, [kcal/mol]
f_v	Maxwell-Boltzmann velocity probability distribution
F_i	Force exerted on atom i in MD simulation, [kcal/mol Å]
$\vec{F}_i^E$	External force exerted by the simulation cell walls, [kcal/mol Å]
$\vec{F}_{ij}^I$	Internal force due to interatomic interactions, [kcal/mol Å]
F_i^T	Total force acting on atom i in equation 2.45, [kcal/mol Å]
F(Q)	Discrete Fourier transform

$g(r)$	Radial distribution function or Pair correlation function
h	Planck's constant, [6.62618×10^{-34} [J s]
H	Hamiltonian operator, [kcal/mol]
H_e	Hamiltonian operator- without the kinetic energy of the nuclei, [kcal/mol]
H_n	Hamiltonian operator- with the kinetic energy of the nuclei, [kcal/mol]
I_{min}	Minimum principle moments of inertia, [amu Å²]
I_{max}	Maximum principle moments of inertia, [amu Å²]
I_1	Diagonalised major principle moments of inertia, [amu Å²]
I_2	Diagonalised intermediate principle moments of inertia, [amu Å²]
I_3	Diagonalised minor principle moments of inertia, [amu Å²]
$J(D_{dip}, \varepsilon')$	Surface correction term in Particle Mesh Ewald
k_B	Boltzmann's constant, [1.38066×10^{-23} J/K]
$\vec{k}$	Reciprocal space lattice vector
k_b	Bond force constant, [kcal/mol Å²]
k_θ	Angle force constant, [kcal/mol rad²]
k_ϕ	Dihedral force constant, [kcal/mol]
k_ω	Improper force constant, [kcal/mol rad²]
k_u	Non-bonded distance force constant for 1, 3 atoms, [kcal/mol Å²]
K	Kinetic energy in equation 2.50, [kcal/mol]
l_c	Length of hydrocarbon chain, [Å]
L	Edge length of the primary cell
L_s	Length of the arc drawn on the molecular surface, [Å]
m_e	Rest mass of an electron, [9.10953×10^{-31} kg]
m_i	Mass of an atom i in MD simulation, [amu]
m_n	Mass of the nuclei, [1.675×10^{-27} kg]
m_p	Mass of a particle in equation 2.2, [9.10953×10^{-31} kg]
$\vec{m}_r$	Reciprocal lattice vector in equation 2.84
M	Number of time steps in the simulation in equation 2.41
MW	Molecular weight of a molecule, [amu]
MW_{mic}	Molecular weight of an OG micelle in MD simulations, [amu]
n	Multiplicity of the function

N	Total number of atoms in the system
P	Pressure of the system, [atm]
P_F	Pressure caused by the intermolecular forces, [atm]
P_M	Pressure caused by the movement of atoms, [atm]
P_O	Target pressure in Berendsen pressure control, [atm]
P_T	Total pressure expressed in terms of the momentum flux, [atm]
q_i, q_j	Charges of the atom i and atom j, [C]
Q	Thermodynamic partition function
r	Bond length in MD simulations (equation 2.29), [Å]
r_a, r_b	Distance between nuclei and the electron, [Å]
r_c	System center of mass in Nose-Hoover thermostat, [Å]
r_{cm}	Micelle center of mass in equation 4.2, [Å]
r_o	Ideal bond length, [Å]
r_{AB}	Distance between the nuclei, [Å]
r_i	Distance of the atom i from the origin in equation 4.2, [Å]
r_j	Position of atom j in equation 2.83, [Å]
r_i^{asa}	van der Waals radii of atom i in surface area calculations, [Å]
r_p^{asa}	Radius of the probe sphere in surface area calculations, [Å]
r_i^{md}	Position of atom i in the MD simulation structure, [Å]
r_i^r	Position of the atom i in the reference structure, [Å]
r_{ij}^t	Position of the atom i with respect to the trajectory frame j, [Å]
r_{ij}	Distance between two charges in equation 2.30, [Å]
R	Total radius of the molecular surface, [Å]
R_g	Radius of gyration, [Å]
R_H	Hydrodynamic radius of the micelle, [Å]
R_{min}	van der Waals distance at Lennard-Jones potential is zero, [Å]
$r_{i(0,0,0)}^{(0,0,0)}$	Primary frame vector
$R_\alpha^{(0,0,0)}$	Image frame vector
s_T	Velocities scale factor
$s_{\alpha j}$	Fractional coordinates of atom j in equation 2.84, [Å]
S	Entropy of the system, [kcal/K]
$S(m)$	Structure factor
S_W	Switching function

t	MD simulation time origin [ps]
T	Temperature of the system, [K]
u	Distance between 1, 3 atoms in MD simulations, [Å]
u_o	Ideal distance between 1, 3 atoms, [Å]
U_{eff}	Total Pair-wise potential energy, [kcal/mol]
U_P	Potential energy in the primary cell, [kcal/mol]
v	Velocity of an atom, [m/s]
v_c	Volume of the hydrocarbon chain, [Å³]
V	Volume of the simulation cell, [Å³]
W	Virial quantity defined in the Clausius virial theorem
Z	Charge number
Z_s	Perpendicular distance from the center of the sphere, [Å]

Greek Notation

α	Cell translation vector
α, β, γ	Primary cell angle geometry of the crystal lattice, [rad]
β_c	Convergence parameter in Particle Mesh Ewald
δ	Phase shift, [rad]
ε	Minimum energy in the potential energy function, [kcal/mol]
ε_o	Dielectric constant of a free space
ε'	Dielectric constant of the surrounding medium
ϕ	Dihedral angle, [rad]
η_{nh}	Barostat friction coefficient
η	Solvent viscosity, [cP]
κ	Width of the Gaussian distribution
κ_P	System compressibility in Berendsen pressure control, [1/atm]
θ	Bond angle in MD simulations (equation 2.29), [rad]
θ_o	Ideal bond angle, [rad]
ρ	Density of the system, [amu/Å³]
ρ_b	Bulk density of water, [amu/Å³]
σ	van der Waals distance between two atoms, [Å]
τ	MD Simulation time, [ps]
τ_P	Pressure coupling constant, [ps]
τ_T	Relaxation time constant (Temperature coupling constant), [ps]

ω_o	Equilibrium out of plane angle, [rad]
ζ	Thermostat friction coefficient
Δt	Integration time step in Berendsen temperature control
ΔZ	Spacing between the sections in surface area calculations, [Å]
$\Theta(N)$	Time scaling factor in electrostatic energy calculations
Ξ	Virial of Clausius
Ψ	Schrödinger wave function
∇_e^2	Laplacian operator with respect to the position of the electron
∇_N^2	Laplacian operator with respect to the position of the nuclei
$\nabla_i U$	Potential energy of atom i, [kcal/mol]

Special

$A(p^n, r^N)$	Property of interest as a function of momentum and coordinates
$< N(r, dr) >$	Average number of atoms in the shell
$\rho(p^N, r^N)$	Probability density function
$\langle \cdots \rangle$	Ensemble average in the MD simulation
$\langle r_{ij} \rangle$	Ensemble average of the lectin structure coordinates in the solution

Abbreviations

ABNR	Adopted basis Newton Raphson
ASA	Accessible surface area
BPTI	Bovine pancreatic trypsin inhibitor
CHARMM	Chemistry at HARvard Macromolecular Mechanics
cmc	Critical micelle concentration
CSFF	Carbohydrate solution force field
DPC	Dodecylphosphocholine
FFT	Fast Fourier transforms
MD	Molecular dynamics
NPT	Constant number, pressure and temperature simulation
NVE	Constant number, volume and energy simulation
NVT	Constant number, volume and temperature simulation
OG	*n*-octyl-β-D-glucopyranoside
PBC	Periodic boundary conditions
PDB	Protein data bank
PME	Particle-Mesh-Ewald
RDF	Radial distribution functions
RHDO	Rhombic dodecahedron
RMS	Root mean square
RMSD	Root mean square deviations
RMSF	Root mean square fluctuations
SANS	Small angle neutron scattering
SAXS	Small angle X-ray scattering
SD	Steepest descent
TETR	Tetragonal
TIP3P	Transferable intermolecular potential 3 point
TIP4P	Transferable intermolecular potential 4 point
TIP4P-Ew	Transferable intermolecular potential 4 point-Ewald
μVT	Constant chemical potential, volume and temperature simulation

List of tables

List of figures

Appendices

Appendix A: Model molecular topology file for OG

```
*********************************************************************
* CHARMM27 Carbohydrate topology file (CSFF_top.inp) for pyranose simulations
* Michelle Kuttel, J. W. Brady and Kevin J. Naidoo
* Aug/Sep 2001
* Modified from the PHLB (Palma-Himmel-Liang-Brady) force field parameters:
* added CPS atom type and altered primary alcohol
* dihedral force constants from CTS values
* to lower energy the barriers to primary alcohol rotation
* Comments to Kevin J. Naidoo email:knaidoo,science.uct.ac.za
*********************************************************************
*
  22   1
!
! References
!
! M. Kuttel and J. W. Brady and K. J. Naidoo. "Carbohydrate Solution
! Simulations: Producing a Force Field with Experimentally Consistent
! Primary Alcohol Rotational Frequencies and Populations",
! J. Comput. Chem., 2002, 23:1236-1243
!
! PHLB force field precursor
!
! R. Palma and M. E. Himmel and G. Liang and J. W. Brady. "Molecular
! Mechanics Studies of Cellulases" in "Glycosyl Hydrolases in Biomass
! Conversion: ACS Symposium Series", published by ACS, 2001, editor
! M. E. Himmel
!
! HBFB force field precursor
!
! S. N. Ha and A. Giammona and M. Field and J. W. Brady, "A revised
! potential-energy surface for molecular studies of carbohydrates",
! Carbohydr. Res., 1988,180, 207-221

MASS    4 HT       1.00800        ! TIP3P water hydrogen
MASS   56 OT       15.99940       ! TIP3P water oxygen
MASS   60 OSPC     15.99940       ! SPC water oxygen
MASS   61 HSPC     1.00800        ! SPC water hydrogen
MASS   73 HAS      1.00800        ! sugar aliphatic hydrogen
MASS   74 HOS      1.00800        ! sugar hydroxyl hydrogen
MASS   75 CTS      12.01100       ! sugar aliphatic carbon
MASS   76 OHS      15.99940       ! sugar hydroxy oxygen
MASS   78 CBS      12.01100       ! C1 in beta sugars
MASS   79 CPS      12.01100       ! primary alcohol carbon (CT3)
MASS   77 OES      15.99940       ! sugar ring oxygen
MASS   89 LP       0.0            ! lone pair for TIP4P


AUTOGENERATE angles dihedrals
DEFA FIRS NONE LAST NONE
```

```
!#######################################################################
!                    CARBOHYDRATE

------------------ Water section ------------------

RESI TIP4        0.000       ! TIP4P water model, generate using noangle nodihedral
GROUP
ATOM OH2  OT      0
ATOM H1   HT      0.52
ATOM H2   HT      0.52
ATOM LP1  LP    -1.04
BOND OH2 H1 OH2 H2 H1 H2                ! The last bond is needed for shake
BOND OH2 LP1 LP1 H1 LP1 H2              ! The last bond is needed for shake
ANGLE H1 OH2 H2
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE


RESI TIP3        0.000       ! TIP3P water model, generate using noangle nodihedral
GROUP
ATOM OH2  OT     -0.834
ATOM H1   HT      0.417
ATOM H2   HT      0.417
BOND OH2 H1 OH2 H2 H1 H2                ! The last bond is needed for shake
ANGLE H1 OH2 H2
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE

----------------- Glycolipid section -------------

RESI BOG   0.0                          ! beta octyl glucoside
GROUP
ATOM O1   OES  -.40
ATOM C1   CBS  0.200
ATOM H1   HAS  0.090
ATOM C5   CTS  0.250
ATOM H5   HAS  0.090
ATOM O5   OES  -.400
ATOM C7   CTS  -0.01
ATOM H7A  HAS   0.09
ATOM H7B  HAS   0.09
GROUP
ATOM C2   CTS  0.140
ATOM H2   HAS  0.090
ATOM O2   OHS  -.66
ATOM HO2  HOS  0.43
GROUP
ATOM C3   CTS  0.140
ATOM H3   HAS  0.090
ATOM O3   OHS  -.66
ATOM HO3  HOS  0.43
GROUP
ATOM C4   CTS  0.140
ATOM H4   HAS  0.090
ATOM O4   OHS  -.66
ATOM HO4  HOS  0.43
```

```
GROUP
ATOM C6   CPS  0.050
ATOM H6A  HAS  0.090
ATOM H6B  HAS  0.090
ATOM O6   OHS  -.66
ATOM HO6  HOS  0.43
! Attach octane chain as C7-C14
GROUP
ATOM C8   CTS  -0.18
ATOM H8A  HAS   0.09
ATOM H8B  HAS   0.09
ATOM C9   CTS  -0.18
ATOM H9A  HAS   0.09
ATOM H9B  HAS   0.09
ATOM C10  CTS  -0.18
ATOM H10A HAS   0.09
ATOM H10B HAS   0.09
ATOM C11  CTS  -0.18
ATOM H11A HAS   0.09
ATOM H11B HAS   0.09
ATOM C12  CTS  -0.18
ATOM H12A HAS   0.09
ATOM H12B HAS   0.09
ATOM C13  CTS  -0.18
ATOM H13A HAS   0.09
ATOM H13B HAS   0.09
ATOM C14  CTS  -0.27
ATOM H14A HAS   0.09
ATOM H14B HAS   0.09
ATOM H14C HAS   0.09

BOND C1   C2    C2   C3    C3   C4    C4  C5   C5  O5    O5  C1
BOND C1   H1    C1   O1    C2  O2   O2  HO2   C3  H3   C3   O3
BOND O3   HO3   C6   H6A   C6   H6B   C6  O6   O6  HO6
BOND C4   H4    C4   O4    O4   HO4   C5  H5   C5  C6    C2  H2
BOND O1   C7    C7   H7A   C7   H7B   C7  C8   C8  H8A  C8  H8B
BOND C8   C9    C9   H9A   C9   H9B   C9  C10  C10 H10A  C10 H10B
BOND C10  C11   C11  H11A  C11  H11B  C11 C12  C12 H12A  C12 H12B
BOND C12  C13   C13  H13A  C13  H13B  C13 C14  C14 H14A  C14 H14B  C14 H14C
DONOR HO2 O2
DONOR HO3 O3
DONOR HO4 O4
DONOR HO6 O6
ACCE   O1
ACCE   O2
ACCE   O3
ACCE   O4
ACCE   O5
ACCE   O6

! Minimized from initial build
IC  C5   O5   C1   C2     1.4206  114.33  -62.14  109.08   1.5339
IC  O5   C1   C2   C3     1.4142  109.08   60.01  108.48   1.5247
IC  C1   C2   C3   C4     1.5339  108.48  -58.41  108.81   1.5297
IC  C2   C3   C4   C5     1.5247  108.81   56.18  109.40   1.5407
IC  C3   C4   C5   O5     1.5297  109.40  -55.29  109.98   1.4206
```

```
IC  C4   C5   O5   C1     1.5407  109.98   59.77  114.33   1.4142
IC  O5   C2   *C1  O1     1.4142  109.08  120.12  107.28   1.4101
IC  O5   C2   *C1  H1     1.4142  109.08 -120.79  110.90   1.0920
IC  C1   C3   *C2  H2     1.5339  108.48  118.82  109.01   1.0955
IC  C1   C3   *C2  O2     1.5339  108.48 -122.51  108.20   1.4145
IC  C2   C4   *C3  O3     1.5247  108.81  121.29  107.95   1.4150
IC  C10  C9   C8   C7     1.5334  112.21  179.22  112.47   1.5299
IC  C11  C10  C9   C8     1.5336  112.46 -179.97  112.21   1.5333
IC  C12  C11  C10  C9     1.5335  112.19  179.23  112.46   1.5334
IC  C13  C12  C11  C10    1.5334  112.44 -179.85  112.19   1.5336
IC  C14  C13  C12  C11    1.5318  112.25  179.27  112.44   1.5335
IC  H14C C14  C13  C12 1.0903  110.94  179.99  112.25   1.5334
IC  O1   C8   *C7  H7A    1.4114  108.74  120.44  109.94   1.0908
IC  O1   C8   *C7  H7B    1.4114  108.74 -120.90  110.46   1.0916
IC  C7   C9   *C8  H8A    1.5299  112.47  121.85  109.68   1.0904
IC  C7   C9   *C8  H8B    1.5299  112.47 -121.63  109.64   1.0905
IC  C8   C10  *C9  H9A    1.5333  112.21  121.74  109.40   1.0907
IC  C8   C10  *C9  H9B    1.5333  112.21 -121.86  109.50   1.0907
IC  C9   C11  *C10 H10A   1.5334  112.46  121.93  109.45   1.0907
IC  C9   C11  *C10 H10B   1.5334  112.46 -121.78  109.43   1.0907
IC  C10  C12  *C11 H11A   1.5336  112.19  121.75  109.41   1.0907
IC  C10  C12  *C11 H11B   1.5336  112.19 -121.86  109.54   1.0907
IC  C11  C13  *C12 H12A   1.5335  112.44  121.91  109.55   1.0907
IC  C11  C13  *C12 H12B   1.5335  112.44 -121.69  109.50   1.0907
IC  C12  C14  *C13 H13A   1.5334  112.25  121.91  109.24   1.0906
IC  C12  C14  *C13 H13B   1.5334  112.25 -121.96  109.22   1.0907
IC  C13  H14C *C14 H14    1.5318  110.94  121.82  107.89   1.0902
IC  C13  H14C *C14 H14B 1.5318  110.94 -121.80  107.85   1.0903

END
```

Appendix B: Model parameter file for OG

```
*******************************************************************
* Carbohydrate parameter file CSFF_parm.inp for pyranose simulations
* Michelle Kuttel, J. W. Brady and Kevin J. Naidoo
* Aug/Sep 2001
* Modified from the PHLB (Palma-Himmel-Liang-Brady) force field parameters:
* added CPS atom type and altered primary alcohol
* dihedral force constants from CTS values
* to lower energy the barriers to primary alcohol rotation
* Comments to Kevin J. Naidoo email:naidoo,science.uct.ac.za
*******************************************************************
*
!
! References
! M. Kuttel and J. W. Brady and K. J. Naidoo. "Carbohydrate Solution Simulations: Producing a
! Force Field with Experimentally Consistent Primary Alcohol Rotational Frequencies and
! Populations", J. Comput. Chem., 2002, 23:1236-1243
!
! PHLB force field precursor
!
! R. Palma and M. E. Himmel and G. Liang and J. W. Brady. "Molecular Mechanics Studies of
! Cellulases" in "Glycosyl Hydrolases in Biomass Conversion: ACS Symposium Series", published
! by ACS, 2001, editor M. E. Himmel
!
! HBFB force field precursor
!
! S. N. Ha and A. Giammona and M. Field and J. W. Brady, "A revised potential-energy surface
! for molecular studies of carbohydrates", !Carbohydr. Res., 1988,180,207-221
!
! NOTE: messages about multiple dihedral terms on reading this file are normal
!
BONDS
HOS  OHS    460.5000          0.9595
HAS  CTS    335.6034          1.1000
HAS  CPS    335.6034          1.1000
HAS  CBS    335.6034          1.1052
CTS  OHS    384.0792          1.4066
CPS  OHS    384.0792          1.4066
CBS  OHS    384.0792          1.3932
CTS  CTS    325.5297          1.5066
CTS  CPS    325.5297          1.5066
CBS  CTS    325.5297          1.5074
CTS  OES    385.3133          1.4165
CPS  OES    385.3133          1.4165
CBS  OES    385.3133          1.4202
HSPC OSPC   450.0             1.0            ! SPC Geometry
HSPC HSPC     0.0             1.6329931      ! SPC Geometry (for SHAKE )
HT   OT     450.0             0.9572         ! TIP3P geometry
HT   HT       0.0             1.5139         ! TIP3P geometry (for SHAKE)

THETAS
HAS  CTS  CTS        42.9062        109.7502
HAS  CBS  CTS        42.9062        109.7502
HAS  CTS  CBS        42.9062        109.7502
```

```
OHS  CTS  CTS        112.2085          107.6019
OHS  CTS  CBS        112.2085          107.6019

DIHEDRALS
CTS  CTS  CTS  CTS    -1.0683   1     0.0
CTS  CTS  CTS  CTS    -0.5605   2     0.0
CTS  CTS  CTS  CTS     0.1955   3     0.0
CPS  CTS  CTS  CTS    -1.0683   1     0.0
CPS  CTS  CTS  CTS    -0.5605   2     0.0

IMPROPER

NONBONDED NBXMOD 5 atom CDIEL shift vatom VDISTANCE VSWITCH -
    CUTNB 13.0 CTOFNB 12.0 CTONNB 10.0 EPS 1.0 E14FAC 1.0 WMIN 1.5
!
!              Emin      Rmin
!              (kcal/mol) (A)
HT        0.00     -0.046    0.2245        ! TIP3P
OT        0.00     -0.1521   1.7682        ! TIP3P
HOS       0.00     -0.0460   0.2245
HAS       0.00     -0.0220   1.3200
CTS       0.00     -0.0200   2.2750    0.000    -0.01000    1.90000 !
CBS       0.00     -0.0200   2.2750    0.000    -0.01000    1.90000 !
CPS       0.00     -0.0200   2.2750    0.000    -0.01000    1.90000 !
OHS       0.00     -0.1521   1.7700        ! from para_na, on5
OES       0.00     -0.1521   1.7700        ! from para_na, on6
!
NBFIX
!
!      NBFIX the TIP3P water-water interactions
OT   OT     -0.152073 3.5365              ! TIPS3P VDW INTERACTION
HT   HT     -0.04598  0.4490
HT   OT     -0.08363  1.9927
!
! This force field treats hbonds implicitly using charges; therefore the
! following section wildcards all the energies to 0. Note that you should
! also set the IHBFRQ to zero in all calculations, to avoid wasting time updating
! the hydrogen bond list!
!
HBOND AEXP 4 REXP 6 HAEX 4 AAEX 0 NOACCEPTORS HBNOEXCLUSIONS ALL-
  CUTHB 0.5 CTOFHB 5.0 CTONHB 4.0 CUTHA 5.0 CTOFHA 90.0 CTONHA 90.0
!
H*   O*     -0.00     2.0

END
```

Appendix C: RMSD

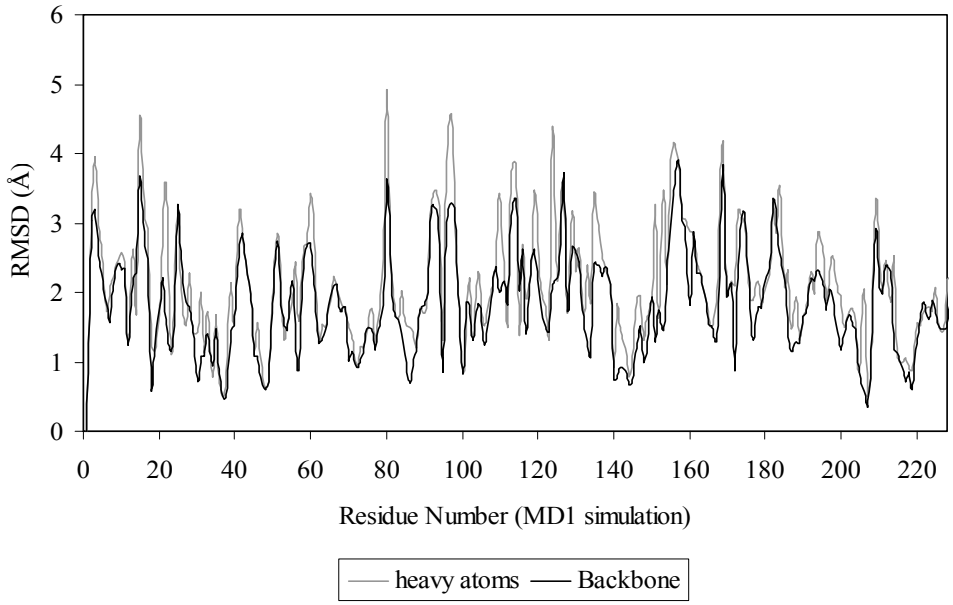

Figure C-1: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD1 simulation.

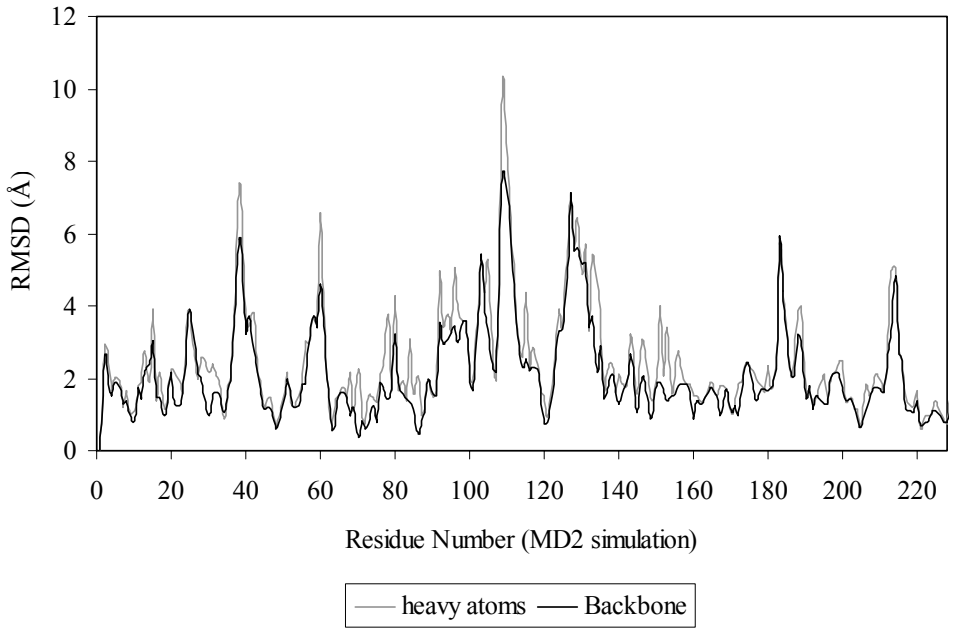

Figure C-2: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD2 simulation.

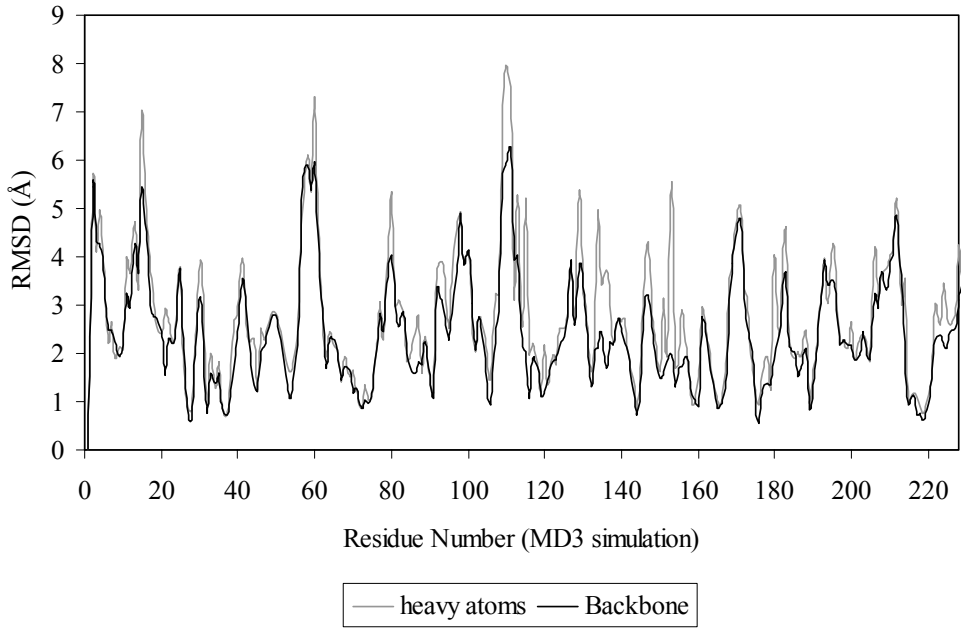

Figure C-3: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD3 simulation.

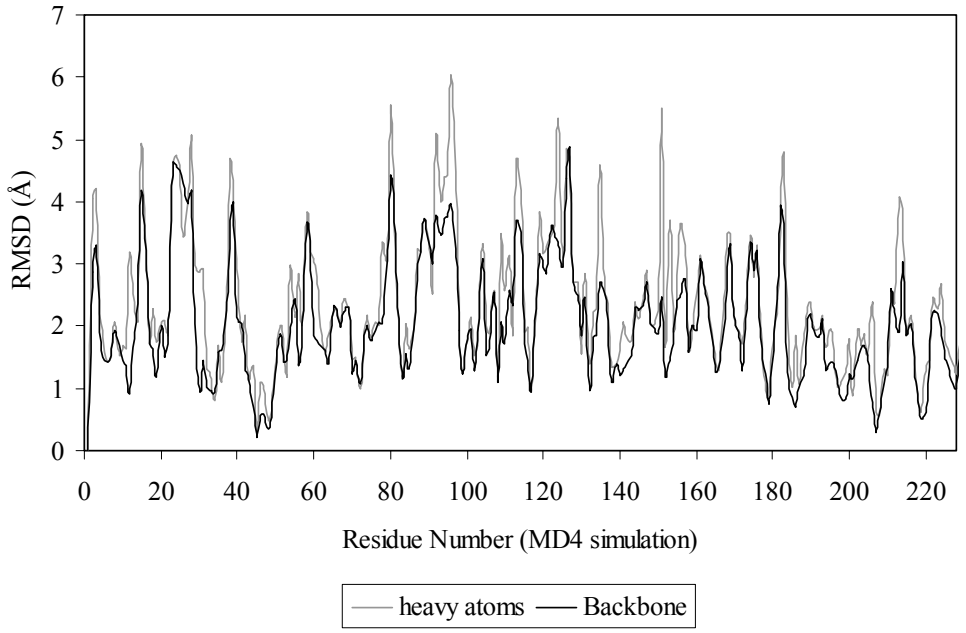

Figure C-4: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD4 simulation.

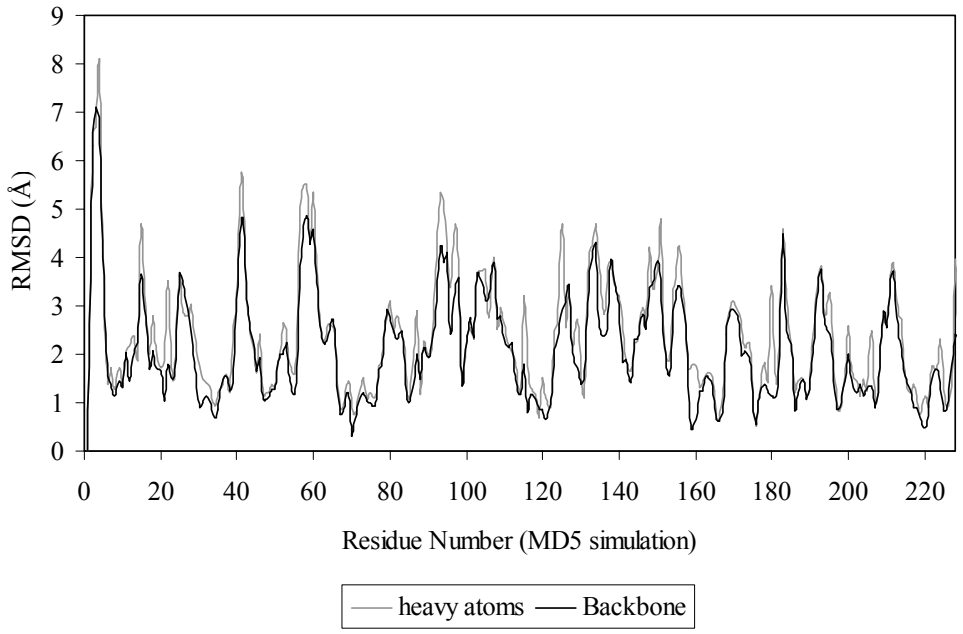

Figure C-5: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD5 simulation.

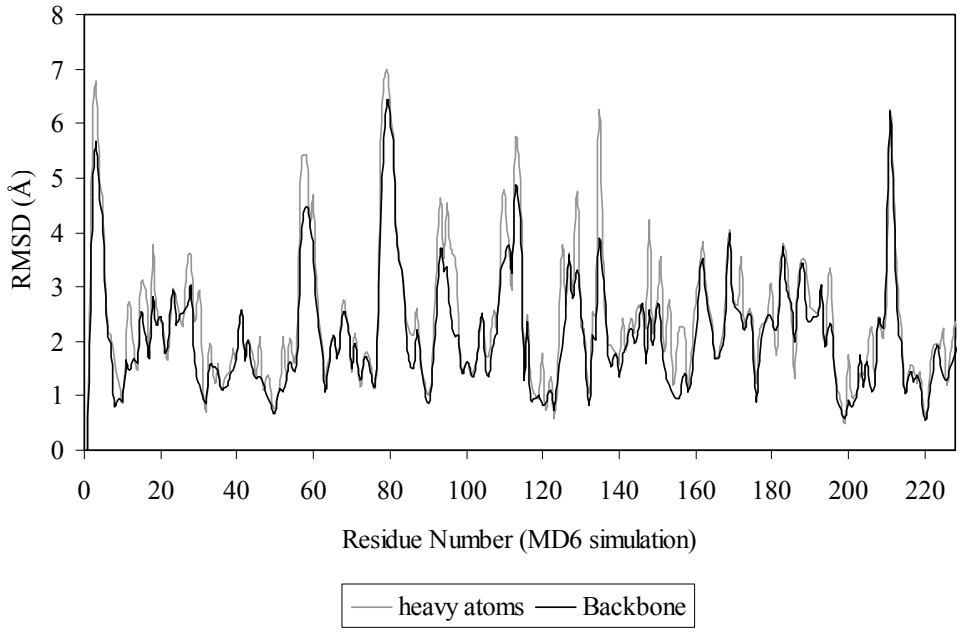

Figure C-6: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD6 simulation.

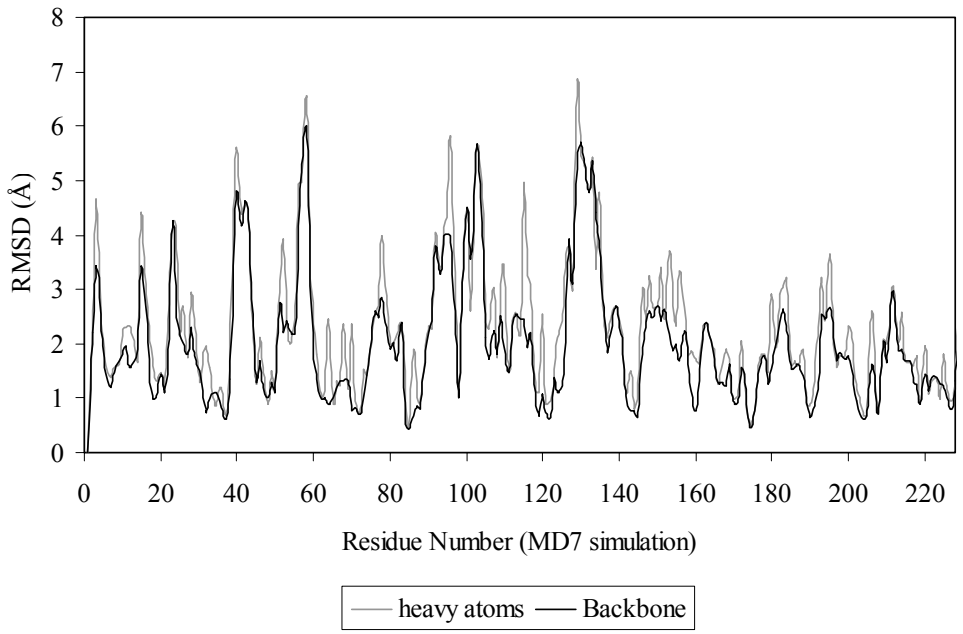

Figure C-7: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD4 simulation.

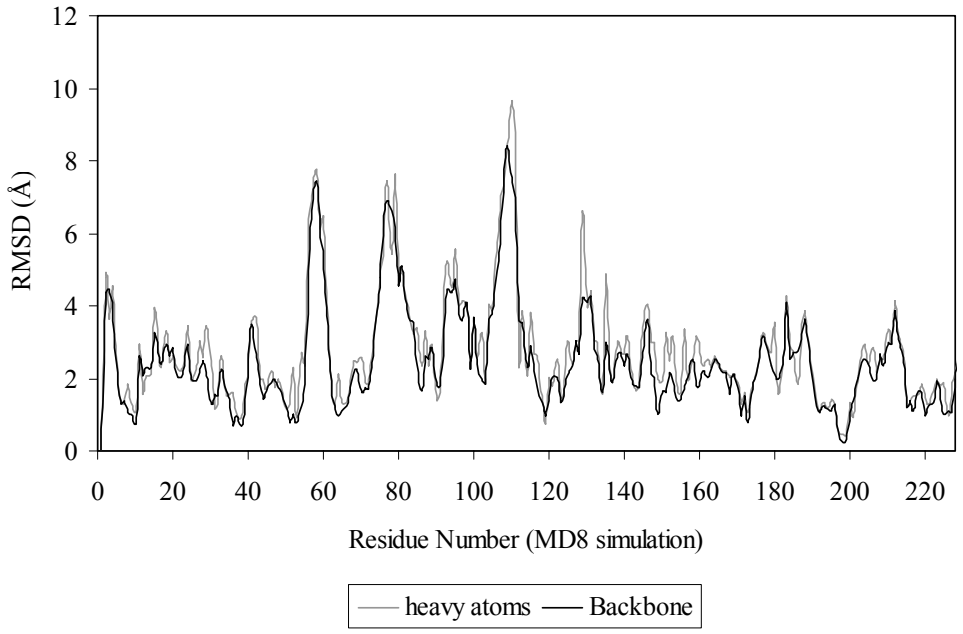

Figure C-8: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD4 simulation.

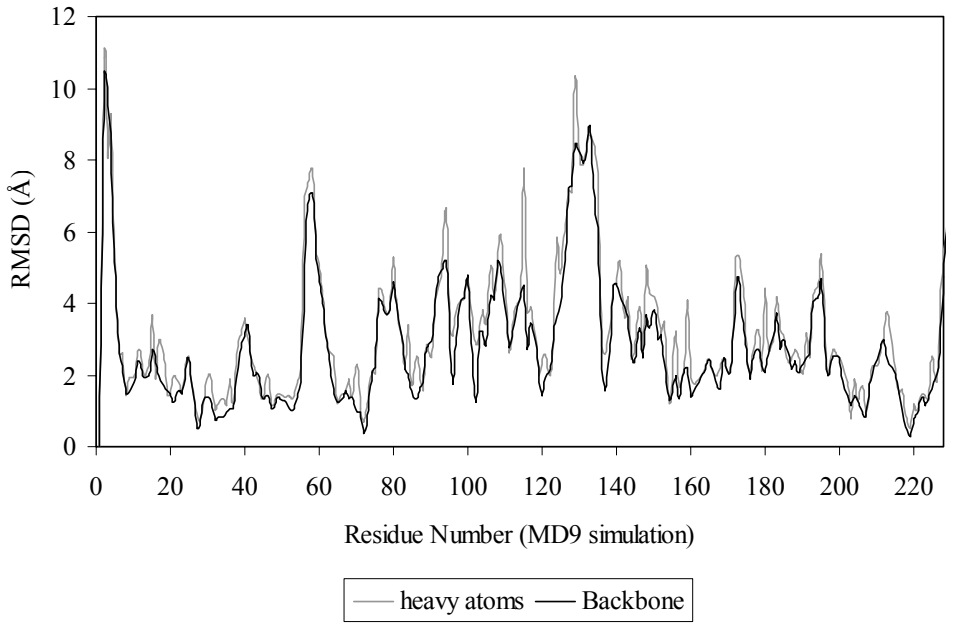

Figure C-9: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD9 simulation.

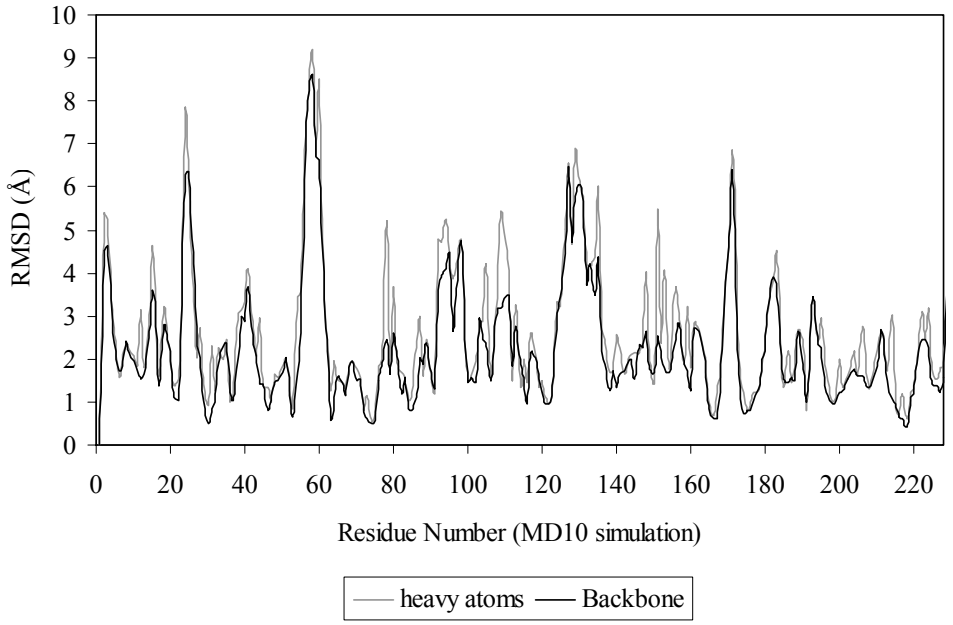

Figure C-10: The average RMSD with respect to the pea lectin residues. The averaged heavy atoms and the backbone atoms residue deviations are shown in gray and black color for MD10 simulation.

Appendix D: RMSF

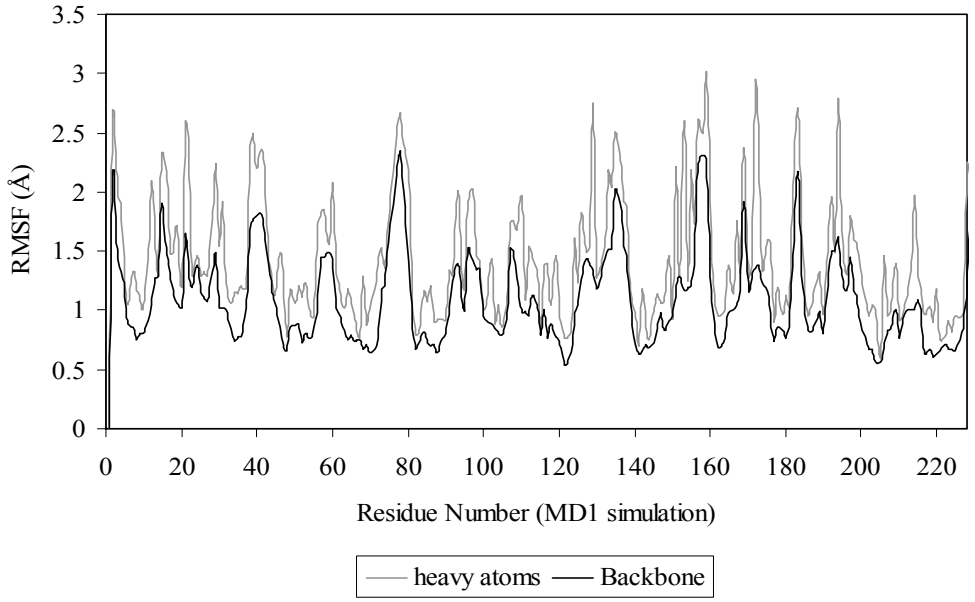

Figure D-1: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD1 simulation.

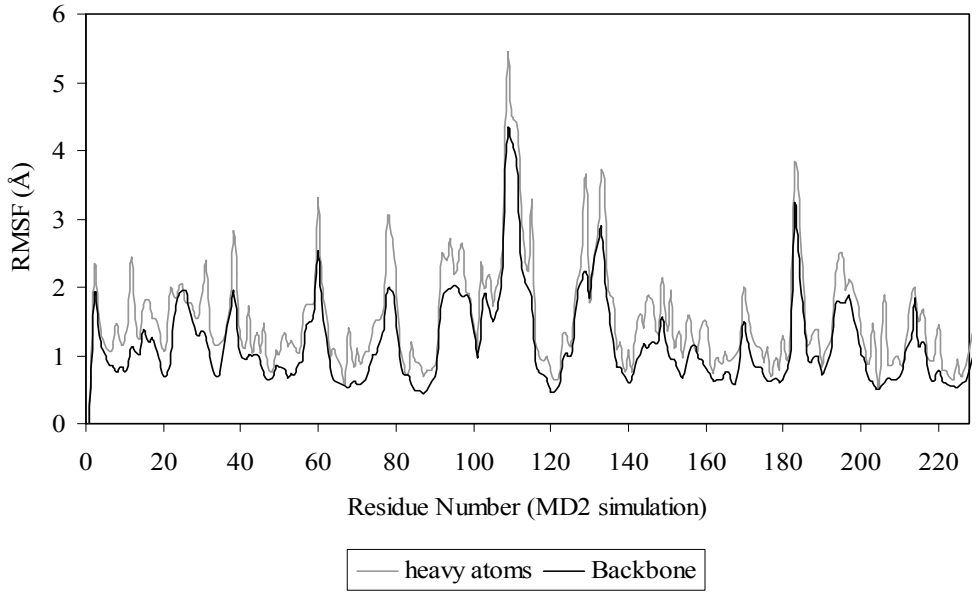

Figure D-2: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in grey and black color for MD2 simulation.

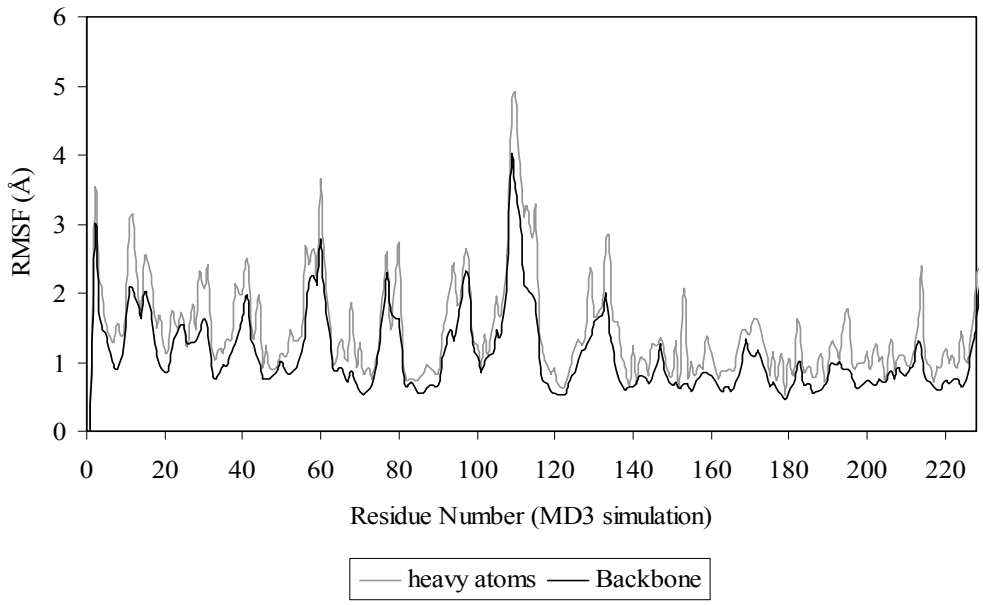

Figure D-3: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD3 simulation.

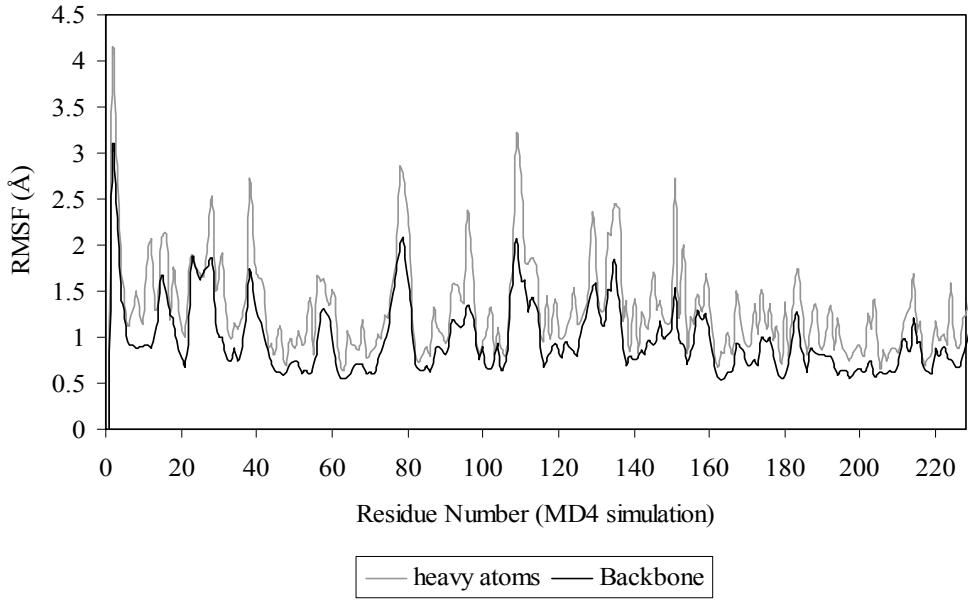

Figure D-4: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD4 simulation.

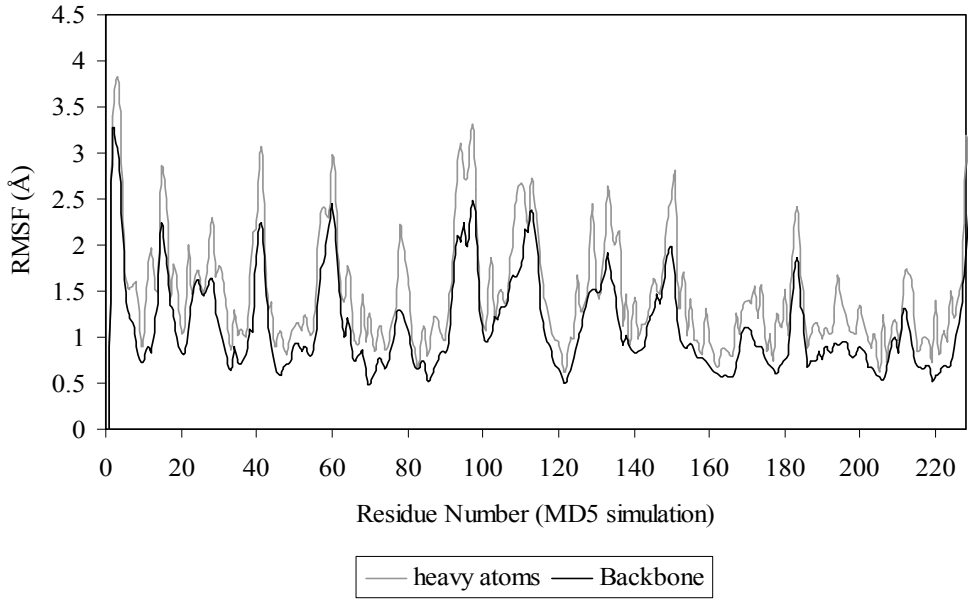

Figure D-5: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD5 simulation.

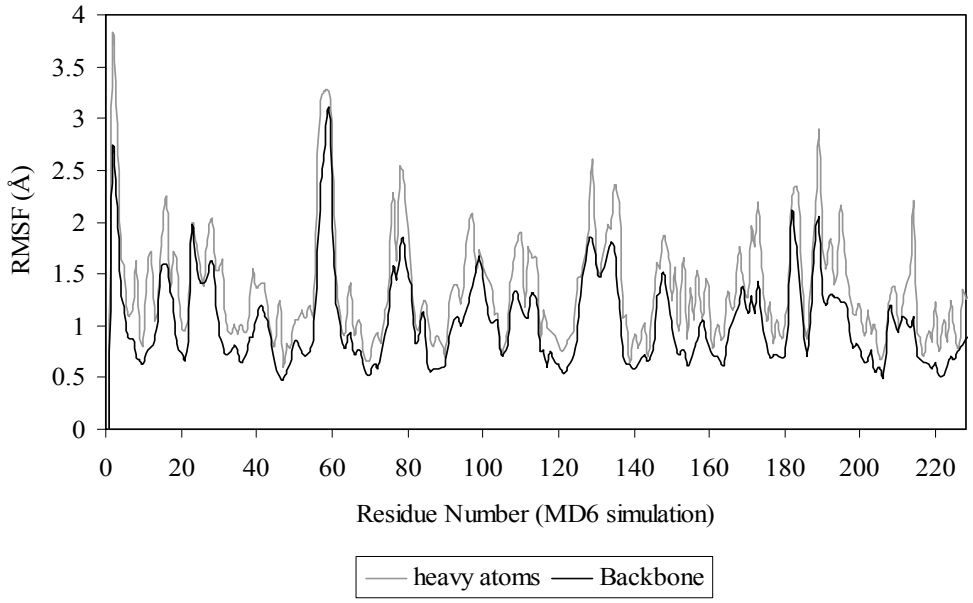

Figure D-6: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD6 simulation.

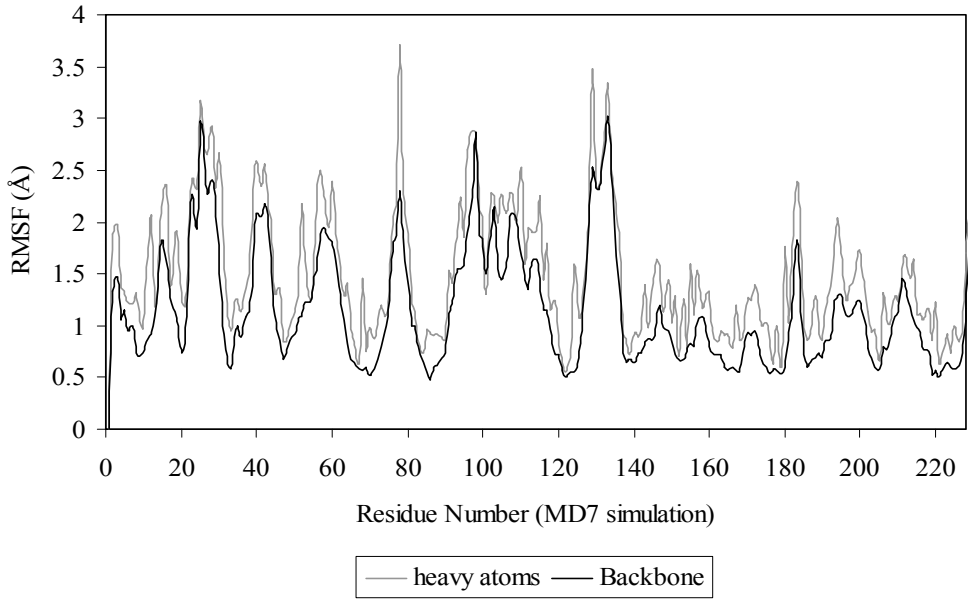

Figure D-7: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD7 simulation.

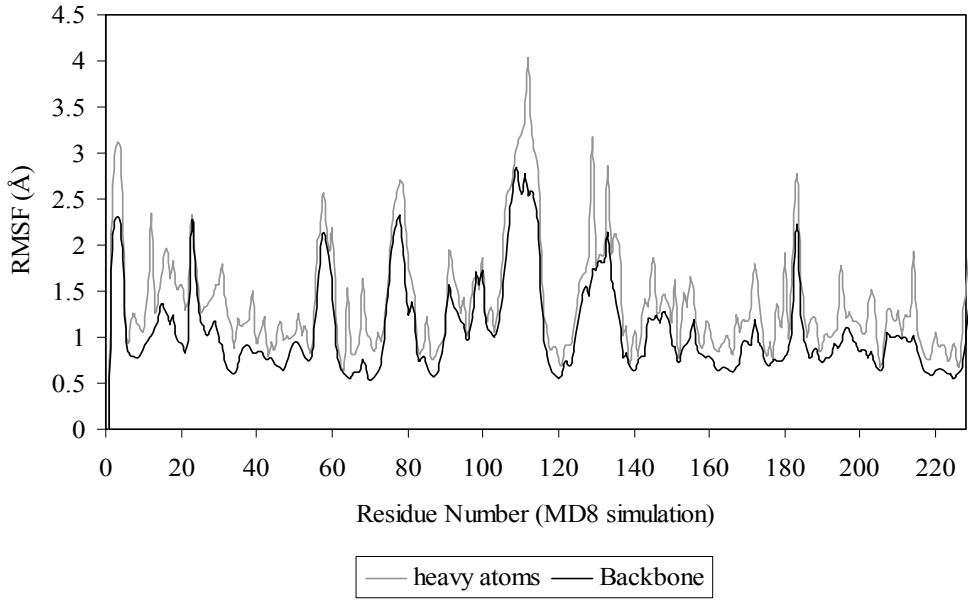

Figure D-8: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD8 simulation.

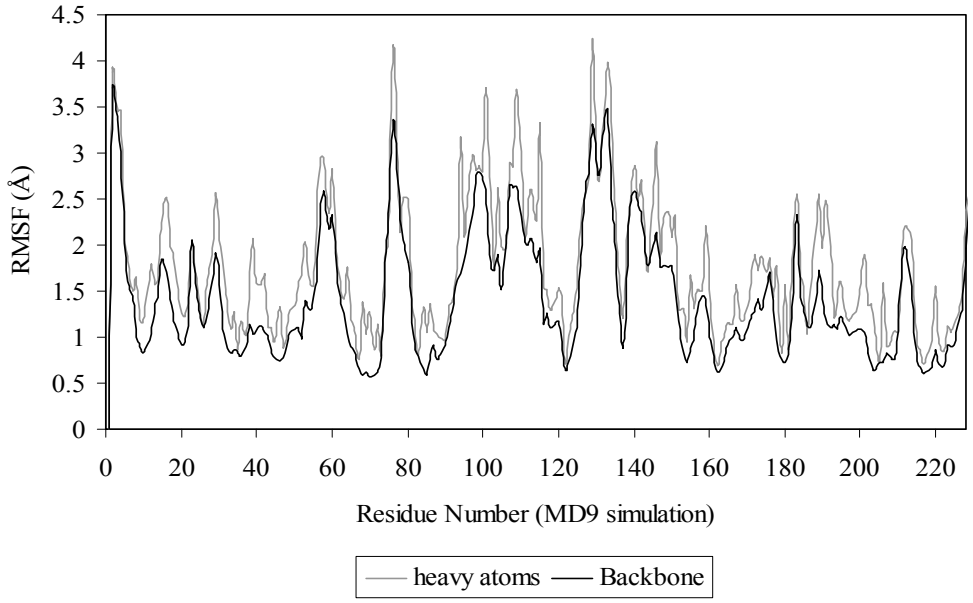

Figure D-9: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD9 simulation.

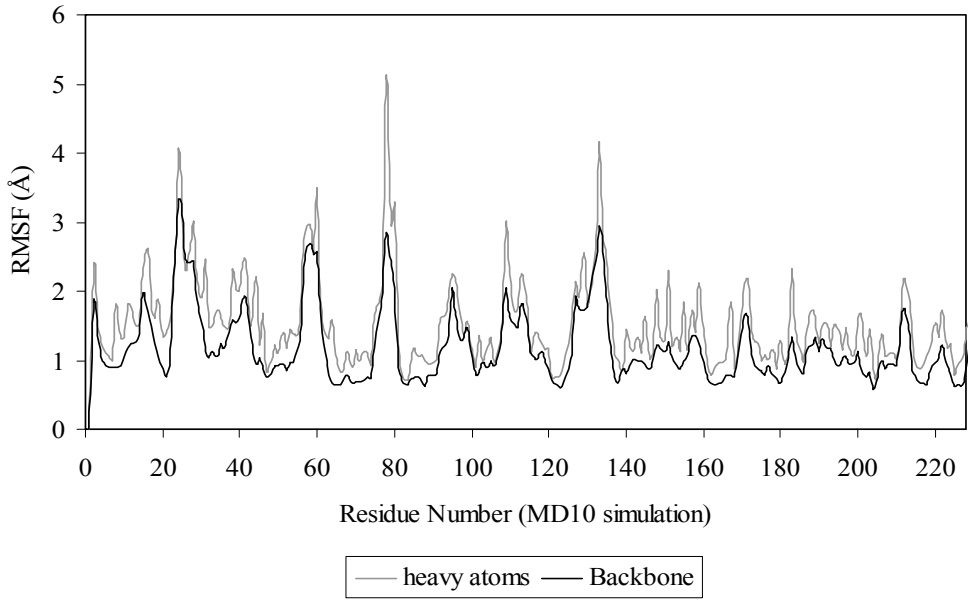

Figure D-10: The average positional fluctuations of the lectin residues. The average heavy atom and the backbone atom fluctuations with respect to the average solution structure are shown in gray and black color for MD10 simulation.

Appendix E: B-factor

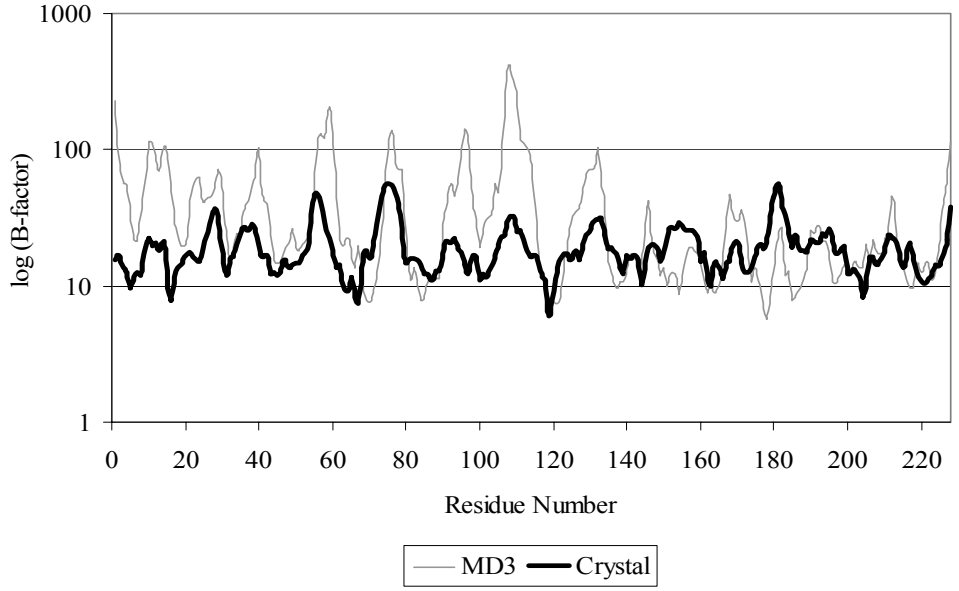

Figure E-1: Temperature B-factor profile computed from the average backbone positional fluctuations. Lectin dynamics are compared between x-ray crystallographic data (thick black) and the MD3 simulation (gray) against the residue number.

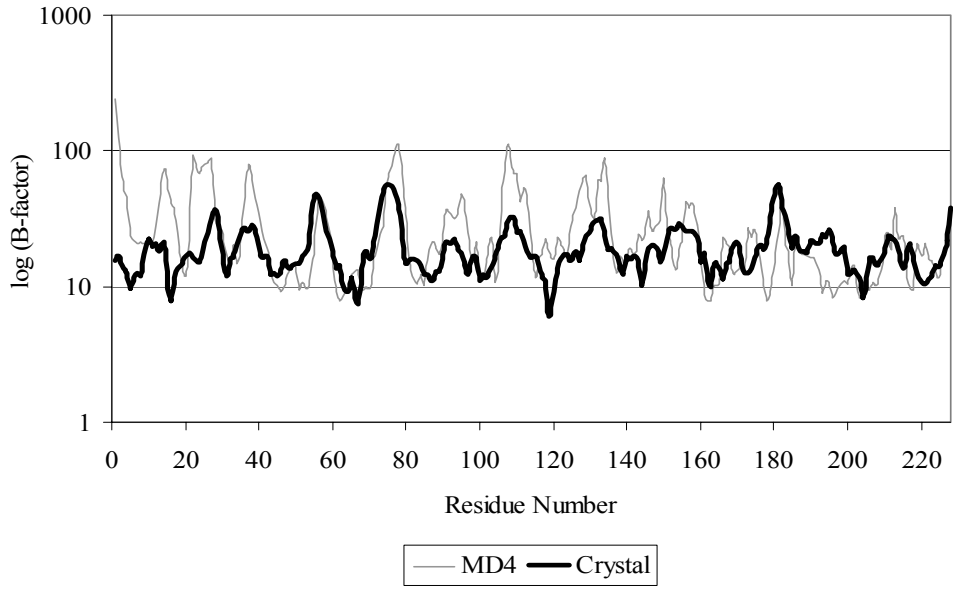

Figure E-2: Temperature B-factor profile computed from the average backbone positional fluctuations. Lectin dynamics are compared between x-ray crystallographic data (thick black) and the MD4 simulation (gray) against the residue number.

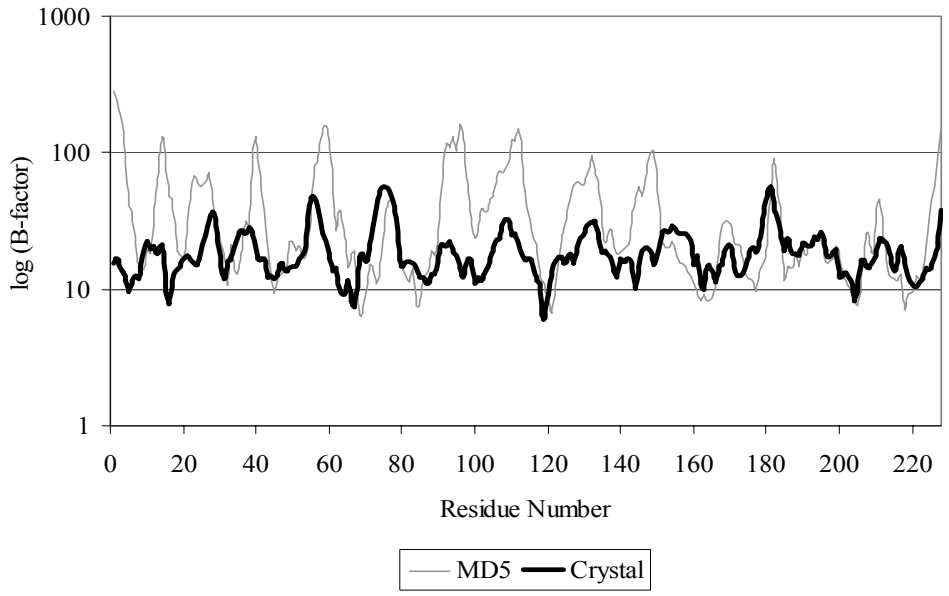

Figure E-3: Temperature B-factor profile computed from the average backbone positional fluctuations. Lectin dynamics are compared between x-ray crystallographic data (thick black) and the MD5 simulation (gray) against the residue number.

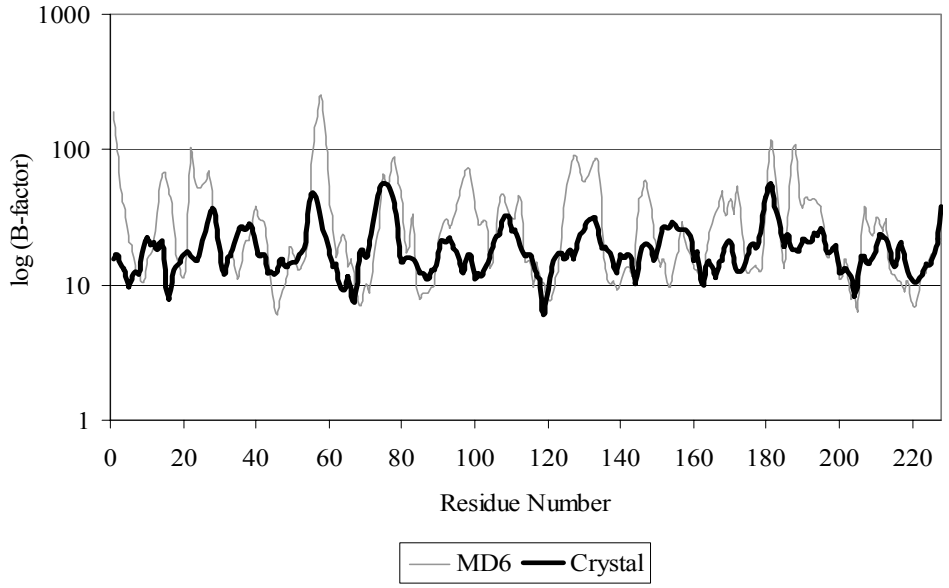

Figure E-4: Temperature B-factor profile computed from the average backbone positional fluctuations. Lectin dynamics are compared between x-ray crystallographic data (thick black) and the MD6 simulation (gray) against the residue number.

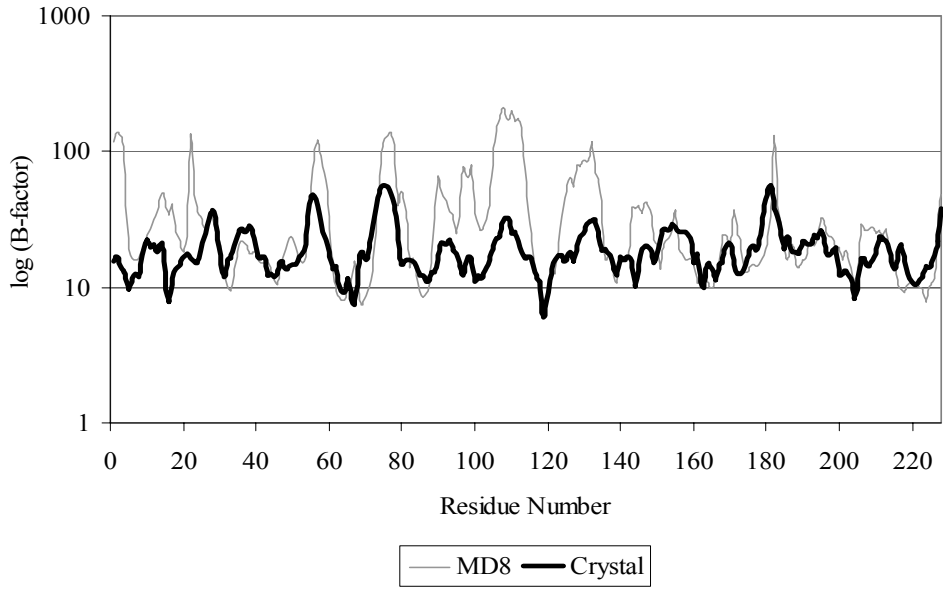

Figure E-4: Temperature B-factor profile computed from the average backbone positional fluctuations. Lectin dynamics are compared between x-ray crystallographic data (thick black) and the MD8 simulation (gray) against the residue number.

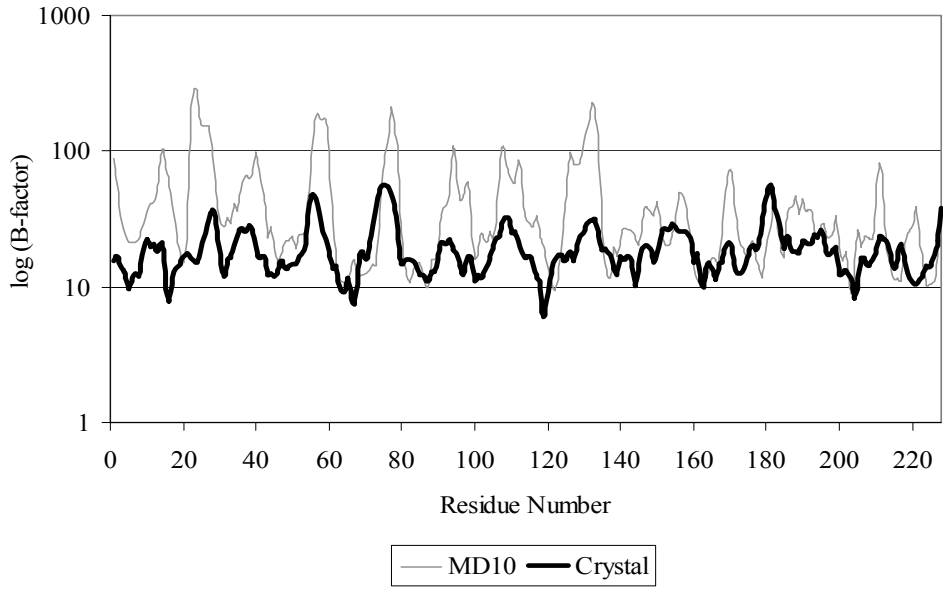

Figure E-4: Temperature B-factor profile computed from the average backbone positional fluctuations. Lectin dynamics are compared between x-ray crystallographic data (thick black) and the MD10 simulation (gray) against the residue number.

Appendix F: Radius of gyration

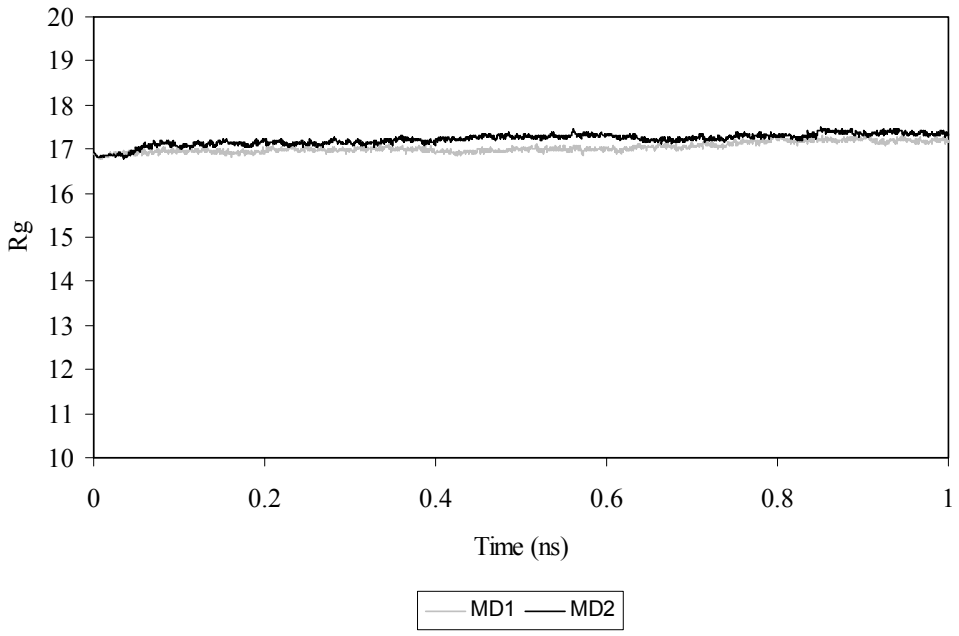

Figure F-1: Radius of gyration of the pea lectin monomer over simulation time in MD1 (grey) and MD2 (black) simulations.

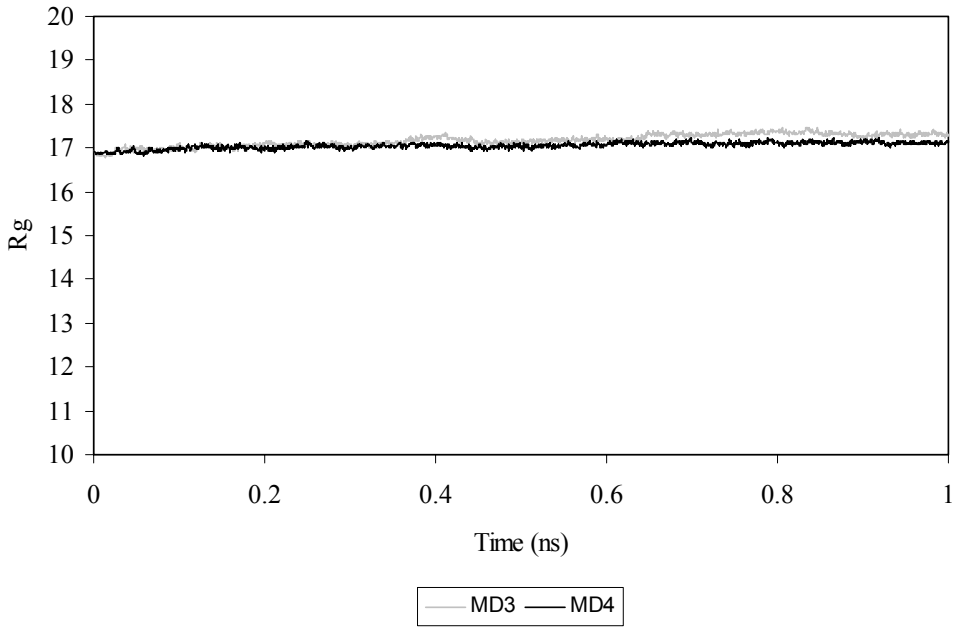

Figure F-2: Radius of gyration of the pea lectin monomer over simulation time in MD3 (grey) and MD4 (black) simulations.

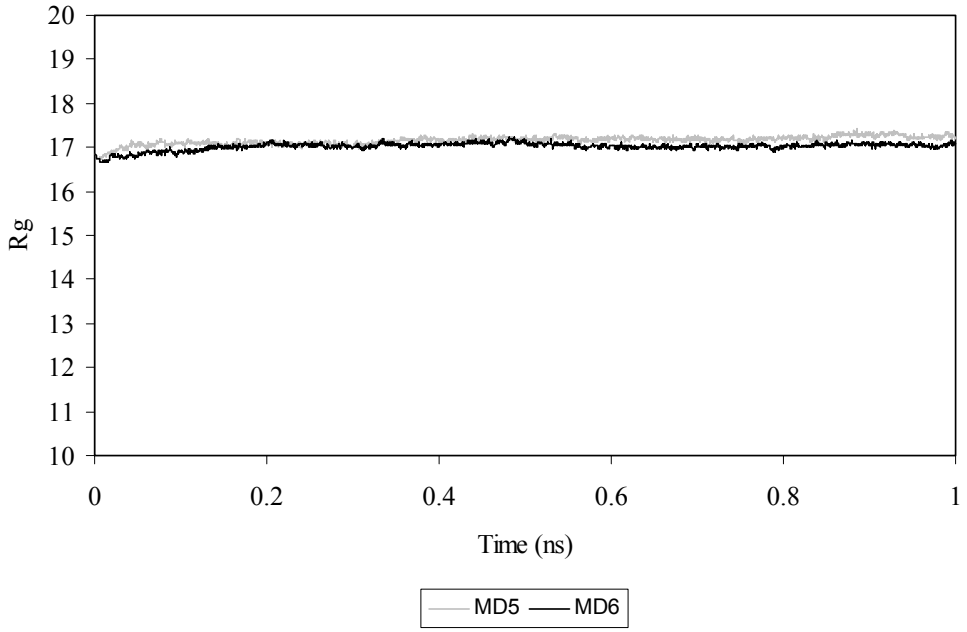

Figure F-3: Radius of gyration of the pea lectin monomer over simulation time in MD5 (grey) and MD6 (black) simulations.

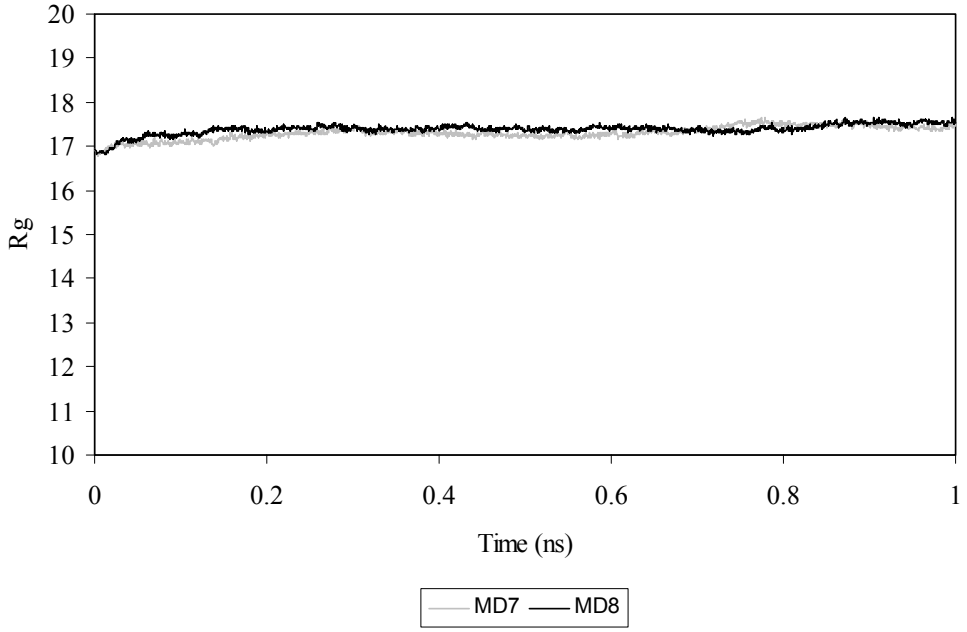

Figure F-4: Radius of gyration of the pea lectin monomer over simulation time in MD7 (grey) and MD8 (black) simulations.

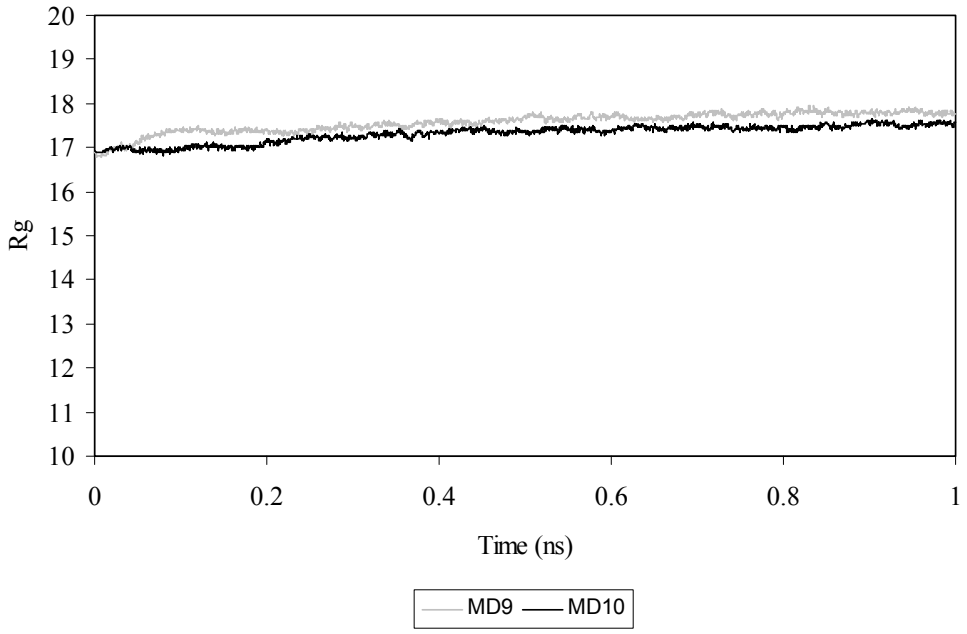

Figure F-5: Radius of gyration of the pea lectin monomer over simulation time in MD9 (grey) and MD10 (black) simulations.

Acknowledgements

I wish to thank my thesis advisor Professor Bernd Niemeyer for his excellent support and introducing me to the competitive field of molecular modelling. I am grateful to him for supervising this work and suggesting with innovative ideas from the beginning till to the successful completion. It is only with his initial motivation, freedom and encouragement which brought me to this success, I really appreciate his help and cooperation.

I would like to extend my special thanks to Professor Stephan Kabelac for his willingness to guide as a second supervisor for this thesis. It was very nice talking to him when I approached and I am grateful to him for accepting this work without any hesitations.

I appreciate the cooperative works with Dr. Lizhong He and Dr. rer. nat. Peter H. Thiesen. My special thanks to both of them for the invaluable discussion and suggestions. My department people in the Helmut-Schmidt-University / University of the Federal Armed Forces Hamburg and the co-workers at the GKSS Research Center, Geesthacht are among several others who have helped me in many ways, so I would like to take an advantage to appreciate their friendly behaviour and kindness.

Finally, though it is mentioned lately, I would like to thank my family and friends for their moral support and advice.

Praveen Konidala

Curriculum Vitae

Personal Details

Praveen Konidala

Date of birth:	9th April 1976
Place of birth:	Chennai, India
Nationality:	Indian

Education

10/1998 - 09/2001	Master of Science (Process Engineering) Technical University Hamburg-Harburg (TUHH), Germany
09/1994 - 04/1998	Bachelor of Technology (Chemical Engineering) University of Madras, Madras, India
06/1992 - 03/1994	Higher Secondary School SRKM Higher Secondary School (Main), Chennai, India
06/1990 - 04/1992	High School SRKM Higher Secondary School (Main), Chennai, India

Professional Experience

03/2003 – to date	Scientific Research Assistant Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg, Institue of Thermodynamics, Hamburg, Germany
12/2001 – 02/2003	Student Research Assistant Helmut-Schmidt-Universität / Universität der Bundeswehr Hamburg, Institue of Thermodynamics, Hamburg, Germany
03/2001 - 09/2001	Diploma Thesis (Master of Science) Siemens AG, Automation & Drives, Karlsruhe, Germany
04/2000 - 06/2000	Master Project Work Technical University Hamburg-Harburg, Hamburg, Germany
12/1999 - 02/2000	Industrial Internship Siemens Axiva GmbH & Co. KG, Frankfurt, Germany
08/1997 - 03/1998	Bachelor Project Work University of Madras, Madras, India