Summer 2011

Water behavior in different biological environments

Ying-Hua Chung
University of Iowa

WATER BEHAVIOR IN DIFFERENT BIOLOGICAL ENVIRONMENTS

by

Ying-Hua Chung

An Abstract

Of a thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Chemistry
in the Graduate College of
The University of Iowa

July 2011

Thesis Supervisor: Associate Professor Claudio J. Margulis

ABSTRACT

In this thesis, we report on our studies of water dynamics and structure in various biological environments which include: the surfaces of proteins and various oligosaccharides, the intervening space between proteins; and in the vicinity of cryoprotectant di-saccharides in the liquid and ice phases. From a theoretical perspective, we propose methodology to compute diffusivity and residence times on the surface of biomolecules. In particular our proposed algorithm to compute residence times appears to be better in dealing with poor statistics associated with the number of water molecules that remain on a surfaces for extended times. The type of linkage between monomers and the anomeric configuration all play a major role in determining the structure and dynamics of water on the surface of carbohydrates.

Abstract Approved: _____
Thesis Supervisor

_____
Title and Department

_____
Date

WATER BEHAVIOR IN DIFFERENT BIOLOGICAL ENVIRONMENTS

by

Ying-Hua Chung

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Chemistry
in the Graduate College of
The University of Iowa

July 2011

Thesis Supervisor: Associate Professor Claudio J. Margulis

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Ying-Hua Chung

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Chemistry at the July 2011 graduation.

Thesis Committee: _____
Claudio Margulis, Thesis Supervisor

_____
Christopher M. Cheatum

_____
Ernesto Fuentes

_____
Amnon Kohen

_____
Jan-Uwe Rohde

ACKNOWLEDGMENTS

First and the most I would like to thank God for everything.

I would like to thank Professor Claudio J. Margulis for his motivation, support and setting the bar for both academic and personal achievement. I also thank Professor Christopher M. Cheatum, Ernesto Fuentes, Amnon Kohen, and Jan-Uwe Rohde for serving on my thesis committee. Their suggestions make the thesis more complete.

I thank my previous and current labmates: Dr. Kasinadar Veluraja, Dr. Zhonghan Hu, Mr. Ryan Daly, Dr. Junchao Xia, Ms. Laura Parker, Mr. Sai Ramadugu, Dr. Hemant Kashyap, and Dr. Harsha Vardhan Reddy Annapureddy for their guidance, support and friendship. Without their support, this thesis and my other two projects that are not included in the thesis might not have been done.

In the view of finishing my Ph.D. program, I have to thank Professor Alberto Segre. He inspired me to start computation-related research. I thank Professor John W. Schweitzer, Jan H. Jensen, and Gerald Payne for giving me a solid theoretical ground. I also have to thank Professor Sarah C. Larson, Mark A. Young, and Vicki H. Grassian. Working with them in teaching Physical Chemistry, I learned many ways and many tricks in teaching. I also would like to thank Mr. Michael Perry and Dr. Marta M. Tryzna for their proofreading and friendship that enriched my life in Iowa City. I thank Mr. Ren-Jie Yu, Dr. Der-Fa Lu, Dr. Li-Kuei Chiou, Dr. Daping Du, Ms. Hui-Yun Wu, Dr. Gopeekrishnan Sreenilayam, Mr. Koi-Hin Samuel Kwok,

Dr. Atsushi Yahashiri, Mr. Ashish Datt, Mr. Chia-Ming Wu, Mr. Wei-Tsung Lee, and other friends that I have not mentioned.

Finally, I thank my parents, Kin-Tong Chung and Yen-Hong Lin, my brother, Ying-Hao, my wife, Ya-Hui, and my daughter Lydia for their moral support and love through every stage of my life. Without my familys support, many things would not have been possible.

ABSTRACT

In this thesis, we report on our studies of water dynamics and structure in various biological environments which include: the surfaces of proteins and various oligosaccharides, the intervening space between proteins; and in the vicinity of cryoprotectant di-saccharides in the liquid and ice phases. From a theoretical perspective, we propose methodology to compute diffusivity and residence times on the surface of biomolecules. In particular our proposed algorithm to compute residence times appears to be better in dealing with poor statistics associated with the number of water molecules that remain on a surfaces for extended times. The type of linkage between monomers and the anomeric configuration all play a major role in determining the structure and dynamics of water on the surface of carbohydrates.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

The interactions of water and biomolecules play a key role in biology at the molecular level. They may affect the functions and structures of proteins and their derivatives. In several cases, the interactions intervene in the docking of ligands and inhibitors to proteins and glycoproteins. On the other hand, biomolecules also affect the structure and dynamics of water in their vicinities. These interactions directly determine the solvation of biomolecules, diffusion controlled reaction rate in biology, and the phase change of water. This thesis includes three major topics: water dynamics in the neighborhood of proteins; the relation of size, linkages, branching and secondary structures of oligosaccharides and the water dynamics in their neighborhood; the relation between the structures of cryoprotectant disaccharides and water structure in their neighborhood in the liquid phase and in crystallization.

In Chapter 2 we introduce atomic molecular dynamic simulations and mathematical methods proposed or applied to analyze the result of simulations. These methods include: a new way to compute residence times in situations of bad statistics; a modified way to compute surface diffusion; the study of rotational diffusion of water and its tetrahedrality, which is an index of the similarity to perfect ice.

In Chapter 3 the study of the diffusivity and residence of water and hydrogen peroxide in the proximity of peroxidase is reported. We compared two force fields for hydrogen peroxide with different partial charges, corresponding to different dipoles. It is interesting to see how the dipole change affects the residence, diffusivity and

spatial distribution.

In Chapter 4 we investigate water behavior on the surfaces of a diverse group of carbohydrates and attempt to explore the effect of size, linkages, branching and secondary structure on the dynamics and structure of water on the surface. We calculate residence times, local solvent occupancy numbers, rotational correlation functions, and diffusivities. We find that due to the secondary structure water residence times are longer and translation and rotational dynamics are retarded when in contact with wide helices and branched sugars. In the case of extended helixes and smaller oligosaccharides, water dynamics is faster and less hindered. This means that branching, anomeric configuration, and type of linkage between monomers all play critical roles in determining the structure and dynamics of water on the surface of carbohydrates.

In Chapter 5 we investigate the effect of two cryoprotectant carbohydrates on the structure of water in their vicinity both in the liquid phase and upon freezing. We do this to explore possible antifreeze mechanisms. It appears that in the liquid phase, water molecules in vicinity of trehalose tend to organize in a more tetrahedral form but in the ice phase, trehalose distorts tetrahedrality the most.

CHAPTER 2
METHODOLOGY

2.1 Atomistic Molecular Dynamics Simulations

Molecular dynamics(MD) is a methodology widely used to compute in atomistic detail trajectories of atoms and molecules. There are many different types of MD simulations that correspond to different thermodynamic ensembles. Some simulations are carried out at constant energy, number of particles, and volume(NVE), others at constant temperature, number of particles and pressure(NPT) and others at constant temperature, number of particles and volume(NVT). NPT and NVT simulations give correct ensemble average thermodynamic properties but NVE simulation reflects the actual dynamics of real physical systems. All that we do when running a NVE molecular dynamics simulation is to solve numerically Hamilton's equations of motion for the system. If the system is ergodic, trajectories can be used to compute thermodynamics averages.

Using state-of-the-art molecular dynamics packages one can simulate systems with hundreds of thousands of atoms. However, this scale is still much smaller than macroscopic sizes in which the number of atoms is of the order of the Avogadro's number. In order to overcome this obstacle, molecular dynamics uses periodic boundary conditions in which the system is infinitely replicated in each direction. Particles leaving the boundary from one end appear back in the simulation box for the opposite end. For this reason, only phenomena corresponding to a wavelength smaller than

half of the size of a simulation box can be captured using this procedure. In general, potential energy terms have the form as follows.

$$\begin{aligned}
U &= U_{stretch} + U_{bend} + U_{torsion} + U_{LJ} + U_{Coulomb} \\
U_{stretch} &= \sum_{bonds} K_r(r - r_{eq})^2 \\
U_{bend} &= \sum_{angles} K_\theta(\theta - \theta_{eq})^2 \\
U_{torsion} &= \sum_{dihedral\phi} \sum_{i=1}^{3} V_i[1 + cos(\phi + f_i)] \\
U_{LJ} &= \sum_{i<j} 4\epsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] \\
U_{Coulomb} &= \sum_{i<j} \frac{q_i q_j}{r_{ij}}
\end{aligned} \tag{2.1}$$

Many force fields have been developed to describe the potential energy between molecules.

2.2 Analysis of residence times and passage time distributions

2.2.1 The survival time correlation function and its pitfalls

The survival time correlation function was first introduced by Impey et. al. [57] in order to study the behavior of the hydration layer surrounding an ion. For an infinite long trajectory, this is defined as

$$N(mdt) = \frac{1}{N_f} \sum_{n=1}^{N_f} \sum_{j=1}^{N_{\rm sys}} p_j(ndt, mdt; dt). \tag{2.2}$$

In a computer simulation this function is more precisely defined in the following way:

$$\begin{aligned} N(mdt) &= \frac{1}{N_f - m} \sum_{n=1}^{N_f - m} \Big(\sum_{j=1}^{N_{\text{surface}}(ndt)} p_j(ndt, mdt; dt) \Big) \\ &= \frac{1}{N_f - m} \sum_{j=1}^{N_{\text{sys}}} \sum_{n=1}^{N_f - m} p_j(ndt, mdt; dt) \end{aligned} \tag{2.3}$$

Here dt is the recording interval, N_f is the total number of trajectory frames recorded, ndt is the starting point of each sliding time window, mdt is the function's variable (i.e. the size of each sliding time window) and $p_j(ndt, mdt; dt)$ takes the value of 1 if molecule j is within the coordination shell at times ndt as well as $(ndt + mdt)$ and has not left the coordination shell for any contiguous period of time longer than dt; otherwise its value is 0.

This survival time correlation function, when correctly converged, quantifies the average number of molecules which at time zero are within the coordination shell of a given object and remain there without ever leaving (within the resolution time dt) at least until time mdt. $N(mdt = 0)$ corresponds to the average coordination number of the object. $N(mdt)$ is always smaller or equal to $N(0)$. The summation index j can either run through all solvent molecules considered (in our case all peroxide molecules or all water molecules) or over all molecules inside the solvation shell at time ndt. This is true because $p_j(ndt, mdt; dt) = 0$ if the j^{th} solvent molecule is not inside the shell at time ndt.

$N(mdt)$ can be normalized to yield the average fraction of solvent molecules

originally within the solvation shell that remain in this region (without ever leaving) for an interval of time at least as long as mdt.

$$f(mdt) = N(mdt)/N(0) \tag{2.4}$$

Different authors have analyzed the behavior of this time correlation function in the case of water solvated proteins (see for example references [13, 18, 88]). If this function is fitted to a single exponential decay, the time constant is commonly described as the "residence time" [57, 51]. Garcia et. al. [51] have applied Impey's formula [57] to the case of water molecules surrounding a protein on a residue base and found that a single exponential was not sufficient to parametrize the data. This has also been observed by several other authors [78, 102, 103]. Some authors have claimed that this function is better analyzed using stretched exponentials or at least multiple exponentials.

When $P(t)$ is well described by two exponentials [13, 18, 88],

$$f(t) = A\exp(-t/\tau_s) + B\exp(-t/\tau_l) \tag{2.5}$$

the decay constants τ_s and τ_l correspond to a fast boundary recrossing process and a more interesting surface residence process. [13, 18, 88, 54] Several authors have calculated survival probability functions on a residue basis. Analysis of simulations in the case of crambin [71], plastocyanin [102], and azurin [72] show, as is to be expected, that the residence time changes according to the chemical character of the residue, $\tau_{\text{charged}} \geq \tau_{\text{polar}} \geq \tau_{\text{nonpolar}}$. In this thesis, we are only interested in the surface averaged survival function but the methodological developments described in

the following subsections apply equally well to the study of properties on a residue or protein region basis.

2.2.2 Probability distribution of first passage times

The concept of first passage times has also been applied in the past to study the mean time that a solvent molecule remains within the solvation shell of a protein. [99, 28, 84].

Here we define the unnormalized distribution of passage times as

$$H(m) = \sum_{j=1}^{N_{\text{sys}}} \sum_{n=1}^{N_f - m - 1} [1 - A_j(n-1)][1 - A_j(n+m+1)] \prod_{k=n}^{n+m} A_j(k). \tag{2.6}$$

Here N_{sys} represents the total number of solvent molecules (or group of molecules of interest) in the system. The function $A_j(k) = 1$ if molecule j lies inside the shell at time $k * dt$. In all the other cases, $A_j(k) = 0$. For simplicity, in subsequent equations and in eqn 2.6 we drop the argument dt. Clearly this distribution of passage times increases with simulation time. A time normalized distribution is defined as

$$\tilde{H}(m) = \frac{H(m)}{N_f - m} \tag{2.7}$$

In the limit of infinite simulation time ($N_f \to \infty$), $N(m)$ can be reconstructed from the histogram of first passage times $\tilde{H}(m)$. When ($M \geq m$), a solvent molecule with first passage time M will contribute $M-m+1$ times to $N(m)$. Hence, the contribution of all solvent molecules with first passage times of M is $(M - m + 1)H(M)$. A sum

over all possible passage times M yields

$$N(m) = \sum_{M=m}^{M=N_f-2} (M - m + 1)\tilde{H}(M) \tag{2.8}$$

This equation is only strictly valid for infinitely long runs, otherwise particles remaining on the surface of a protein past the end of the simulation or particles that entered the solvation shell before time zero are neglected. The upper limit is explicitly written as $(N_f - 2)$ to emphasize that solvent molecules must have passage times that are at least $2 \times dt$ shorter than the full duration of the MD run in order to be counted. Clearly, so long as $\tilde{H}(M)$ is non-negligible, because of the term $M - m + 1$ in Eq. 2.8, solvent molecules that remain on the surface for longer times (i.e. with large M) will have larger weights on the short time behavior (small m) of the survival function $N(m)$. From a practical computational perspective we have found that for a given protein, $N(m)$ and correspondingly $f(m)$ can be quite different for simulations with the same average pressure and temperature. This is particularly true even at very short times $m * dt$. A single molecule staying inside the hydration shell for a long time contributes an amount equivalent to that of hundreds or thousands of molecules that enter and leave the shell for a short period of time. The problem with this is that for typical finite time computer simulations the statistics on solvent molecules that remain within the solvation shell of a protein for long times is extremely poor! From experiments we know that some water molecules that are not "permanently bound" to the protein can still remain in its vicinity for times much longer than $500ps$ [83, 125, 124, 87]. In our simulations we found that different runs yield different number of long staying molecules. Clearly the long time behavior of this function

only depends on the statistics of those long staying molecules. Therefore both long time behavior and short time behavior are strongly affected by molecules on which one has extremely poor statistics!

A filtering threshold time after which molecules in a simulation are deemed permanently bound to the protein and not counted is also not appropriate since there is no obvious definition for this time. Survival probability functions for the same run look widely different depending on which threshold time is chosen[31].

Instead, it is useful to define an expression for the cumulative probability of first passage times.

$$\tilde{H}_{accum}(\geq m) = \sum_{m'=m}^{N_f-2} \tilde{H}(m') = \sum_{m'=m}^{N_f-2} \frac{H(m')}{N_f - m'} \tag{2.9}$$

Substituting Eq. 2.6 into Eq. 2.9, we obtain

$$\tilde{H}_{accum}(\geq m) = \sum_{j=1}^{N_{\rm sys}} \sum_{m'=m}^{N_f-2} \frac{1}{N_f - m'} \sum_{n=1}^{N_f-m'-1} [1-A_j(n-1)][1-A_j(n+m'+1)] \prod_{k=n}^{n+m'} A_j(k) \tag{2.10}$$

The single particle survival correlation function $p_j(ndt, mdt; dt)$ in Eq. 2.3 can be represented in terms of A_j as follows,

$$p_j(ndt, mdt; dt) = \prod_{k=n}^{n+m} A_j(k) \tag{2.11}$$

Using this relation, the histogram $H(m)$ of first passage times in Eq. 2.6 can be written as

$$H(m) = \sum_{j=1}^{N_{sys}} \sum_{n=1}^{N_f-m-1} [1 - A_j(n-1)][1 - A_j(n+m+1)] p_j(n, m) \tag{2.12}$$

Correspondingly, the cumulative probability distribution of first passage times in Eq. 2.10 can be rewritten as

$$\tilde{H}_{accum}(\geq m) = \sum_{j=1}^{N_{sys}} \sum_{m'=m}^{N_f-2} \frac{1}{N_f - m'} \sum_{n=1}^{N_f-m'-1} [1 - A_j(n-1)][1 - A_j(n+m'+1)]p_j(n, m') \tag{2.13}$$

Finally we define the fraction of molecules with first passage time greater or equal to mdt as

$$\tilde{f}(\geq m) = \frac{\tilde{H}_{accum}(\geq m)}{\tilde{H}_{accum}(\geq 0)} \tag{2.14}$$

Notice that in this equation only particles than enter and leave the solvation shell are counted, therefore permanently bound molecules do not enter this equation. In Eq. 2.14, a particle with passage time $M \geq m$ is only counted once. One would therefore expect that the influence of poor statistics on molecules with very long passage times will affect the short time behavior of $\tilde{f}(\geq m)$ very little. This is in contrast to the case of $f(m)$ in which the contribution is $M - m + 1$.

2.2.3 The mathematical relation between the survival and first passage time correlation functions

In order to obtain the mathematical relationship between these two functions, it is advantageous to take the limit $dt \to 0$. The function $p_j(m, n, dt)$ can be written as:

$$\begin{aligned} p_j(m, n, dt) &= e^{\ln\left(\prod_{k=n}^{n+m} A_j(kdt)\right)} \\ \lim_{dt\to 0} e^{\sum_{k=n}^{n+m} \ln(A_j(k*dt))} &= e^{\int_{t_0}^{\tau+t_0} \ln(A_j(t))dt} = p_j(t_0, \tau) \end{aligned} \tag{2.15}$$

Here $t_0 = ndt$, $\tau = m * dt$ and $A_j(t)$ is zero if particle j is outside of the solvation shell at time t, otherwise its value is one. With these definitions the survival time correlation function becomes

$$N(\tau) = \frac{1}{T_f - \tau} \sum_{j=1}^{N_{sys}} \int_0^{T_f - \tau} dt_0 \; p_j(t_0, \tau) \tag{2.16}$$

where $T_f = N_f * dt$. By explicit evaluation of the product in eq. 2.6 we obtain

$$H(\tau) = \sum_{j=1}^{N_{sys}} \int_{dt}^{T_f - \tau - dt} \{p_j(t_0, \tau) - p_j(t_0 - dt, \tau + dt) - p_j(t_0, \tau + dt) + p_j(t_0 - dt, \tau + 2 * dt)\} \, dt_0 \tag{2.17}$$

In the continuous limit we can drop the time differentials dt from the integration limits. Furthermore since we are integrating over all possible values of t_0 the first argument in the second and fourth terms under the integral can simply be replaced by t_0. The only relevant variable is the second argument of p_j which is the first passage time. We therefore have

$$H(\tau) = \sum_{j=1}^{N_{sys}} \int_0^{T_f - \tau} \{p_j(t_0, \tau) - 2 * p_j(t_0, \tau + dt) + p_j(t_0, \tau + 2 * dt)\} \, dt_0 \tag{2.18}$$

This clearly corresponds to

$$H(\tau) = \sum_{j=1}^{N_{sys}} \int_0^{T_f - \tau} \frac{\partial^2 p_j(t_0, \tau + dt)}{\partial \tau^2} * \Delta^2 dt_0 \tag{2.19}$$

where Δ is a time differential. This equation can be approximated as

$$H(\tau) = \Delta^2 * \sum_{j=1}^{N_{sys}} \int_0^{T_f - \tau} \frac{\partial^2 p_j(t_0, \tau)}{\partial \tau^2} dt_0 \tag{2.20}$$

It will be clear from our results section that the functions $\tilde{H}(\tau)$ and $N(\tau)$ are very rapidly decaying function of τ. The simulation time T_f is commonly much longer

(on the order of nanoseconds) than the relevant values of τ (on the order of a few picoseconds) for which these functions have an appreciable value different from zero. It is therefore a good approximation to drop τ from the upper limit of the integrals in eqn.2.16 and eqn.2.20 as well as from the denominator $\frac{1}{T_f-\tau}$ in eqn. 2.16. Under these approximations we have

$$\tilde{H}(\tau) = \frac{\Delta^2}{T_f} * \frac{\partial^2 \sum_{j=1}^{N_{sys}} \int_0^{T_f} p_j(t_0,\tau)dt_0}{\partial\tau^2} = \Delta^2 \frac{\partial^2 N(\tau)}{\partial\tau^2} \tag{2.21}$$

Finally,

$$\tilde{H}_{accum}(\geq \tau) = \int_\tau^{T_f} \tilde{H}(\tau_0)d\tau_0 = \Delta^2 \left(\frac{\partial N(\tau_0)}{\partial \tau_0} \mid_{\tau_0=T_f} - \frac{\partial N(\tau_0)}{\partial \tau_0} \mid_{\tau_0=\tau} \right) \tag{2.22}$$

Since $N(t)$ commonly decays exponentially or as a sum of decaying exponentials for large T_f eqn. 2.22 becomes

$$\tilde{H}_{accum}(\geq \tau) = -\Delta^2 \frac{\partial N(\tau)}{\partial \tau} \tag{2.23}$$

Finally, we have

$$\tilde{f}(\geq \tau) = \frac{\tilde{H}_{accum}(\geq \tau)}{\tilde{H}_{accum}(\geq 0)} = \frac{\frac{\partial N(\tau)}{\partial \tau}}{\frac{\partial N(\tau)}{\partial \tau} \mid_{\tau=0}} = \frac{\frac{\partial f(\tau)}{\partial \tau}}{\frac{\partial f(\tau)}{\partial \tau} \mid_{\tau=0}} \tag{2.24}$$

Equation 2.24 is one of the important results in this thesis which shows the relationship between the fraction of molecules with passage times longer or equal to τ and the survival probability time correlation function. As we will demonstrate in subsequent sections, it turns out that $\tilde{f}(\geq \tau)$, the fraction of molecules with first passage time greater or equal than τ is easier to converge computationally than $f(\tau)$ the fraction of molecules originally within the solvation shell that never leave during an interval

at least as long as τ. From the equations derived above one could compute $\tilde{f}(\geq \tau)$ and if desired obtain the corresponding time constants for $f(\tau)$.

2.3 Surface Diffusion

2.3.1 Bulk Diffusion and the Einstein Relation

In analyzing bulk diffusion, the solvent mobility is conveniently described by diffusion coefficients and the mean square displacement of a particle in a liquid. [5, 13, 18, 54, 17, 75, 84].

The macroscopic term "matter flux" is defined as the "number of particles passing through a unit area during a unit time interval. Experimental observations of transport properties show that the flux in the direction of a certain coordinate is proportional to the first derivative of the concentration with respect to that coordinate.

$$J = D \cdot \frac{\partial c}{\partial x} \tag{2.25}$$

where J: matter flux, D: diffusion coefficient, c: concentration, x: the coordinate

The equation above is called Fick's first law of diffusion. By combining Fick's law with the equation for conservation of matter:

$$\frac{\partial c(\vec{r},t)}{\partial t} + \nabla \vec{J}(\vec{r},t) = 0 \tag{2.26}$$

one obtains:

$$\frac{\partial c(\vec{r},t)}{\partial t} - D \cdot \nabla^2 c(\vec{r},t) = 0 \tag{2.27}$$

With the definition of the mean square displacement of a particle, the Einstein relationship in d dimensions can be deduced. [5]

$$D = \frac{1}{2d} \lim_{\Delta t \to \infty} \frac{< \Delta r^2(t) >}{\Delta t} \tag{2.28}$$

At times long compared to typical correlation scales in the velocity auto-correlation function, $< \Delta r^2(t) > \propto 2dDt$. At very short times $< \Delta r^2(t) > \propto t^2$ but in supercooled liquids or in general glassy systems an intermediate cage regime bridges the gap between the fast ballistic behavior and the long term diffusive regime.

2.3.2 Diffusion on Arbitrary Complex Surfaces

The calculation of the diffusive behavior of solvent molecules in the hydration shell of a macromolecule is problematic since only particles in the vicinity of the surface should be considered but most particles constantly enter and leave the first solvation shell. Therefore if one only looks at the behavior in the first solvation shell, the statistics for the mean square displacement are only good at short times. For longer times one can only take into account the few particles that remain on the vicinity of the protein for extended periods. Two problems arise because of this, poor statistics and biasing in favor of the slow diffusive particles. In a way the long time behavior of the mean square displacement only represents the dynamics of those particles strongly bound to the surface.

We calculate the mean square displacement (MSD) on a surface as follows,

$$\langle \Delta r(t)^2 \rangle = \frac{1}{N_w\prime(t)} \sum_{t_0}^{T_{run}-t} \sum_{j=1}^{N_w(t_0,t)} [r_j(t+t_0) - r_j(t_0)]^2 \tag{2.29}$$

where N_w is the number of water molecules inside the hydration shell at t_0, which remained so until $t_0 + t$, and

$$N_w\prime(t) = \sum_{t_0}^{T_{\rm run}-t} N_w(t_0, t). \tag{2.30}$$

Notice that our formula in Eq. 2.29 is different from the previously used in Ref. [78] which is displayed as below,

$$\langle \Delta r(t)^2 \rangle = \frac{1}{T_{\rm run} - t + 1} \sum_{t_0=0}^{T_{run}-t} \frac{1}{N_w(t_0, t)} \sum_{j=1}^{N_w(t_0,t)} [r_j(t + t_0) - r_j(t_0)]^2. \tag{2.31}$$

Equation 2.31 first takes an average over all particles which are in the solvation shell during the period t_0 t_0+t and then performs an average over all possible initial points t_0. Equations 2.29 and 2.31 should be almost identical in the case of a solute in a single solvent since $N_w(t_0, t)$ should be nearly constant independent of the time origin t_0 and only dependent on the time difference t. Consequently,

$$\frac{1}{(T_{\rm run} - t + 1) \times N_w(t_0, t)} \approx \frac{1}{N_w\prime(t)}.$$

This however is not the case for our study of the system of a protein in a 5% solution of hydrogen peroxide in chapter 3. In the case of a solution with more than one solvent component the number of peroxide molecules $N_{per}(t_0, t)$ can be different for different initial t_0 values. Therefore an average of an average as in Eq. 2.31 is not accurate and Eq. 2.29 should be used instead.

2.4 Rotational Correlation Functions

Similar to translational correlation function (mean square displacement) of solvent discussed in last section, rotational correlation function reflects solvent mobility

in rotation. The autocorrelation function in rotation is defined as:

$$C(t) = \frac{< \vec{v}(t_o + t) \cdot \vec{v}(t_0) >}{< \vec{v}(t_o) \cdot \vec{v}(t_0) >}. \tag{2.32}$$

where $\vec{v}$ is the unit vector directed along the dipole moment vector of polar solvent (water) at time t.

2.5 Parameters Characterizing Water Structure

In our study of water structure in chapter 5, we calculate radial distribution functions of oxygen on water and tetrahedralities of the oxygen network. Here we briefly introduce radial distribution functions and the calculation of tetrahedrality.

Radial distribution functions (RDF, also called pair correlation functions) are commonly used in the analysis of the structure of a liquid. A radial distribution function describes how the particle density varies as a function of the distance from one particular particle.

Water at low temperature or at room temperature in a hydration shell has tetrahedral ordering. Order parameters specific for tetrahedral configurations are introduced in this section.

2.5.1 Order Parameters for Tetrahedral Configuration

The geometric study of liquid structure was pioneered by Bernal. [15] He suggested using the Voronoi tessellation to describe the structure of simple liquids. Subsequently, various order parameters have been proposed to describe the structure of

simple molecules and their phase transitions [110, 73, 27, 22].

Order parameters specific for tetrahedral configurations are of specific interest for studying liquid water. Conceptually, ideal ice is of perfect tetrahedral nature. The four nearest neighboring waters to each water molecule are located at the four vertices of a tetrahedron because of the angles or relative orientations of the O-H bonds and lone pairs. These bonds and lone pairs are responsible for hydrogen bonding, which is important for the structure of liquid water.

Steinhardt proposed quadratic and the third-order invariants formed from bond spherical harmonics to measure cluster symmetries in supercooled liquid [110]. For a tetrahedral configuration there are five points: the four vertices of the tetrahedron and the center. Chau proposed the use of six angles to calculate a quantity S_g and the use of four distances between central atoms and vertices to calculate S_k in order to measure the tetrahedrality of ice I_h in aqueous solutions [27]. The mathematical form of S_k is:

$$S_k = \frac{1}{3}\sum_{k=1}^{4}\frac{(r_k - \bar{r})^2}{4\bar{r}^2} \tag{2.33}$$

The mathematical form of S_g is:

$$S_g = \frac{3}{32}\sum_{k=1}^{4}\sum_{j>k}(cos\psi_{ijk} + \frac{1}{3})^2 \tag{2.34}$$

where ψ_{ijk} is the angle between a central oxygen atom i and its neighboring oxygen atoms j and k. This summation is over the six possible O-O-O angles involving the central molecule i and pairs of its four nearest neighbors. The factor of 3/32 normalizes S_g to the range $0 \leq S_g \leq 1$. S_g is zero if the cosines of all angles are -1/3,

i.e., the tetrahedron is perfect. S_g is unity if all four vectors are overlapping.

Debenedetti rescaled S_g and defined a new parameter q in the following equation.

$$q_i = 1 - \frac{3}{8}\sum_{k=1}^{4}\sum_{j>k}(cos\psi_{ijk} + \frac{1}{3})^2 \tag{2.35}$$

Here $< q >= 1$ corresponds to perfect tetrahedral structure; $< q >= 0$ corresponds to the case of an ideal gas, in which the four vectors arrange randomly. In most of cases q_i falls in the interval (0,1). However, it is possible to have negative value of q_i [44].

CHAPTER 3
DYNAMICS OF WATER AND HYDROGEN PEROXIDE IN CROWDED PROTEIN ENVIRONMENTS

In Chapter 2, we described the survival time correlation function, first passage time correlation function and their mathematical relation. In this chapter we apply these two methods to the study of water dynamics on the surface of a protein to demonstrate that the first passage time correlation function is more robust in characterizing the time a particle stays on an arbitrary surface. We can solve the problem caused by poor statistics of long-term staying particles by using the first passage time correlation function. Besides, we also report residence times and surface diffusion coefficients of water and hydrogen peroxide.

3.1 Introduction

Reactive oxygen species (ROS) have important functions in cell signaling and, when present at overly high levels, may cause oxidation of important biological molecules. Among ROSs, Hydrogen peroxide is commonly used as an index of oxidative stress [134, 101, 38, 11, 123]. Kinetic models to study diffusion of hydrogen peroxide and other ROS species inside of mitochondria often assume dynamics similar to that in solutions. Without a better model for diffusion, researchers often assume that the rate of appearance scales with the gradient across cell boundaries. Furthermore the diffusion constant and permeability coefficients are taken to be the same or very similar to those of water. See citation [24, 107] and references therein. However,

it is well-known that separation of proteins in the cytosol or inside of mitochondria, where ROS are most predominant, can be smaller than 1 nm. In an intracellular matrix there are sources and sinks of ROS and since environments are crowded, motion is most similar to a percolation process[24, 107] . In this thesis we study similarities and differences in the structure and dynamics of water and peroxide when in contact with proteins. The protein we have chosen is PRDX5, a small globular peroxiredoxin present in mitochondria.

Residence time calculations have shed light on the time scales on which water molecules stay inside the hydration layer of a protein. They also provide insight into the structure and dynamical behavior of inter-facial water in the hydration shells of different proteins [13, 12, 95, 88, 18, 57]. Many studies have shown that the dynamical behavior of water molecules closely associated with proteins is not simply diffusive. Experimentally, the residence time is extracted from time correlation functions or derived from dielectric dispersion measurements [13]. Early studies from nuclear Overhauser effect provided rather long residence times for water on the surface of proteins with an upper limit of 500 ps [13, 87]. Later it was pointed out that this result might be incorrect due to an overestimation caused by water molecules strongly bound to proteins [52]. More recent studies from magnetic relaxation dispersion gave much shorter times in the range of 10-50 ps [83]. Clearly, as it has been shown in different simulation studies [94, 29, 102, 51, 71, 21], the residence time will depend on factors such as the local polarity of the surface, and its curvature. Residence time at different locations on the surface can be quite different.

In computer simulations, the residence time is usually obtained from the single or multiple exponential time constant fitted to the decay of the survival time correlation function [18, 102, 90, 16, 45, 72]. As we showed in Chapter 2, water molecules with long residence times have important contribution to the short time behavior of this correlation function. The problem from a computational perspective is that one has very poor statistics on molecules that are not"permanently bound" [87, 83, 124, 125] to the protein but that remain on the surface for hundreds or thousands of picoseconds.

A different method used in order to study the time scale on which a solvent molecule remains in the vicinity of a protein is the calculation of mean first passage times [99, 28, 84].

We have introduced these two methods in Chapter 2 and also derive expressions linking these two functions. We also discussed the algebra involved in each calculation and found that the distribution of first passage times is a much more reliable tool to study the short time dynamics on the surface of proteins. In this chapter, we will compare real data and show that the results based on the probability distribution of first passage time are reproducible across simulations and mostly independent of those molecules that remain on the surface of the protein for long times for which one has poor statistics.

The effects induced by a protein on its surrounding liquid environment have been studied in detail in the past. Local information correlating with the effect of different residues, surface curvature [94, 29, 102, 51, 71, 21] and polarity have provided

important information[133, 26] for the understanding of folding, drying, hydrophobic hydration and the kinetics of aggregation. The self-diffusion coefficient in the vicinity of a protein surface has been shown to exhibit a wide range of dynamic behaviors [13, 88, 18]. Many studies [13, 18, 54, 78, 75, 108, 113, 20, 69] have investigated the effect of different protein groups on the motility of surrounding water, in particular, how hydration water diffusion is affected by polar, nonpolar or charged protein groups. Our approach here is different, in our study, the protein is regarded as a whole and the water and peroxide surrounding it are studied as a function of the distance from the surface. The goal is to understand the difference between water and hydrogen peroxide in regards to their residence in the vicinity of a protein or the intervening space between proteins. Since water and hydrogen peroxide are quite similar in size and composition it is important to understand whether there is an enrichment of peroxide on the surface of a protein and if the diffusivity of water and peroxide can be regarded as identical as has been done in the past in kinetic models.[24, 107]

3.2 Force Field and Simulations

3.2.1 Force Field

Unlike water for which many accurate potential energy functions exist for simulating bulk behavior, for hydrogen peroxide different models exist but none has been fully tested against thermodynamic data in water solutions. Simulations can be found with partial charge for oxygen ranging from -0.36 to -0.54 in units of electrons. [4, 3, 118, 77].

In order to be sure that our results are robust and model independent we have used two different models. Model 1 with oxygen partial charge of -0.46[31] and Model 2 with oxygen partial charges of -0.36 derived from reference [3] in which DFT calculations have been performed to fit the potential energy surface for the $H_2O - H_2O_2$ dimer. For water we used the flexible SPC model [46] and for the protein the OPLS-AA [59] force field.

Other terms in the peroxide potential were constructed in the following way: bend, stretch and Lennard Jones parameters were adopted from the standard AMBER[64] library. For intermolecular interactions the set of point charges listed in Table 3.1 were used. Intramolecular interactions were obtained from high level calculations reported in reference [65]. All intramolecular interactions except for the bending and stretching potential are accounted for by using the Rykaert Bellemans potential form [65]. Table 3.1 summarizes all force field parameters for hydrogen peroxide. Standard geometric combination rules were used to determine Lennard Jones parameters.

	Bond/angle/diherdral	Coefficient/equilibrium position	Source
Stretch	OO bond	$K_r = 294440$ KJ nm^2/mol $r_{\text{eq}} = 0.14600$ nm	AMBER
	OH bond	$K_r = 462308$ KJ nm^2/mol $r_{\text{eq}} = 0.09600$ nm	AMBER
Bend	HOO	$K_\theta = 414.6$ KJ degree^2/mol $\theta_{\text{eq}} = 98.72$ degree	AMBER
Torsion (Rykaert Bellemans)	HOOH	$C_0 = -7.7099$ KJ/mol $C_1 = -10.7269$ $C_2 = -15.2261$ $C_3 = -2.2442$ $C_4 = -0.25839$ $C_5 = 0.00$	Ref. [65]
Coulomb	Charges on atom	$O_1 = -0.460$ $O_2 = -0.460$ $H_1 = 0.460$ $H_2 = 0.460$	3-21G* with given initial coordinates by Koput [65]

Table 3.1: Force field parameters for hydrogen peroxide.

3.2.2 Simulations

In order to compare bulk hydrogen peroxide solution behavior with the situation in a crowded protein environment three different types of simulations were performed.

1. Aqueous hydrogen peroxide solution (System A): The system consisted of 5812 water molecules and 306 hydrogen peroxide molecules in a periodically replicated simulation box with approximate dimensions of $40 \times 40 \times 40\text{Å}^3$. The concentration of hydrogen peroxide was approximately 5% (1 peroxide molecule for each 19 water molecules).

2. Periodically replicated single protein solution (System B): a PRDX5 protein was soaked in a simulation box of approximate dimensions: $57.6 \times 56.6 \times 63.6\text{Å}^3$, containing 5673 water molecules and 297 hydrogen peroxide molecules (peroxide concentration close to 5%). The initial structure of the protein was obtained from the Protein Data Bank entry corresponding to Ref. [38].

3. Two-protein system (System C): two identical PRDX5 proteins were fixed in space by position constraints of all non-hydrogen atoms in aqueous hydrogen peroxide solution of 5% concentration with a box of the following approximate dimensions: $107.5 \times 56.8 \times 64.0\text{Å}^3$. The inter-protein separation was $17\mathring{A}$. The number of peroxide and water molecules was 574 and 10893, respectively.

In all simulations, the initial equilibration time in the constant pressure, temperature and number of particles (NPT) ensemble was 200 picoseconds. Following

this, data was collected from constant energy simulations after 100 ps of equilibration in the constant energy (NVE) ensemble. The initial temperature in all cases was set at 300K and the average temperature during NVE runs remained close to this value. In order to compare the accuracy of the residence time calculations, we simulated two replicas of System B with different initial conditions. Each production run used for data collection was 2.5 nanoseconds in duration. A water or peroxide molecule was defined to be inside the hydration shell if the distance between its center of mass and any atom of protein was smaller than 4 $\mathring{A}$. This is consistent with first solvation shells obtained from radial distribution g(r) between atoms in the surface of the protein and the solvent.

3.3 Results

3.3.1 Comparison and Validation of Methodology

In subsection 2.2 we derived Eq. 2.24 which establishes an approximate relationship between the normalized survival probability function $f(\tau)$ corresponding to the fraction of solvent molecules around an object that are in its solvation shell at time zero and remain there without leaving for at least a time τ with that of $\tilde{f}(\geq \tau)$ the fraction of solvent molecules with first passage time greater or equal to τ. In order to establish the numerical quality of this approximation we have computed the first derivative of $f(\tau)$ and also $\tilde{f}(\geq \tau)$ for water around PRDX5. The results are shown in Fig. 3.1. Clearly $\tilde{f}(\geq \tau)$ is an excellent approximation to the derivative of $f(\tau)$ over all relevant values of the variable τ.

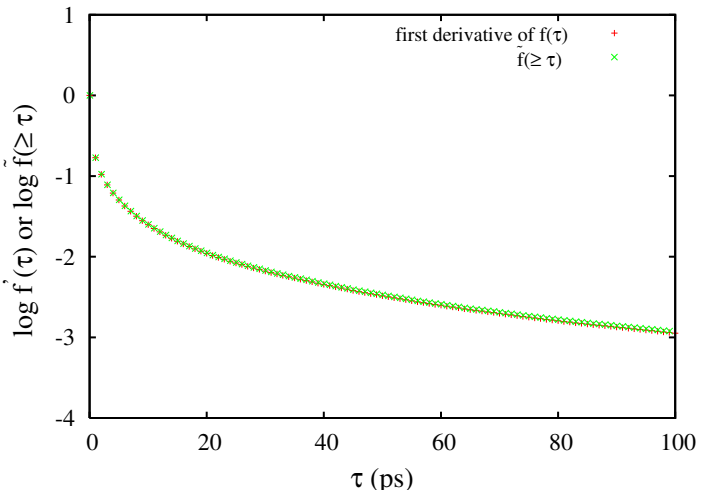


Figure 3.1: First derivative of the normalized survival time correlation function ($f(\tau)$) and fraction of water molecules with first passage time larger or equal to τ ($\tilde{f}(\geq \tau)$) in the first hydration shell of PRDX5 (System B).

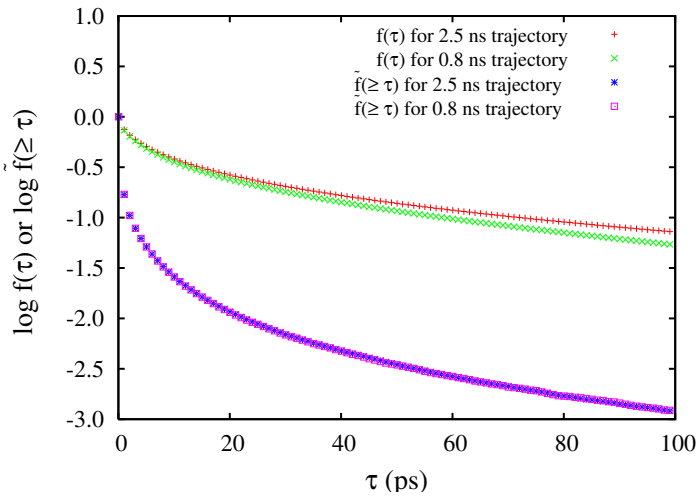
In section 2.2, we stressed the need to work with the cumulative distribution of first passage times as opposed to the survival probability correlation function because of the large fractional contribution that solvent molecules with long residence times, for which statistics are very poor, can have. Fig. 3.2(a) and Fig. 3.2(b) clearly support this analysis. Fig. 3.2(a) shows that $\tilde{f}(\geq \tau)$ is almost identical when analyzed over a period of 0.8 ns and 2.5 ns, while $f(\tau)$ is not. The trajectory length has much less of an effect on $\tilde{f}(\geq \tau)$. Clearly the effect of solvent molecules with large residence times strongly affects one function and not the other. The same phenomenon can be observed from the analysis of two independent molecular dynamics simulations with the same length but different number of solvent molecules with long residence times. Fig. 3.3 shows that $\tilde{f}(\geq \tau)$ is almost trajectory independent while $f(\tau)$ is clearly trajectory dependent. Fig. 3.4 shows the fractional contribution to $f(\tau)$ of

water molecules with first passage times larger than 1ns in two independent molecular dynamics trajectories. Clearly while for the fully converged $f(\tau)$ these should be identical, it is clear that contributions are very different across trajectories. Since $f(\tau)$ and $\tilde{f}(\geq \tau)$ are simply linked by differentiation, it only makes sense to work with $\tilde{f}(\geq \tau)$, the one that is less prone to statistical error.

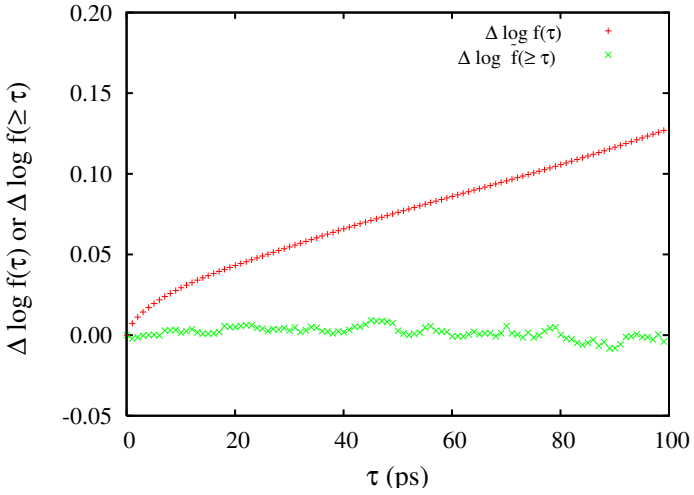
3.3.2 Comparison between water and hydrogen peroxide

Figure 3.5 shows $\tilde{f}(\geq \tau)$ for water and the two models for hydrogen peroxide investigated in this paper. It is clear from this plot that, the fraction of water molecules with first passage times larger than τ is always smaller than that of peroxide independent of the model used for peroxide and independent of τ. All the curves in Fig. 3.5 display, fast and a slow components reflected in the corresponding initial and asymptotic slopes of $\log \tilde{f}(\geq \tau)$. Time constants are tabulated in Table 3.2. We have made no attempt to fit the whole range of $\tilde{f}(\geq \tau)$, which appears from figure 3.5 not to be the sum of two simple exponentials. Instead we have computed time constants by fitting the apparent initial and final asymptotic behavior of $\tilde{f}(\geq \tau)$. τ_s is obtained from a fit in the range 0.5-5 ps while τ_l is obtained from the range 30-100 ps. τ_s only increases slightly as one increases the thickness of the shell considered, no obvious difference exist between H_2O and H_2O_2. On the other hand, in all cases, increasing the shell thickness results in a much larger τ_l and it is obvious that independent of the model, hydrogen peroxide appears to have a larger value of τ_l than water. This

indicates that peroxide remains longer on the surface of the protein than water.

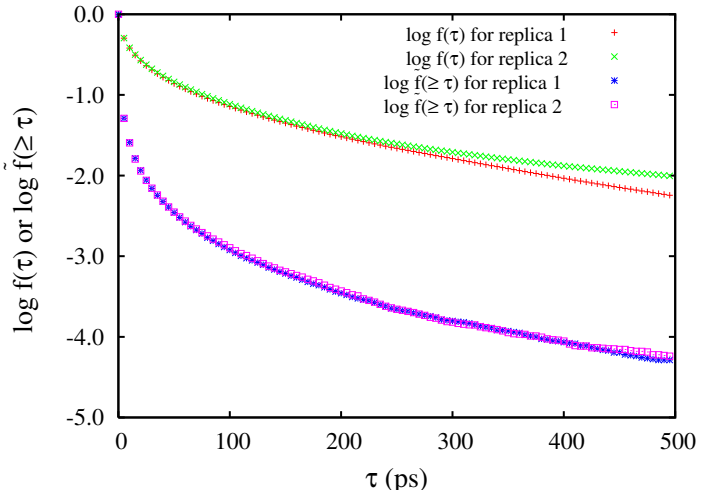


(a) semi-log plots of $f(\tau)$ and $\tilde{f}(\geq \tau)$

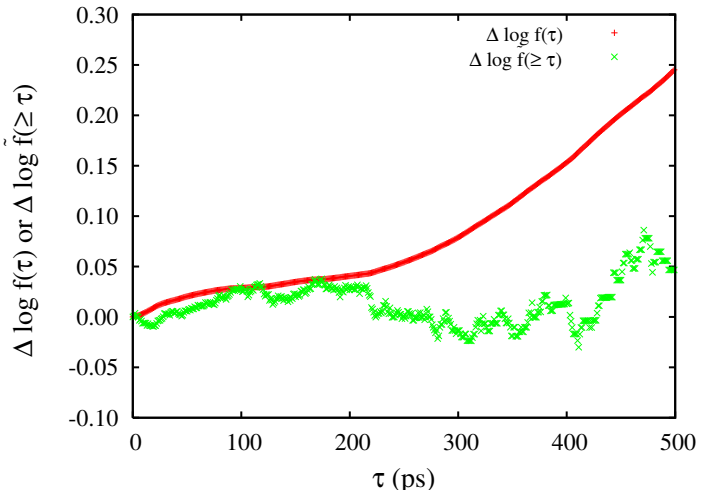


(b) linear plots of $\Delta \log f(\tau)$ and $\Delta \log \tilde{f}(\geq \tau)$

Figure 3.2: Comparison between the normalized survival time correlation functions $f(\tau)$ and the fraction of solvent molecules with first passage time greater or equal to τ, $\tilde{f}(\geq \tau)$ for the first hydration shell (4 $\mathring{A}$) in System B. Figure a corresponds to results of a single trajectory analyzed using its first 0.8 ns and its full 2.5 ns duration. Figure b shows the variation in $f(\tau)$and $\tilde{f}(\geq \tau)$. Obviously $\tilde{f}(\geq \tau)$ has a much more robust behavior as a function of time.



(a) semi-log plots of $f(\tau)$ and $\tilde{f}(\geq \tau)$



(b) linear plots of $\Delta \log f(\tau)$ and $\Delta \log \tilde{f}(\geq \tau)$

Figure 3.3: Comparison between the normalized survival time correlation functions ($f(\tau)$) and the fraction of solvent molecules with first passage time greater or equal to τ, $\tilde{f}(\geq \tau)$ for the first hydration shell (4 $\mathring{A}$) in System B. Figure a corresponds to results of two MD trajectories with the same length (2.5 ns) but different initial configurations. Figure b shows the variation in $f(\tau)$and $\tilde{f}(\geq \tau)$ from these two trajectories.

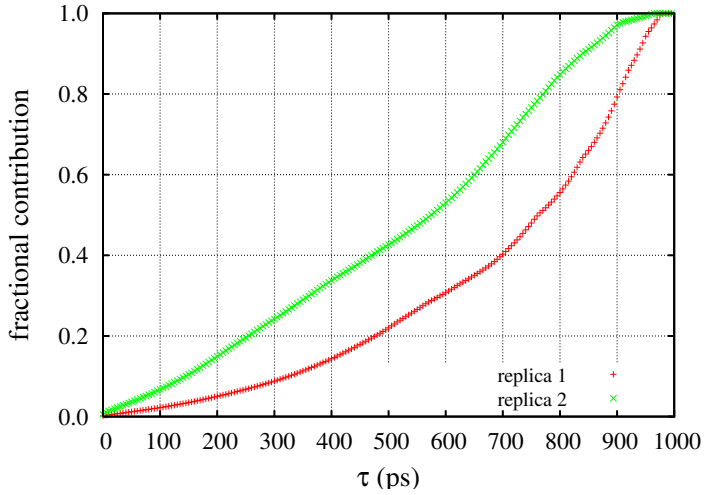


Figure 3.4: Fractional contribution to $f(\tau)$ of water molecules with first passage time larger than 1ns compared for two independent molecular dynamics trajectories.

thickness	H_2O		H_2O_2 model 1		H_2O_2 model 2	
(Å)	τ_s	τ_l	τ_s	τ_l	τ_s	τ_l
$0-4$	2.3 ± 0.06	39.3 ± 0.10	2.2 ± 0.05	46.6 ± 0.18	2.4 ± 0.06	40.8 ± 0.10
$0-6$	2.5 ± 0.05	52.9 ± 0.08	2.4 ± 0.05	71.7 ± 0.21	2.7 ± 0.05	66.5 ± 0.10
$0-8$	2.8 ± 0.07	64.5 ± 0.14	2.7 ± 0.07	82.4 ± 0.25	2.9 ± 0.07	78.3 ± 0.23

Table 3.2: The characteristic times τ_s and τ_l (in units of ps) for water and two different models for hydrogen peroxides around a single PRDX5 protein from our simulations of System B.

Note: The errors reported here are those from fitting to an exponential form.

Characteristic τ_s for water are consistent with those reported in references [104, 13] (around 3-5 ps) and most likely represent boundary recrossing situations. The constant τ_l mainly characterizes those molecules associated with the protein surface. Experimental and computational articles [85, 88, 13, 54, 74, 72, 83] report that τ_l may range from 10 to 100 ps. These values are in agreement with our findings in Tab. 3.2. It should be noted that, if $(\tilde{f}(\geq \tau))$ was truly a sum of exponential factors, then the time constants τ_s and τ_l should be identical to those of the survival time correlation function $f(\tau)$. However we find that $f(\tau)$ is clearly non-exponential. Since $f(\tau)$ can not be realiable fit as a sum of two exponentials we do not report here values for τ_s and τ_l for this function.

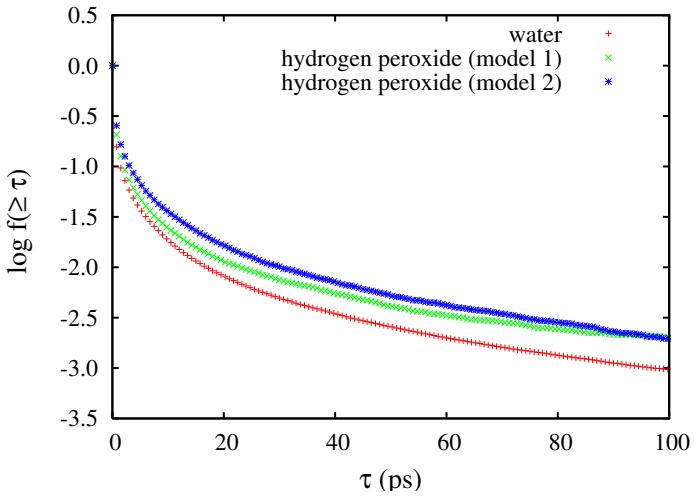


Figure 3.5: Cumulative probability distributions of first passage time $(\tilde{f}(\geq \tau))$ for water and hydrogen peroxide in the first hydration shell (0 - 4 $\mathring{A}$) of system B.

Figure 3.6(a) shows the mean square displacements $(< \Delta r^2 >)$ of water and

the two different models for hydrogen peroxide in hydration shells of different size from our simulations of System B. It is clear that the diffusivity of hydrogen peroxide is highly dependent on the model. For the model with larger partial charges the diffusivity is smaller than that of water while for that with the smaller partial charges diffusivity is larger. It is interesting, however, that this does not appear to affect the residence times on the surface of the protein since both models give similar results. In all cases water and peroxide diffuse slower when close to the protein. This is consistent with previous observations [84, 75].

Another very interesting finding from our simulations is that regardless of the model, hydrogen peroxide has a concentration enhancement on the surface of the protein. This can be seen from Figure 3.7. Since the volume element must be taken into account in order to compute a density, and since the volume element is difficult to define for a system with rough surfaces; in Figure 3.7 we plot the ratio of the number (or of the densities) of peroxide molecules to water molecules $\frac{n_{wat}}{n_{per}} = \frac{n_{wat}/V}{n_{per}/V}$. This quantity is a measure of the relative concentration of the two species. A water molecule or a peroxide molecule is considered to be in a given volume shell, V_d, if its center of mass is within a distance "d" from the protein. It is clear that at long distances away from the protein, the ratio of concentrations is that expected in the bulk while at short distances (around the first hydration shell) there is a clear enhancement of the concentration of peroxide at the expense of a depletion in the case of water. We also notice a slight depletion in the peroxide concentration at around 4.5Å.

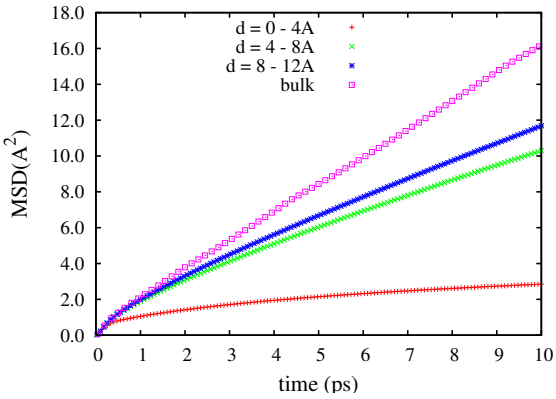
We have also studied the distribution of water and hydrogen peroxide in the intervening space between two PRDX5 proteins (System C). Figures 3.8(a) and 3.8(b) show that independent of the model used for hydrogen peroxide there is an enhancement of water concentration at the center of the intervening space while there is a depletion of peroxide in the center and an enhancement in contact with the protein. The results for one protein and two proteins are consistent and imply that hydrogen peroxide not only resides longer on the surface of the protein but also is more likely to non-specifically bind to proteins when compared to water. It is worth mentioning that in this simulation the active site of PRDX5 is not contained in the volume element depicted in Figure 3.8(c). This indicates that the results are not due to peroxide being trapped at the active site.

3.4 Conclusion

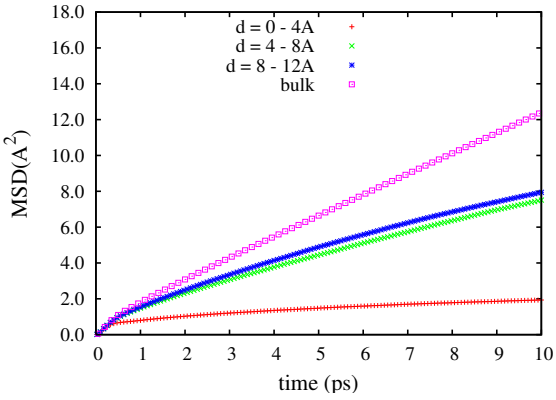
We have compared the structure and dynamics of hydrogen peroxide and water on the surface a single protein and in the intervening space between two proteins. Regardless of the model used for hydrogen peroxide, it appears that H_2O_2 has higher affinity than water for the protein. The distribution of passage times shows that peroxide molecules stay longer on the surface of the protein and there is a density enhancement of peroxide on the surface at the expense of a depletion in water.

Theoretical analysis and simulation results show that the cumulative probability distribution of first passage times has less statistical error than other methodologies previously used to study residence times of molecules on different surfaces.

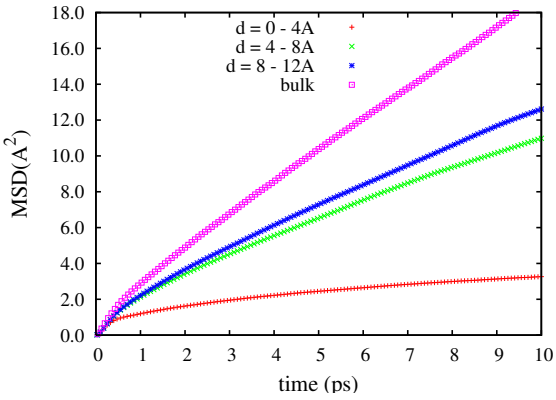
This is due to the better treatment of molecules with long times of residence for which simulations can not provide good statistics. Our approach presents a better way to estimate the times of residence of a molecule on a rough molecular surface and we present a clear analytical link between this formulation and previously derived survival time correlation functions.



(a) water

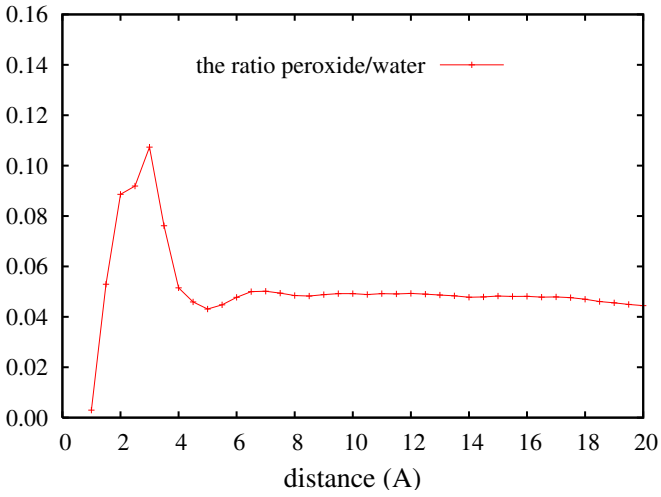


(b) hydrogen peroxide, model 1

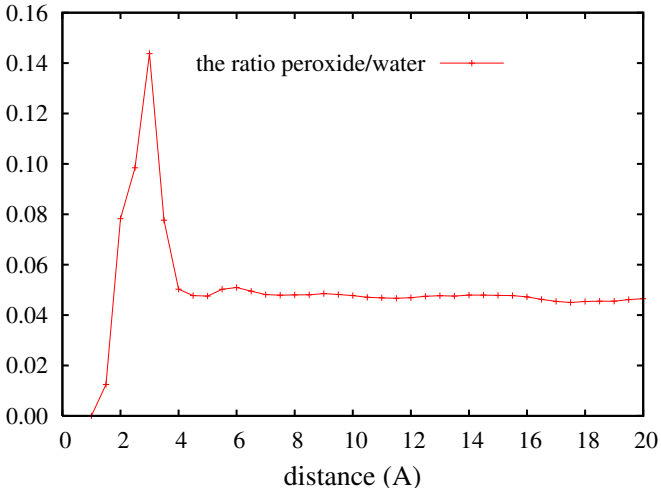


(c) hydrogen peroxide, model 2

Figure 3.6: Mean square displacement for water and hydrogen peroxide in hydration shells of different size in system B.

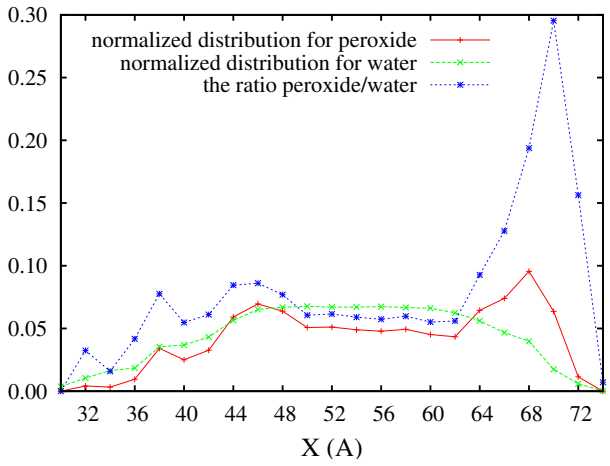


(a) model 1

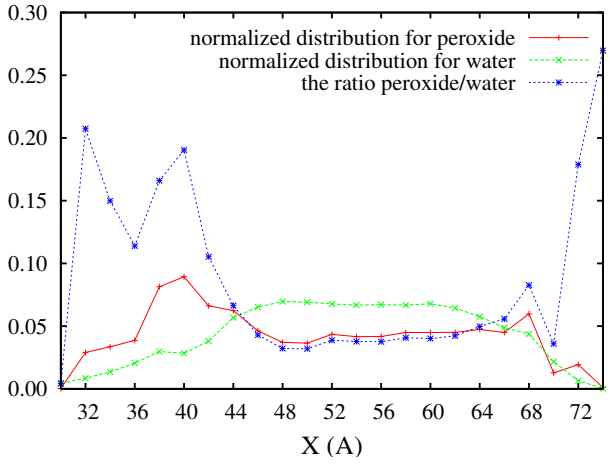


(b) model 2

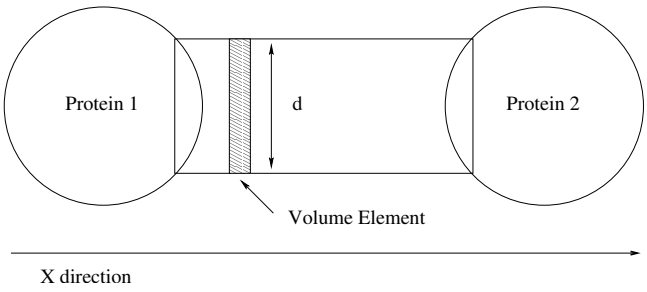
Figure 3.7: Ratio of hydrogen peroxide to water concentration as a function of the shortest distance (r) to the protein in system B.



(a) model 1



(b) model 2



(c) Two Proteins Diagram

Figure 3.8: Normalized distribution of water molecules, peroxide molecules and corresponding number density ratios in the intervening space between two PRDX5 proteins. In order to construct this system, two overlapping proteins were displaced along the X direction in such a way that the surfaces are approximately 17 $\mathring{A}$ apart. The X axis simply corresponds to the X coordinate in the system. Figure 3.8(c) shows a schematic diagram of the system. Water and peroxide molecules are only counted if they fall within a cylindrical volume of diameter $d = 29\mathring{A}$.

CHAPTER 4
THE BEHAVIOR OF WATER ON THE SURFACE OF OLIGOSACCHARIDES

In this chapter we characterize the behavior of water on the surface of a diverse group of carbohydrates and attempt to determine the role of saccharide size, linkage and branching as well as secondary structure on the dynamics and structure of water at the surface. In order to better understand the similarities and differences in the behavior of the solvent on the carbohydrate surface we explore residence times, rotational correlation function, local solvent occupancy numbers and diffusivities. We find that due to the differences in secondary structure water residence times are longer and translational and rotational dynamics are retarded when in contact with wide helices and branched sugars. In the case of extended helices and smaller oligosaccharides, water dynamics is faster and less hindered. This indicates that branching, the type of linkage between monomers and the anomeric configuration all play a major role in determining the structure and dynamics of water on the surface of carbohydrates.

4.1 Introduction

Carbohydrates are an important class of biomolecules that can be found free in the cytoplasm, decorating the surface of proteins and as parts of glycolipids. They are involved in cell adhesion, immune responses, protein trafficking and signal processing. Understanding the structure and function of oligosaccharides is very important because these molecules are exquisite biological recognition agents. This unique-

ness stems from the large number of chiral centers, the presence of branching and their conformational variability. As an example, in diseased states, glycans are expressed differently and act as biomarkers in cancer, AIDS and rheumatoid arthritis [120, 10, 41, 40].

Dashnau *et al.* showed that the orientation of hydroxyl groups in axial and/or equatorial positions in aldohexopyranoses affects the water structuring in the first hydration shell. Aldohexopyranoses such as β-glucose, β-mannose and β-galactose have hydrophobic and hydrophilic hydration sites that play a role in aromatic interaction during carbohydrate-protein recognition [35, 112]. Resonance two photon ionization and ultraviolet and infrared ion-dip spectroscopy of hydrated mono- and disaccharides have shown that water on the surface of carbohydrates helps these biomolecules achieve the conformations that are recognized by proteins i.e., water is not a mute spectator but it actively participates in molecular recognition events (see reference [109] and citations therein). In contrast with the large amount of information available for the role of water in contact with proteins and nucleic acids, no comprehensive study of the role of water on the surface of glycans is available. This is perhaps because of the topological complexity of sugars. It is therefore of crucial interest [115, 116] to shed light on the water structure patterns and the diffusive dynamics on the surface of carbohydrates as a function of the key elements present in carbohydrates but absent in proteins such as branching, linkage pattern and anomeric configuration [56, 39, 2].

Several experimental and theoretical studies have reported on the behavior of water at the interface with carbohydrates. Because of the complexity of these systems,

most of these studies have been carried out on monosaccharides, disaccharides or small model oligosaccharides. Kirchner and Woods have performed high level quantum and molecular dynamics simulations and have shown that the conformational preferences for the 1→6 linkage are correctly reproduced only in the presence of water [62]. In exploring the role of water in the vicinity of simple monosaccharides such as α-d-glucopyranose and α-d-xylopyranose, Leroux *et al.* observed that the hydroxyl groups of the monosaccharide units align in such a way that they form hydrogen bonds with water instead of intra-molecular hydrogen bonds [68]. Consistent with these computational predictions, recent depolarized Rayleigh scattering (DRS) and low-frequency Raman Spectroscopy experiments performed on an aqueous glucose solution by Paolantioni *et al.* showed that a solute with the ability to have multiple hydrogen bonds disrupts the tetrahedral geometry of water in its first hydration shell. The loss of hydrogen bonding between water molecules is compensated by sugar-water hydrogen bonds leading to the denser water environment around the sugar [89].

Lee *et al.* have shown that disaccharides such as sucrose and trehalose not only disrupt the tetrahedral geometry of water in their vicinity but also its translational and rotational dynamics. Dynamics on the surface of these dissaccharides is much slower than that on the surface of glucose [67]. Liu *et al.* have shown that α,α-trehalose imposes a strong anisotropic structuring of solvent that extends up to three solvation shells away from the sugar due to the formation of water mediated H-bonds. In computational studies the self-diffusion coefficient of water in 87 μM α,α-trehalose solution was found to be 20% smaller than that in neat water simulations

[70]. Englesen *et al.* studied disaccharides maltose, sucrose and trehalose in dilute aqueous solutions. The H-bonding pattern, the solvent residence times and the solvent density around these disaccharides were observed in simulation to be different. Water surrounding trehalose displayed the longest residence times and was clearly more structured than in the vicinity of maltose and sucrose [42, 43].

Almond *et al.* have elucidated the role of water on the surface of small oligosaccharides composed of α- and β-linkages of glucose and mannose. In the case of α-linkages, weak intramolecular hydrogen bond interactions along contiguous residues and many water mediated hydrogen bonds were observed contrary to the situation when β-linkages are present. In the case of β-linkages water mediated hydrogen bonds appeared not to be favored; instead strong direct hydrogen bonds with water were observed [6, 7].

Although the above studies are very important, as far as we know there have been no systemic and comprehensive analysis of the water structure and dynamics on the surface of moderate to large oligosaccharides consisting of various linkages and monosaccharide compositions. Several outstanding questions arise; how do water structure and dynamics differ around a linear homopolymer of glucose or mannose and across $1\rightarrow2$, $1\rightarrow3$ and $1\rightarrow4$ linkages? How does the anomeric configuration of the different monosaccharides affect surrounding water dynamics? How do branched oligosaccharides differ from linear oligosaccharides with the same monosaccharide composition in their ability to affect water structure and dynamics?

In order to address the effect of branching on a given type of oligosaccharides,

we have studied several variations of the Man_9 sugar. In order to address issues related to the anomeric configuration, linkage point and monosaccharide identity, we have studied homopolymers of 9 monomeric units of 1→2, 1→3, and 1→4 linkages of α and β glucose and mannose sugars (Fig. 1 and Table 1). The effect of polymer length is addressed by comparing the results for nonamers with two selected twenty-mers.

We have recently written several articles [130, 128, 129, 76, 122] addressing the problem of sugar conformational variability in solution. This chapter does not attempt to address this issue but instead is written from the point of view of the solvent around well defined sugar structures. In subsequent sections we will show that water structure and dynamics on the surface of sugars depends on the overall structure of the biomolecule, its anomeric configuration and types of linkages. We will also show that some simple rules of thumb can be derived by systematically analyzing the results derived from our simulations.

4.2 Materials and Methods

The models chosen for this study are categorized into 4 groups. Homopolysaccharides of glucose, mannose, branched sugars and a trisaccharide. Within the homopolysaccharides we have studied 6 glucose and 6 mannose nonamers and 2 mannose twenty-mers. For easy reference, the branched structures and the glucose homo nonamers as well as the trimer are depicted in Figure 4.1. Common names for some of these molecules can be found in Table 4.1.

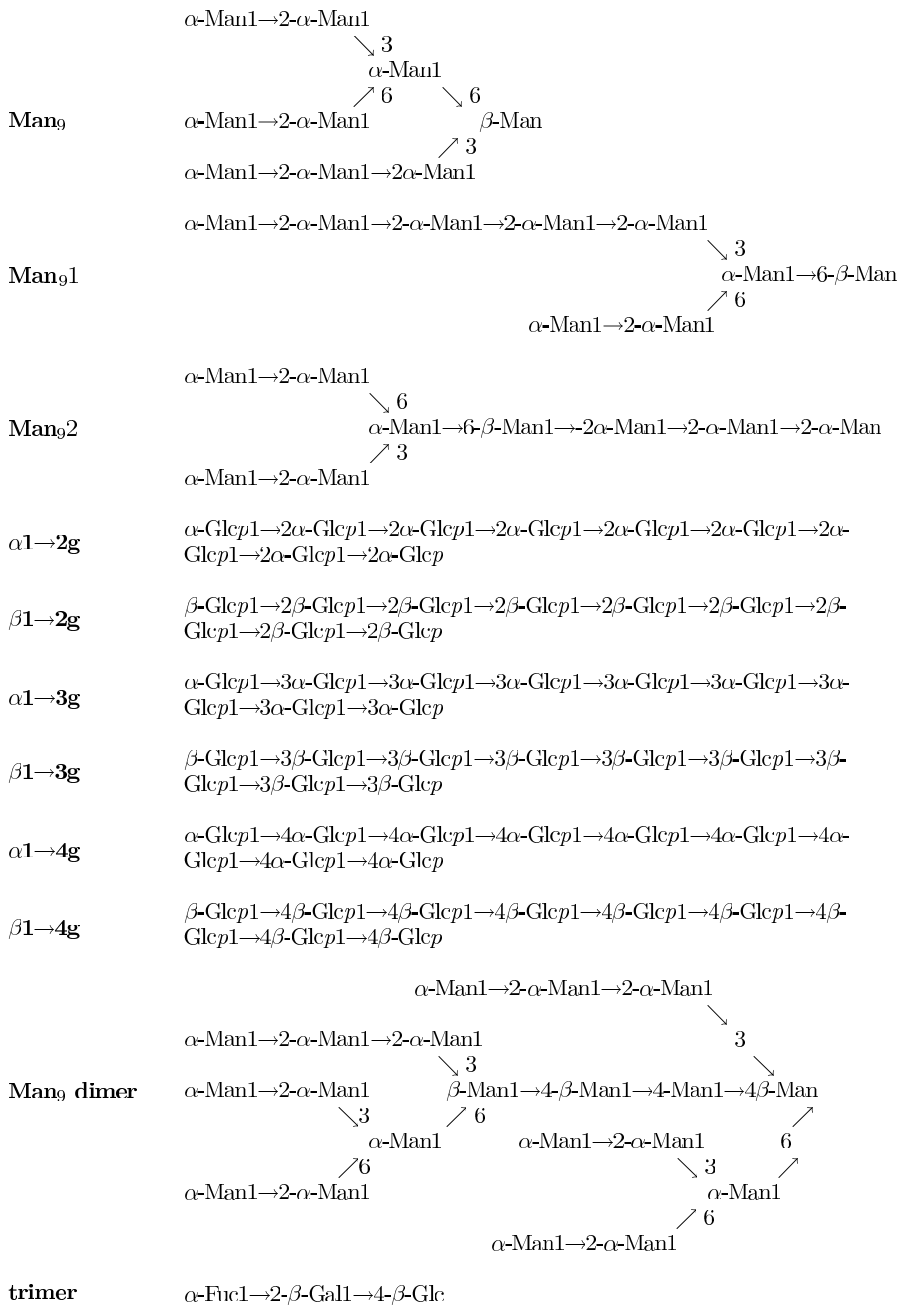


Figure 4.1: Some of the various oligosaccharide models used in the study. The Man_9 oligosaccharide was manipulated to generate the different variants $Man_9$1, $Man_9$2 and Man_9 dimer

Common Name	residue	linkage
Amylose	Glucose	$\alpha 1 \rightarrow 4$
Cellulose	Glucose	$\beta 1 \rightarrow 4$
Crown Gall	Glucose	$\beta 1 \rightarrow 2$
Laminaran	Glucose	$\beta 1 \rightarrow 3$
Mannan	Mannose	$\beta 1 \rightarrow 4$

Table 4.1: Common names of some of the sugars chosen for this study [96]

All oligosaccharide models were built using the xleap module in AMBER 9.0 The choice of the initial conformations was based on previous experimental information [86, 105, 106, 132, 8, 91]. We used the GLYCAM [62, 127] force field for all our calculations. The total number of atoms in the system ranged from 5000-14000 depending on the model studied. All the oligosaccharides were solvated using SPC [58] water in truncated octahedron boxes. Long range electrostatic interactions were treated using the Particle Mesh Ewald summation [34]. Molecules were energy minimized using the steepest descent method followed by the conjugate gradient method. After constant pressure and temperature equilibrations for at least 250ps at ambient conditions, production molecular dynamics simulations in the NVE (constant energy, volume and number of atoms) ensemble of 5 ns in duration were carried out with the Sander module of AMBER 9.0. Because of the time scale separation between solvent dynamics and sugar conformational dynamics 5 ns was enough to converge all correlation functions. Data was collected every 50 fs for subsequent analysis.

In Chapter 2 we introduced our theoretical methodology and a computational algorithm that significantly improve on the accuracy and reliability of residence time

calculations on complex surfaces [30]. Besides, these methods, we also use solvent occupancy plots to describe the water structure and helical parameters, n and h, to describe helical structures of oligo-saccharides. One thing we need to mention here is that in the calculation of residence times, MSDs and rotational autocorrelation functions, the hydration shell is defined as the region within 4Å from the solute.

3-D solvent occupancy plots were generated on typical 50ps fragments of our trajectories using the Chimera software from UCSF [92]. Grid cells of $1Å^3$ appear colored only when they are occupied by solvent molecules 10% of the time. This particular value was selected because it allows for optimal discrimination between our sugar models. At higher values, almost no occupancy can be detected for $\beta 1 \rightarrow 4$ homopolysaccharides.

In order to classify our linear saccharides, helical parameter, n (number of residues per pitch) and h (advancement per monomer unit) were calculated according to the definition in the following references [96, 49, 100]. According to Rees'[100] classification of perfectly periodic helices our studied sugars fall into the ribbon family n=2-$\pm$4 and h=4-6Å (this is what Rao's book[96] calls extended helices)and the hollow helix family n=2-$\pm$10 and h=0-4Å (this is what Rao's book[96] calls wide helices).

4.3 Results and Discussion

Several X-ray and conformational studies have been reported on homopolymers of $1 \rightarrow 2$, $1 \rightarrow 3$ and $1 \rightarrow 4$ glucose and $1 \rightarrow 4$ mannose. [86, 106, 105, 132, 8, 91] In

order to validate our simulation methodology we have computed torsional angles as a function of time for these sugars and found that both the minimum energy configurations as well as fluctuations are consistent with previously reported values [86, 106, 105, 132, 8, 91]. We are therefore confident that sugar conformations are those experimentally reported and they do not significantly change on the time scale of our 5 ns simulations.

In order to classify the morphology of our saccharides, we computed helical parameters n-h along simulation using the method described by French. [50, 49] Results are presented in Fig. 4.2. Though results are similar, the dispersion in our data is larger than that in reference [50] because our simulations do not force the saccharides to remain in a perfect helical configuration. According to reference [50] and our findings (see representative structures in Fig. 4.3), as the homopolysaccharide chain length is increased, these sugars form either wide helices (helices with large diameter that have a large number of residues per turn) or extended helices (helices with small diameter and low number of residues per turn). Based on the definitions in Section 4.2 and Fig. 4.2 we find that linear homopolysaccharides with linkages $\alpha 1 \rightarrow 2$, $\beta 1 \rightarrow 2$ $\alpha 1 \rightarrow 4$ and $\beta 1 \rightarrow 3$ tend to form wide helices and $\alpha 1 \rightarrow 3$ and $\beta 1 \rightarrow 4$ linkages tend to form extended helices

4.3.1 The Effect of Saccharide Structure on the Surrounding Solvent Structure and Residence Time

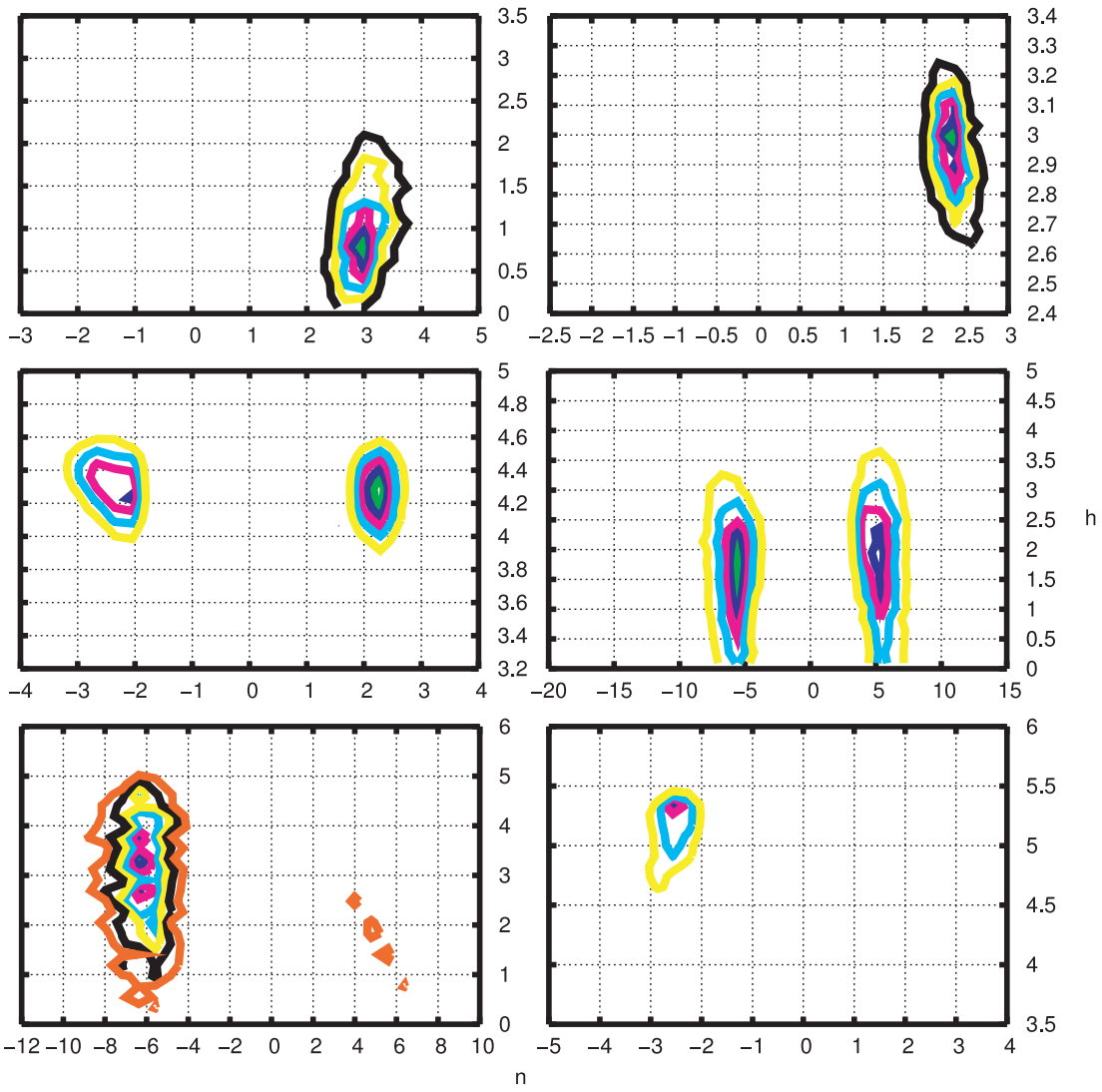


Figure 4.2: Helical parameters n (number of residues per pitch) and h (advancement per monomer unit in Å) from our production runs of α and β glucan homopolysaccharides. From top to bottom, L-R are shown models of $\alpha 1{\rightarrow}2$, $\beta 1{\rightarrow}2$, $\alpha 1{\rightarrow}3$, $\beta 1{\rightarrow}3$, $\alpha 1{\rightarrow}4$, $\beta 1{\rightarrow}4$-glucans. Negative values of n represent a left-handed helix whereas positive values represent right-handed helix. These figures can be compared with French's work [50] Figures 2-6.

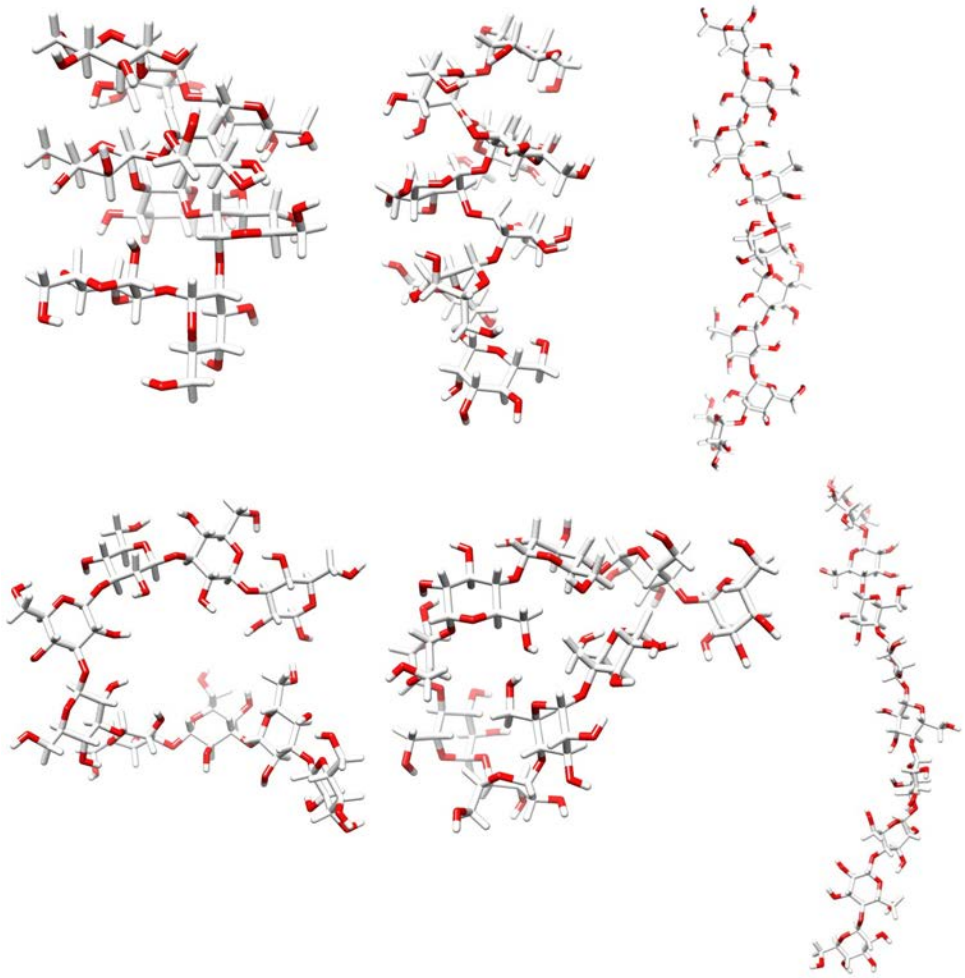


Figure 4.3: Representative structures from our production runs of α and β glucan homopolysaccharides. From top to bottom, L-R are shown models of $\alpha 1 \rightarrow 2$, $\beta 1 \rightarrow 2$, $\alpha 1 \rightarrow 3$, $\beta 1 \rightarrow 3$, $\alpha 1 \rightarrow 4$, $\beta 1 \rightarrow 4$-glucans. Mannose linear homopolymers are not shown in this figure but they have very similar secondary structures to the glucose analogues. In general, we find that linear homopolysaccharides with $\alpha 1 \rightarrow 2$, $\beta 1 \rightarrow 2$, $\beta 1 \rightarrow 3$ and $\alpha 1 \rightarrow 4$ linkages form wide helices while those with $\alpha 1 \rightarrow 3$ and $\beta 1 \rightarrow 4$ linkages tend to form extended helices in solution.

The fraction of molecules with first passage time greater than or equal to t ($\tilde{f}(\geq t)$) is a measure of the behavior of water in contact with a surface and can be used to compare the nature of different sugar-solvent interfaces. In this study, we start by attempting to address the effect of linkage, anomeric configuration, secondary structure, branching, size and monosaccharide identity on $\tilde{f}(\geq t)$. Fig. 4.4 shows the logarithm of $\tilde{f}(\geq t)$ for a broad set of sugars with widely different structural characteristics ranging from a trisaccharide to a dimer of Man_9 and two structurally very different linear mannose twenty-mers. Some of these saccharides, are linear, others are highly branched and they have different linkage and anomeric configurations. For clarity this figure also shows typical conformations in solution for each of these molecules and surrounding solvent occupancy isosurfaces.

Since Fig. 4.4 is on a logarithmic scale, the slope of these curves correspond to the inverses of the characteristic water residence times on the surface of the different saccharides. Table 4.2 shows the value of these inverse slopes between 20 and 30ps. In broad terms, it is clear from Fig. 4.4 and Table 4.2 that branched or in general crowded sugars as well as sugars forming wide helices have longer solvent residence times. The residence time on the surface of extended helices is in general much shorter. Secondary structure plays a crucial role in defining the residence time of water on the sugar surface. Size appears to be less of a factor since for example $Man_{20}\beta 1\rightarrow 4$ and a $\beta 1\rightarrow 4$ nonamer have almost identical residence times.

Model	Secondary Structure	τ (ps)	MSD ($\mathring{A}^2/ps$)	$\tau_{rot,fit\ 0-10}$ (ps)	$\tau_{rot,fit\ 4-10}$ (ps)
β1→2m	Wide Helix	34.3 ± 0.13	0.872	10.43 ± 0.12	13.30 ± 0.08
man_9 dimer	Branched	33.3 ± 0.11	0.829	11.23 ± 0.16	15.28 ± 0.14
α1→3m	Extended Helix	32.9 ± 0.11†	0.928	10.57 ± 0.13	13.63 ± 0.09
α1→2g	Wide Helix	32.6 ± 0.09	0.889	10.21 ± 0.13	13.14 ± 0.07
β1→3g	Wide Helix	31.7 ± 0.11	0.977	9.31 ± 0.09	11.31 ± 0.05
β1→2g	Wide Helix	31.0 ± 0.07	0.855	10.95 ± 0.15	14.86 ± 0.13
Man_9	Branched	30.9 ± 0.07	0.859	10.58 ± 0.13	13.64 ± 0.09
α1→2m	Wide Helix	30.6 ± 0.06	0.903	10.58 ± 0.13	13.60 ± 0.07
Man20α1→4	Wide Helix	30.4 ± 0.05	0.917	10.28 ± 0.12	13.15 ± 0.08
β1→3m	Wide Helix	29.6 ± 0.10	0.981	9.84 ± 0.11	12.26 ± 0.07
α1→4g	Wide Helix	27.7 ± 0.06	0.905	9.59 ± 0.10	12.00 ± 0.08
α1→3g	Extended Helix	27.4 ± 0.06	0.946	9.58 ± 0.11	12.09 ± 0.07
Man20β1→4	Extended Helix	24.5 ± 0.03	1.105	8.63 ± 0.07	9.95 ± 0.02
α1→4m	Wide Helix	24.5 ± 0.03	1.051	8.94 ± 0.10	11.23 ± 0.07
β1→4g	Extended Helix	23.7 ± 0.06	1.034	9.01 ± 0.09	11.20 ± 0.07
Trisaccharide	No Secondary Structure	23.4 ± 0.06	0.988	9.23 ± 0.11	12.14 ± 0.12
β1→4m	Extended Helix	23.0 ± 0.02	1.082	8.80 ± 0.08	10.80 ± 0.06

Table 4.2: Characteristic residence times (τ values), mean square displacements, and time constants for rotation of our model saccharides derived from simulation. The categorization of helices is based on the definitions in Section 4.2. The mean square displacements were obtained via curve fitting in interval 4 to 10 ps. The time constants for rotation were obtained via curve fitting in intervals 0 to 10 and 4 to 10ps. Errors reported are those of the least squares fit. α1→3m (labeled with a †) shows interesting off pattern behavior due to a very particular epimeric effect that allows for highly localized water mediated hydrogen bonds.

Heyden *et al.* have studied solvation dynamics of water surrounding saccharides of different size using terahertz spectroscopy. In a comparison of a monosaccharide with two disaccharides the authors observed that water hydrogen bond dynamics in the first solvation shell of the sugar was significantly less affected in the case of the monosaccharide [53]. Larger oligosaccharides with particular secondary structure patterns favor water mediated hydrogen bonds. In calculating the residence times of water in aqueous solution of sucrose, Englesen and Pérez [43]found that the presence of water mediated hydrogen bonds results in longer water residence times on the sugar surface. In particular, they observed that two water mediated hydrogen bonds, namely O-2g...Ow...O-3f and O-2g...Ow...O-lf were present in their simulations for more than 40% of time. More recently, Veluraja and Margulis found the same type of behavior in Sialyl LewisX containing oligosaccharides [122].

Among all the homopolysaccharide models selected for this study, the residence times of $\beta 1 \rightarrow 2$-m and $\beta 1 \rightarrow 4$-g or mannose constitute two extremes. Based on our distributions of first passage times, we are able to determine that residence times in the case of $\beta 1 \rightarrow 2$-mannose (wide helix) is largest while those for $\beta 1 \rightarrow 4$-glucose and mannose (extended helices) are shortest. The residence times for the solvent around all other nonamer homopolysaccharides fall in between these two limits. The distinction between wide and extended helices also explains the results in the case of our mannose twenty-mers in Fig. 4.4. The twenty-mer forming a wide helix has longer water residence time.

The effect of branching can be gauged by analyzing a group of different

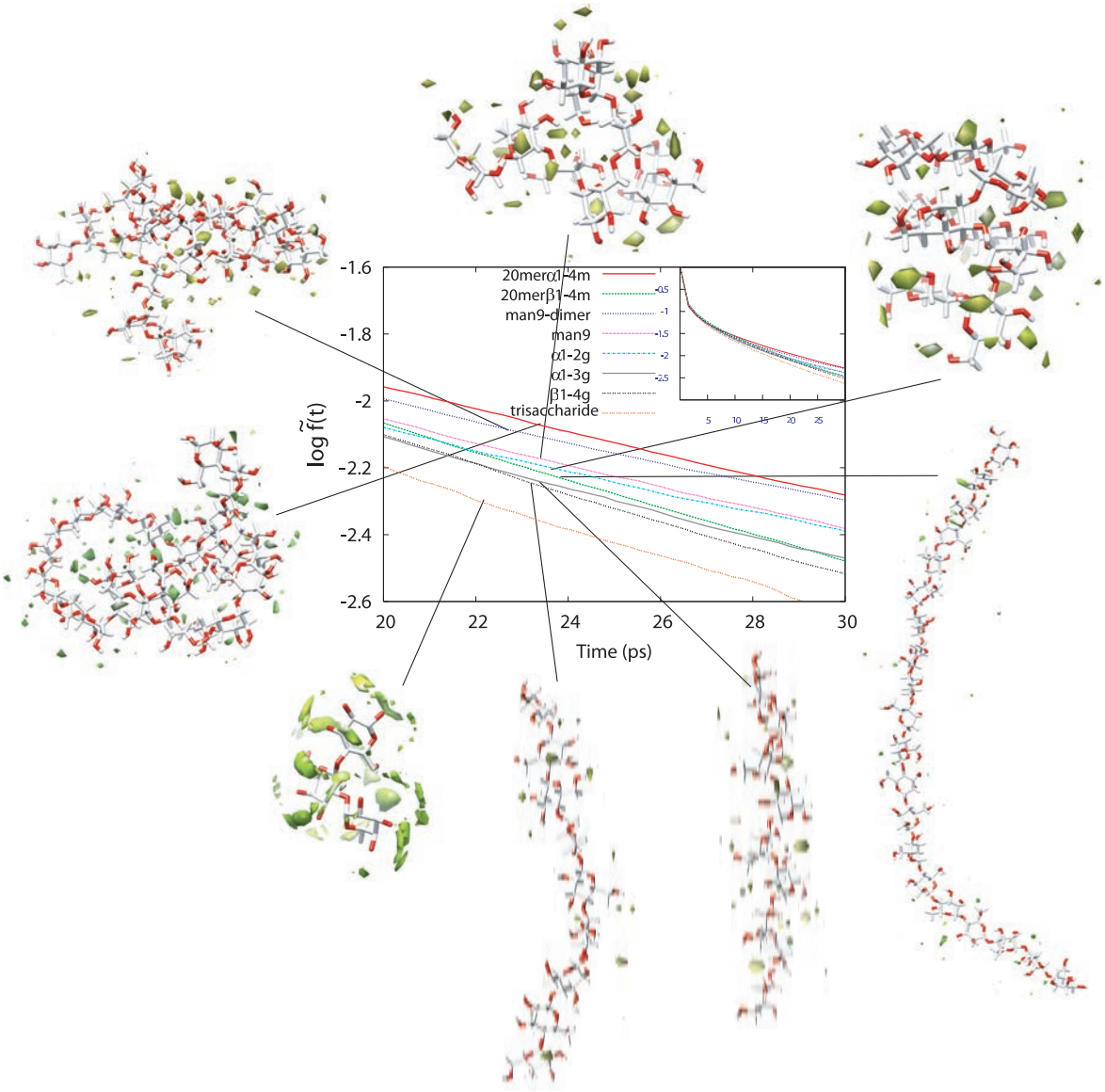


Figure 4.4: Logarithm of the fraction of molecules with first passage time greater or equal to t as a function of t. The slopes of these plots correspond to the residence times.

oligosaccharides belonging to the Man_9 family. The function of $\tilde{f}(\geq t)$ for all of these is fairly similar and the τ values are similar to that for Man_9 in Table 4.2.

Most of our findings regarding residence times on the surface of homopolysaccharides of glucose and mannose can be explained in terms of their secondary structure and the extent to which they are able to trap water on their surface[70, 42, 6, 33]. As previously discussed, the $\alpha 1 \rightarrow 2$-, $\beta 1 \rightarrow 2$-glucose/mannose, $\beta 1 \rightarrow 3$-[19] glucose/mannose and $\alpha 1 \rightarrow 4$-glucose/mannose tend to form wide helices while $\alpha 1 \rightarrow 3$-glucose and mannose and $\beta 1 \rightarrow 4$-glucose [98] and mannose tend to form extended helices. Our analysis of local water occupancies appear to indicate that in wide helical conformations, the orientation of the pyranose rings is such that small pockets of water can get trapped giving rise to loci of high local density. On the contrary, in the case of extended helical conformation, solvent residence times are shorter, and for the most part solvent occupancies do not show appreciable water trapping though some interesting exceptions exist.

In Fig. 4.5 we show five characteristic examples of water occupancy isosurfaces in contact with saccharides of different sequence, linkages and secondary structure. The top two saccharides $\alpha 1 \rightarrow 2$-glucose, and a modified Man_9 correspond to a wide helix and a branched oligosaccharide respectively. In both cases significant water trapping and slow down occurs on the surface. From top to bottom, the following two structures, $\alpha 1 \rightarrow 3$-glucose and $\beta 1 \rightarrow 4$-glucose are extended helices; for these the residence time of water on the surface is shorter and water dynamics is faster. There are clear differences between the two extended sugars. In the case of $\alpha 1 \rightarrow 3$-glucose

(see Fig. 4.5) some water trapping can clearly be observed due to the particular structural pattern that facilitates the formation of water mediated hydrogen bonds between O6-O2-O4 in three consecutive residues. In the case of β1→4-glucose almost no water trapping can be observed at the same iso-solvent occupancy contour level. A very interesting exception to the wide/extended rule appears to be α1→3-mannose (bottom of Fig. 4.5). Table 4.2 clearly shows that α1→3m has one of the largest τ values. This is due to a combination of secondary structure similar to that in α1→3-glucose and an epimeric effect that favors water trapping (see Figure 4.6).

4.3.2 Solvent Rotational and Translational Motion at the Saccharide Surface

While the distribution of first passage times and corresponding residence times provide information about how long a typical solvent molecule will remain in contact with the saccharide surface, it does not provide detailed information regarding its mobility. The calculation of mean square displacements and rotational correlation functions for water molecules on the saccharide surface convey information regarding the degree to which solvent motion is hindered and enables us to obtain deeper understanding of the similarities and differences between saccharide surfaces.

Figure 4.7 shows rotational correlation functions for water molecules in the first solvation shell around selected saccharides of different characteristics. Consistent with our previous analysis, the fastest relaxation occurs in the case of β1→4-mannose twenty-mer followed by β1→4-glucose nonamer. Fig. 4.7 clearly shows that these two

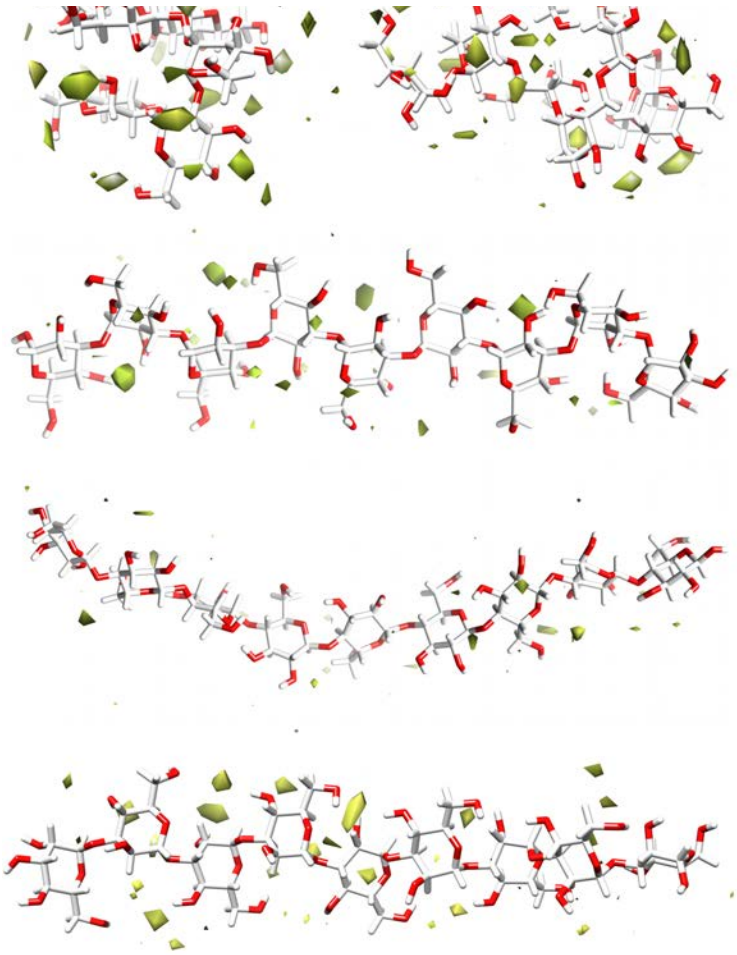


Figure 4.5: Water density around saccharides with varied sequence, branching, linkage and secondary structure. From left to right top to bottom the saccharides are: α1→2-glucose, modified Man_9, α1→3-glucose, β1→4-glucose and α1→3-mannose

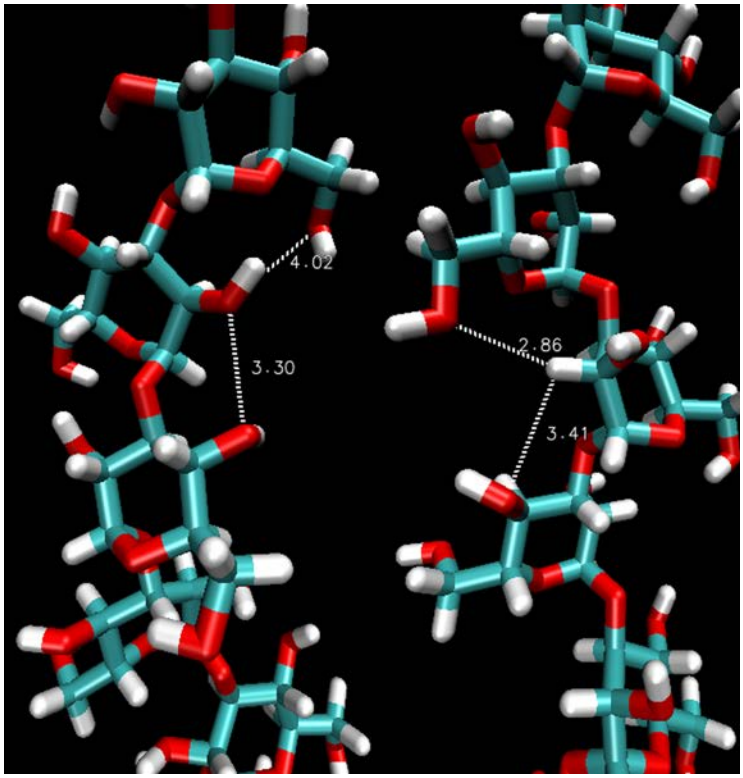


Figure 4.6: Comparison of typical snapshots along simulation for $\alpha 1 \rightarrow 3$-glucose (left) and $\alpha 1 \rightarrow 3$-mannose (right) showing the loci of enhanced water trapping. The dashed lines indicate the distance between O6.....O2 and O2.....O4, showing the epimeric effect. In case of mannose the O2 is axial which makes the water molecule trapped in comparison to glucose. The epimeric difference between Man and Glu results in tighter solvent configurations in contact with the surface in the case of Mannose.

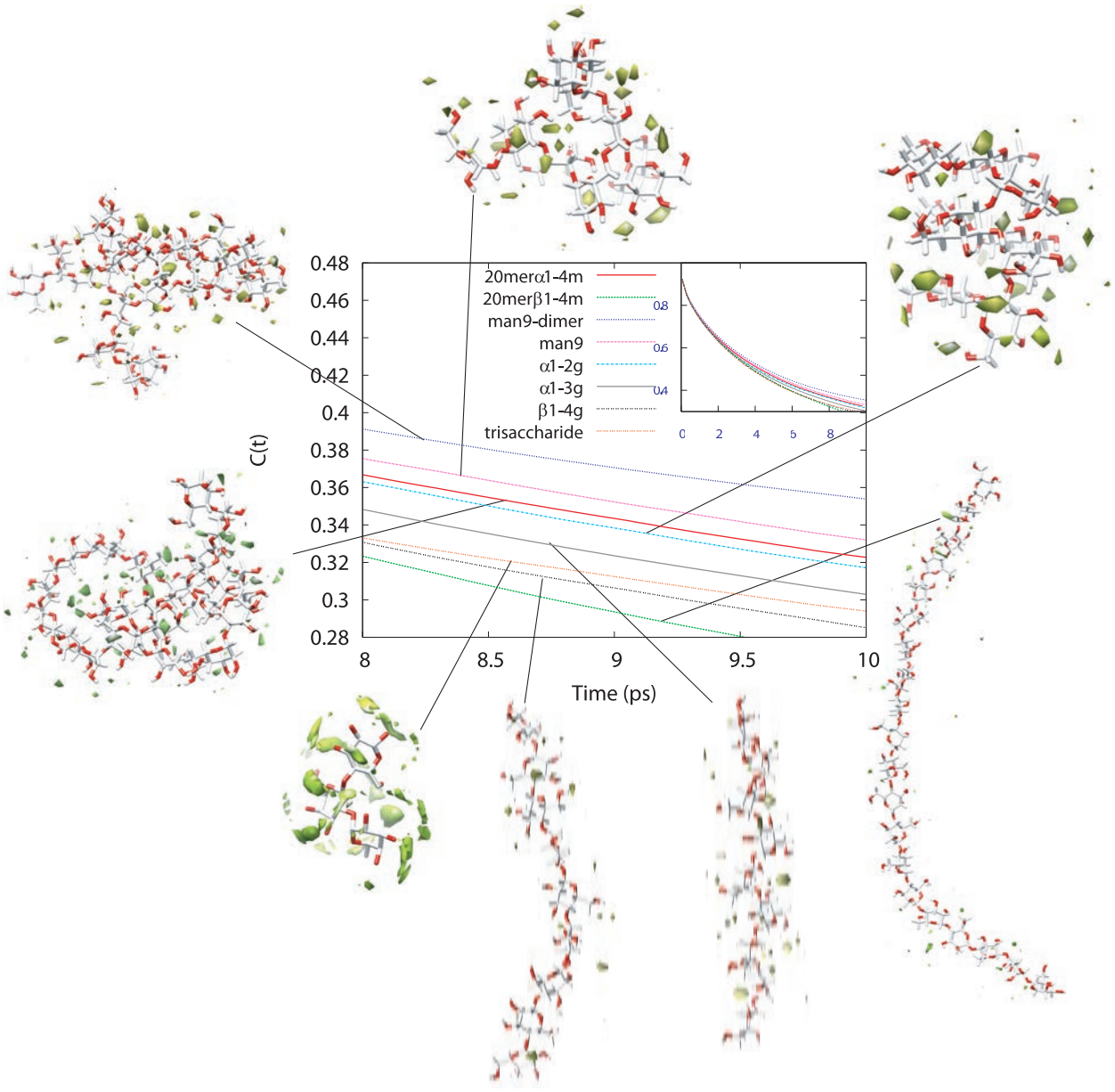


Figure 4.7: Water rotational correlation functions in the first solvation shell on the saccharide surface. Inset shows the complete decay of this function in the range from 0 to 10ps.

are extended helices with modest ability for water trapping. Similarly fast decay of correlation is observed in the case of the trisaccharide. Intermediate behavior is observed in the case of α1→3-glucose an extended helix that shows moderate ability to trap water as demonstrated by solvent isosurfaces as well as residence time. The decay of correlation is slowest in the case of wide helices and complex branched oligosaccharides such as Man_9dimer and the α1→4-mannose twenty-mer.

A more detailed analysis of the rotational behavior of water surrounding homopolysaccharides can be obtained from Figure 4.8. Figure 4.8(A) and (B) compare the rotational behavior of glucose and mannose nonamers. The first notable difference between epimers can be seen in the case of α1→3 and α1→4 linkages. In the case of glucose, the water rotational autocorrelation functions at the surface of α1→3-glucan decays at an intermediate rate, faster than a typical wide helix such as β1→2-glucose and slower than an extended helix such as β1→4-glucose. In the case of mannose the water rotational correlation functions on the surface of α1→3-mannose shows uncommonly slow decay while that of α1→4-mannose displays a fast decay. The case of α1→3-m and α1→3-g are specially interesting because of the particular arrangement of atoms O6-O2-O4 in three consecutive rings that make water residence time on the surface unusually large compared to the case on the surface of other extended helices. The epimeric configuration at O2 in mannose makes this effect even more pronounced. The epimeric configuration of these two sugars can be clearly seen in Fig. 4.6 while the different local 3D solvent occupancies are shown in Fig. 4.5. The configuration of O2 in the case of mannose facilitates the formation of a significantly

rigid water structure that promotes water mediated hydrogen bonds between O6 and O4.

The case of α1$\rightarrow$4-linkage is peculiar since it is known that amylose (α1$\rightarrow$4-glucan) attains a wide range of helical conformations. [91] This may explain why water rotational correlation functions have dissimilar decays for glucose and mannose homopolymers. It is clear that 5ns is enough to study water structure and dynamics on systems with well defined helicity, however when the separation of time scales between water dynamics and saccharide conformational dynamics is smaller our results may reflect a combination of the water dynamics on two different saccharide secondary patterns.

Figures 4.8 (C) (D) and (E) compare the rotational correlation functions of water on the surface of α and β homopolymers at different linkage positions. It is clear as we have discussed in previous paragraphs that α1$\rightarrow$3-m is special in the slow decay of water rotational correlation as can be appreciated from Figures 4.8 (C). The rest of the 1$\rightarrow$3 homopolymers have water rotational correlation functions on their surface that decays at faster rates. In general, we find that regardless of whether they are α or β the 1$\rightarrow$2 (Figures 4.8 (D))homopolymers have water rotational correlation functions that decay slowly. This is because they all tend to form wide helices with many locations suitable for tight water mediated hydrogen bonds. Figures 4.8 (E) can also be easily understood. Except for α1$\rightarrow$4-g which is in the wide helix configuration during simulation all other homopolymers form extended structures that are less suitable for water trapping.

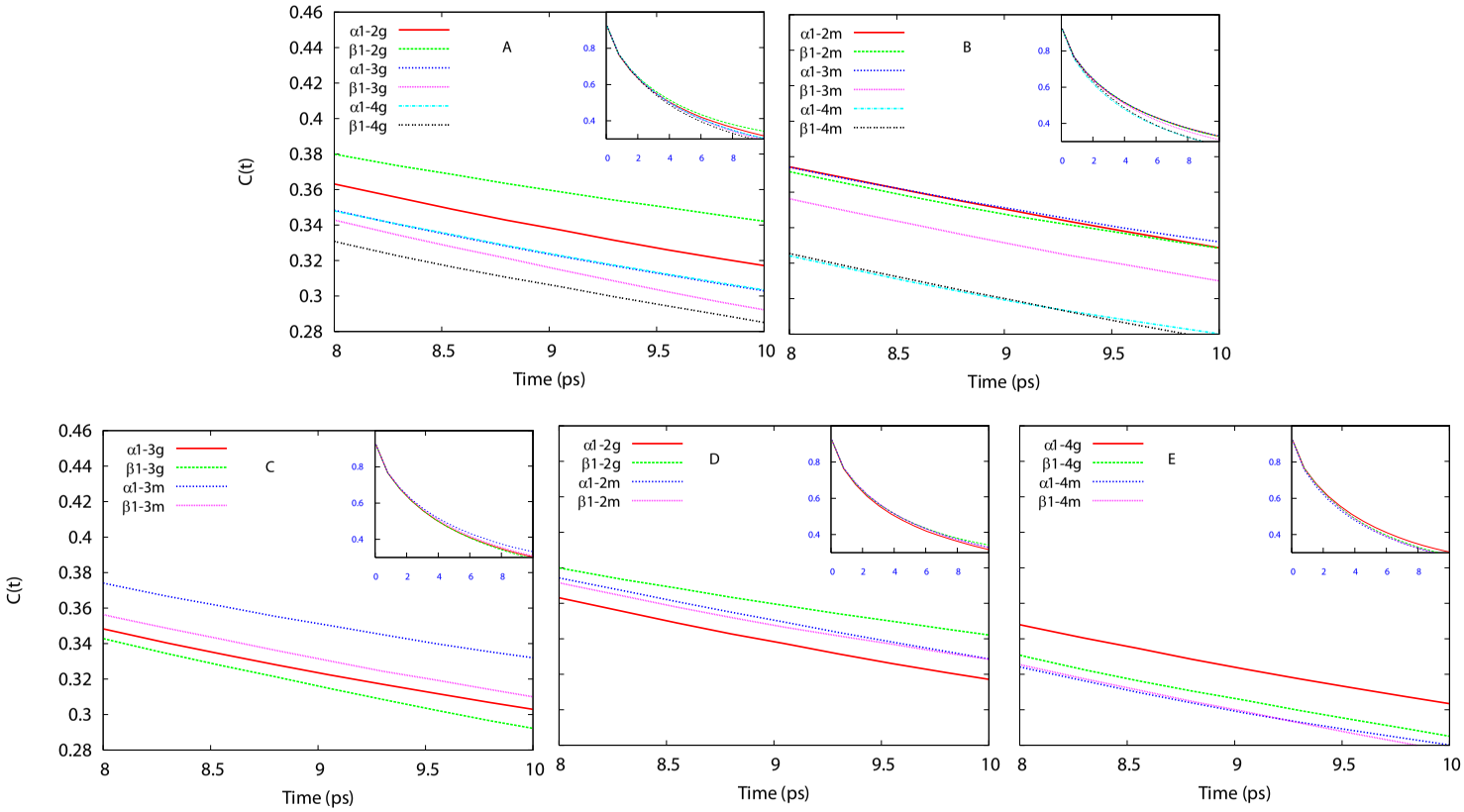


Figure 4.8: Comparison of rotational correlation functions for water on the first solvation shell of homopolysaccharides. (A) shows a comparison between different glucose nonamers, (B) shows a comparison between different mannose nonamers, (C) through (E) compare rotational correlation functions of water surrounding glucose and mannose nonamers with different linkages.

The relationship between the ability of a sugar to form water mediated hydrogen bonds and the decay of surrounding solvent rotational correlation functions has been observed before. It has been shown earlier for the pentasaccharides of cellulose ($\beta 1 \rightarrow 4$-glucose) and maltose ($\alpha 1 \rightarrow 4$-mannose) that solvent rotational correlation functions are slow decaying when compared to bulk water and also when compared to the decay on the surface of disaccharides [117].

The translationally diffusive behavior of water on the surface of saccharides can be studied using equation 2.29. We can get an accurate picture of the similarities and differences in the behavior of water on the surface of wide or extended helices as well as complex branched and small oligosaccharides by analyzing Fig. 4.9. The conclusions that can be derived from this plot are very similar to those obtained while studying rotational diffusion. In general complex branched sugars and wide helices which as we have demonstrated can trap significant amounts of water on their surface due to water mediated hydrogen bonds, show both the slowest decay of solvent rotational correlation and smallest solvent mean square displacements.

A detailed comparison of epimeric differences between mannose and glucose still show that $\alpha 1 \rightarrow 3$-mannose has uncommonly slow water surface diffusion on its surface due to the particular arrangement of atoms displayed in Fig.4.6. In general, just as we found in the case of the rotational correlation functions, the MSD of water on the surface of nonasaccharides with $1 \rightarrow 4$ linkages appear to be largest. In particular $\beta 1 \rightarrow 4$-mannose and glucose appear to be less able to slow down the motion of water in contact with their surface.

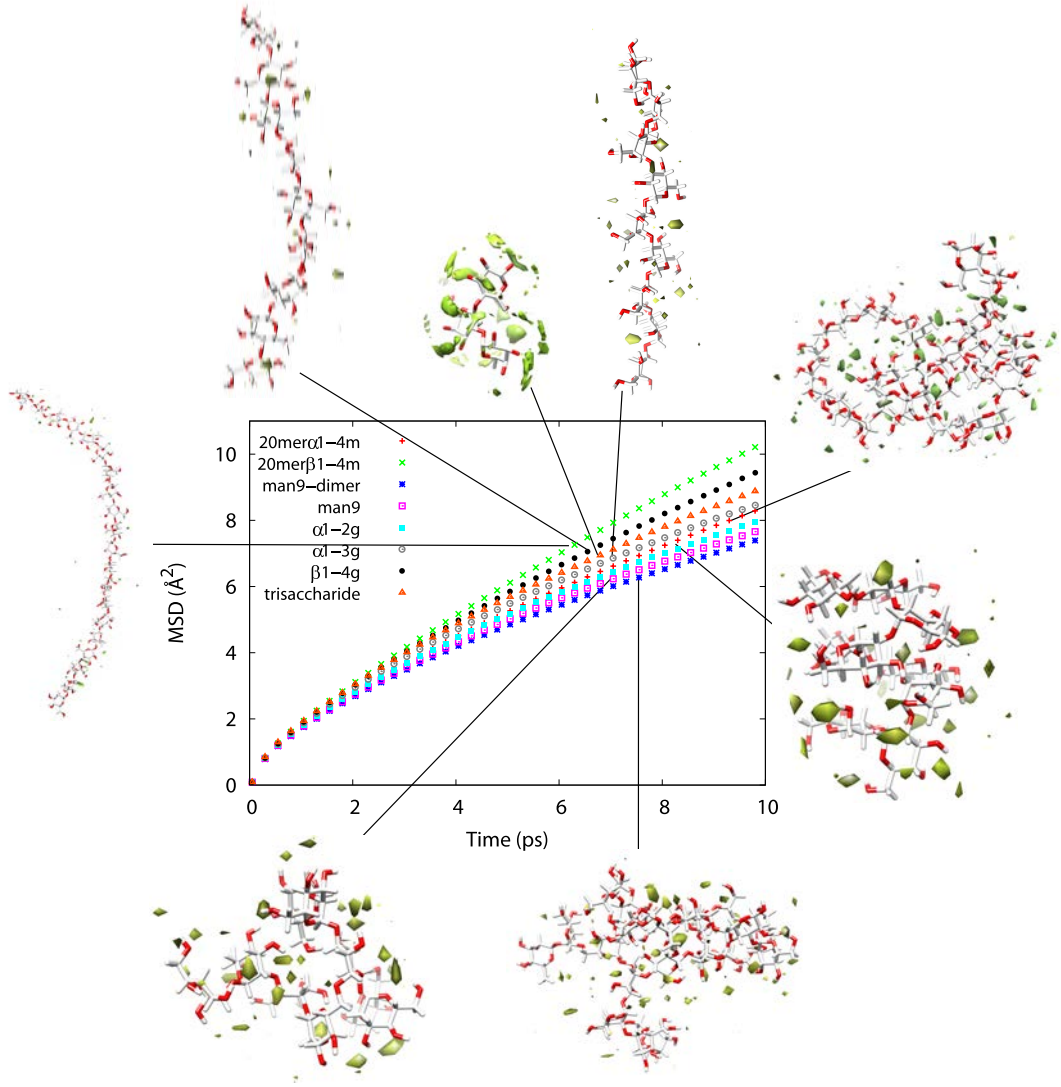


Figure 4.9: The mean square displacement of water around selected saccharide models chosen for this study.

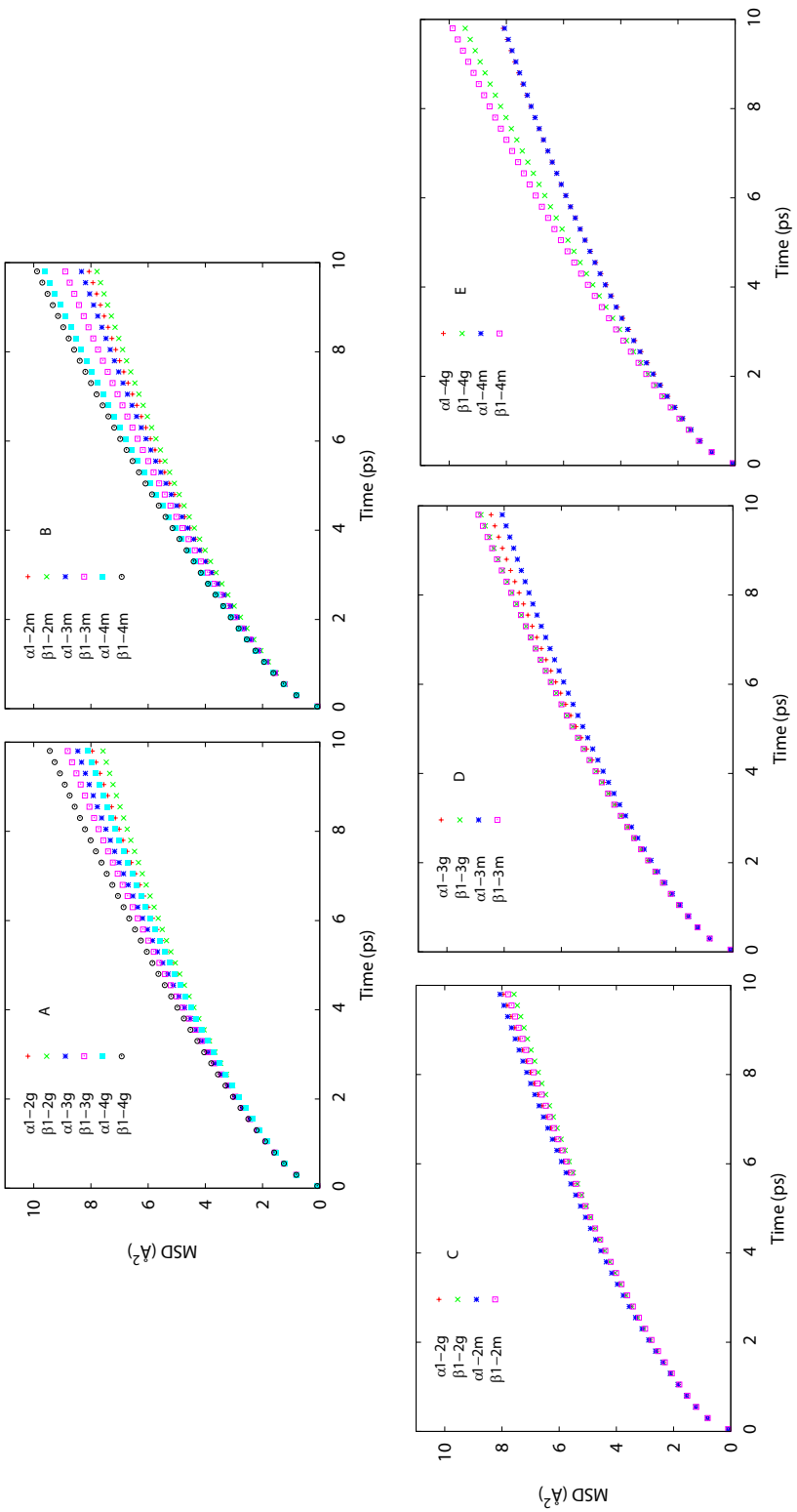


Figure 4.10: A comparison of the the mean square displacement in units of $Å^2$ of water around different nonamers models characterized by their linkage position, anomeric and epimeric configuration.

4.4 Conclusions

As we have shown in previous subsections, geometric effects dramatically affect the behavior of water on the surface of sugars. Branched sugars are particularly well suited to trap water and so are wide helices. In general we find that $\alpha 1 \rightarrow 3$ and $\beta 1 \rightarrow 4$ homopolysaccharides of glucose and mannose form extended helices and it is likely that other homopolysaccharides with the same linkage and anomeric configuration as well as O4 of the second residue in equatorial position will form similar secondary patterns. Though we were not able to find in the literature the structure of homopolysaccharides of allose and altrose, these should form similar patterns to those found in mannose and glucose and based on Table 4.2 we predict that they will have low or intermediate ability to trap water on their surface.

Glucose and mannose are epimers at O2; it is clear that all $\alpha 1 \rightarrow 2$ and $\beta 1 \rightarrow 2$ homopolysaccharides studied form wide helices which are able to trap water for long periods of time and slow down its rotational and translational diffusion. Therefore it appears that both $\alpha 1 \rightarrow 2$ and $\beta 1 \rightarrow 2$ configurations irrespective of whether the second residue is axial or equatorial will form such structures. We therefore predict that all aldohexoses that have a $1 \rightarrow 2$ linkage will form compact helices that are likely to strongly associate with water and slow down its dynamics.

Based on our studies, the $1 \rightarrow 3$ linkages are interesting since both extremes of very low and very high water trapping and slowing can be observed. While the $\beta 1 \rightarrow 3$ version of mannose and glucose homopolysaccharides form wide helical structures in solution with intermediate water trapping and slowing abilities the $\alpha 1 \rightarrow 3$

homopolysaccharides form well defined extended structures. Since mannose and glucose have equatorial O3 configurations, we expect $\alpha 1 \rightarrow 3$ galactose and talose to also form well defined extended helices. Interestingly $\alpha 1 \rightarrow 3$-g has moderate ability to trap and slow down surface water dynamics while $\alpha 1 \rightarrow 3$-m is as highly effective as the branched sugars at modifying the behavior of water on its surface. This is because of the unique arrangement of O6, O2 and O4 in consecutive residues depicted in Fig. 4.6 which is only present in the case of mannose. It would be interesting to study the case of talose, which is an epimer of mannose in O4. Unfortunately this is an uncommon sugar for which a $\alpha 1 \rightarrow 3$ homopolysaccharide has not been studied. In the case of linear saccharides we find that beyond certain number of residues size does not significantly effect residence times as well as other properties.

CHAPTER 5
STRUCTURAL ANALYSIS OF CRYOPROTECTANT SUGAR AQUEOUS SOLUTIONS

5.1 Introduction

Cryoprotectant, as a synonym of antifreeze, is a substance preventing ice formation to protect biological components from freezing damage. Cryoprotection is often noncolligative. Many cryoprotectant proteins, polyols and sugars such as glycerol are naturally used in animals and plants to prevent freezing damage [93, 111]. The mechanism often involves the prevention of crystallization by inducing glassy behavior, or vitrification [32]. For example in the case of stem cells trehalose was studied as a possible replacement for DMSO because of its lower toxicity [23]. Antifreeze glycoproteins are also an important family of bio-relevant cryoprotectants where disaccharides attached to repeating Thr groups appear to play a significant role [119]. The link between cryoprotectant activity and polyols conformation has been previously studied [79, 14], however to the best of my knowledge little is known about the relation between cryoprotection, and saccharide conformations during the process of crystallization.

Three mechanisms of cryoprotection have been proposed in the past. These involve either adsorption inhibition [114, 97, 63], nucleation inhibition [131, 48, 126] or recrystallization inhibition [55]. All of these mechanism are potentially concurrently active. Our studies in the following subsections will attempt to shed light on possible antifreeze mechanisms at very low saccharide concentrations. It is worth mentioning

that most biological systems and medical applications use fairly large concentrations of cryoprotectant agents. Because of this we also explore the effect of inclusion of several saccharides on the structure of water during freezing simulations. This is in line with what we have already presented in Chapter 3 regarding diffusion in crowded environments. The main point of this chapter is to study these systems upon freezing and to compare the situation in the liquid and ice thermodynamic phases.

5.1.1 Previous Simulation Studies

The first successful simulation of homogenous nucleation of ice was first reported in 2002 [80]. Simulation studies dealing with the crystallization of aqueous sodium chloride have also been reported [25, 61]. Debenedetti provided structural analysis of water surrounding trehalose in aqueous liquid phase at 30 and 0 degrees Celsius [67]. Kim reported simulations on the effect of short peptides on the growth of ice crystals [60]. However, to the best of my knowledge, no one has explored the behavior of disaccharides on the growth of ice crystals computationally.

5.2 Force Field and Simulations

Compared to other liquids, many properties of water are abnormal. These include density, specific heat, isothermal compressibility, and thermal expansion [66, 36, 37, 82, 81, 9]. In simulations of water-ice phase changes, it is an important issue to decide which water model to use because all of the available force fields of water perform well only for certain properties. For example, TIP4P/Ice model was specifically designed to account for the properties of the solid amorphous water phases.

TIP4P/2005 has the best fit with real density and thermal expansion coefficients at 1 bar as temperature varies [121].

In this study, we use the TIP5P model for water, which has been shown to provide a reasonable description of the properties of water near its freezing point. This model gives a freezing point of 274 K for ice I_h, which is much closer to the experimental value 273.15 K than other popular water models such as the TIP4P model, which has a freezing point of 232 K, or the SPC/E model, which has a freezing point of 215 K [47, 1, 61].

Our systems were composed of an infinite ice-slab and a body of liquid water with 0, 1, or 4 disaccharides dissolved in it. The body of liquid water was placed between the ice-slab and its periodically replicated image. Configurations of hexagonal ice (with 2048 water molecules) and liquid water (with 10248-10305 water molecules) were prepared separately by using a GROMACS library. Disaccharides were placed in the liquid phase at a distance between 14 and 16 Å away from ice in the x-direction (the normal direction to the original ice-water interface). Overlapping water molecules were removed. In all cases, at the end of the preparation stage, we checked that the shortest distance between the ice-lab and any sugar was in the range between 10 and 12 Å. This way, we ensured that at least at the beginning of our production run, disaccharides were almost free from interactions with the ice slab and the small temperature difference between the ice slab and the water had little effect on the structure and dynamics of the disaccharides. We emphasize that the dissaccharides were not only far from the original slab but also its periodic image.

Systems were equilibrated in two stages: First a 150 ps simulation in the NPT ensemble was run under anisotropic compressibility. Second a 100 ps NPT simulation was run under isotropic compressibility. During the anisotropic-compressing run, the system is only compressible along the normal direction to the ice-slab (x-axis). This guarantees that the liquid achieves proper density in contact with the ice slab without perturbing the ice density. To prevent ice melting during the preparation stage, the temperature of ice was kept at 250 K and liquid water was kept at 260 K.

The size of the system in the initial frame was approximately 160 X 61X 60 Å in size. During production runs the system expands only in the x direction which is the direction of crystallization. This is achieved by running non-equilibrium NPT runs with anisotropic compressibility in which the volume of the system is allowed to change only perpendicular to the ice slab.

All of the MD simulations were carried out using GROMACS v4.0.4. The equations of motion were integrated numerically using a leap-frog algorithm with a time step of 1 fs. The Nose-Hoover coupling method was used to keep the temperature at the desired value with coupling constants of 0.1 ps and Parrinello-Rahman coupling was used for pressure coupling at 1 bar. A cutoff distance of 0.9 nm was used for the LJ interactions, and the particle-mesh Ewald method was used for the long-range electrostatic interactions with a real space cutoff distance of 0.9 nm.

The trajectories were recorded every 25 ps in the first 30 ns and every 5 ps during 30-40 ns.

5.3 Results and Discussions

Figure 5.1 and Fig. 5.2 are examples of typical configuration at the beginning and end of our production runs. Though saccharides appear to be close by in these figures, this is partly because of poor depth perception. For example in Fig. 5.2 the closest saccharides are separated by three ice layers. Fig. 5.3 shows the potential energy change of a typical systems as a function of time during a 44 ns run. As it is clear from this figure, the total potential energy of the system continuously decreases until about 35 ns when the curve becomes flat. This reflects its almost complete crystallization.

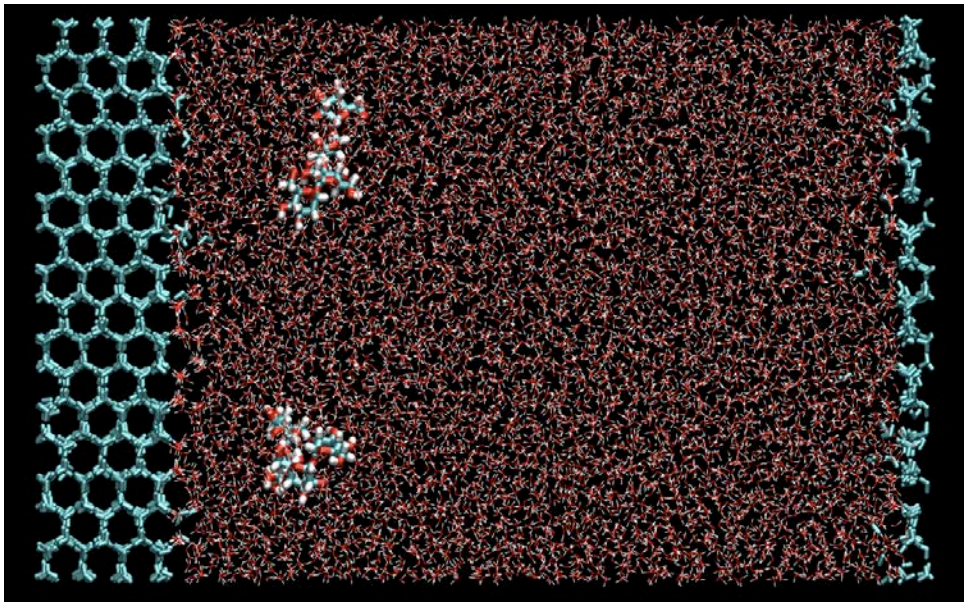


Figure 5.1: Snapshot of a typical simulation before freezing

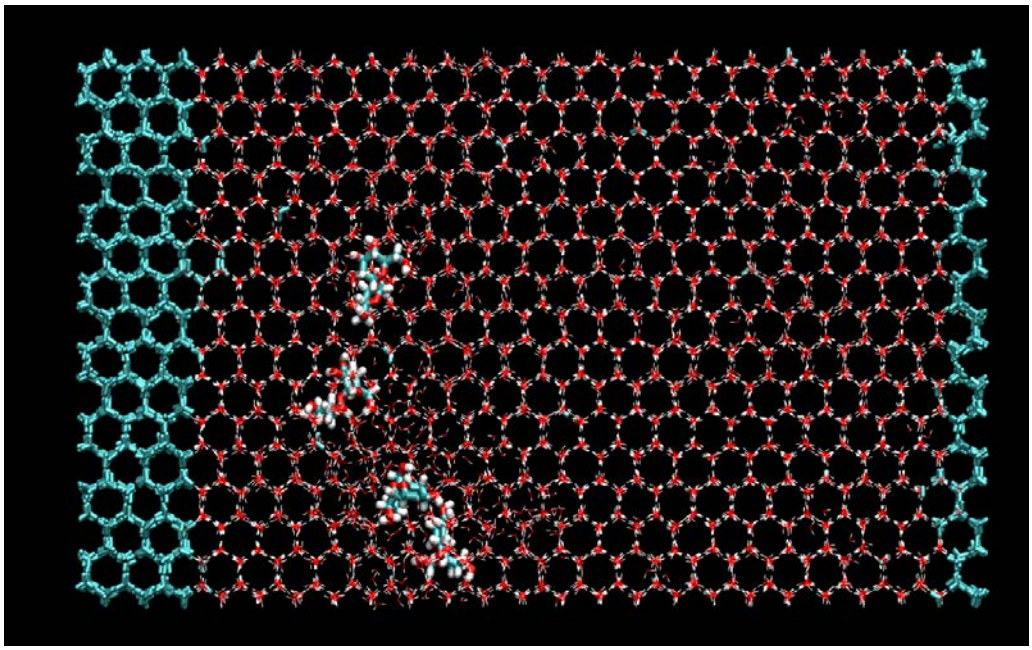


Figure 5.2: Snapshot of a typical simulation after the freezing process is complete

5.3.1 Structural comparison of dimannose and trehalose dissolved in the liquid and solid water phases

Figure 5.4(a) shows that in the liquid state dimannose has a bimodal conformation in its ψ distribution. In contrast, the distribution of ψ of trehalose is centered at a single value. Trehalose is more rigid than di-mannose. This bimodality is absent in the ice state. When water crystallizes the energy barrier for conformational exchange is high and dimannose is trapped in a single preferred conformation.

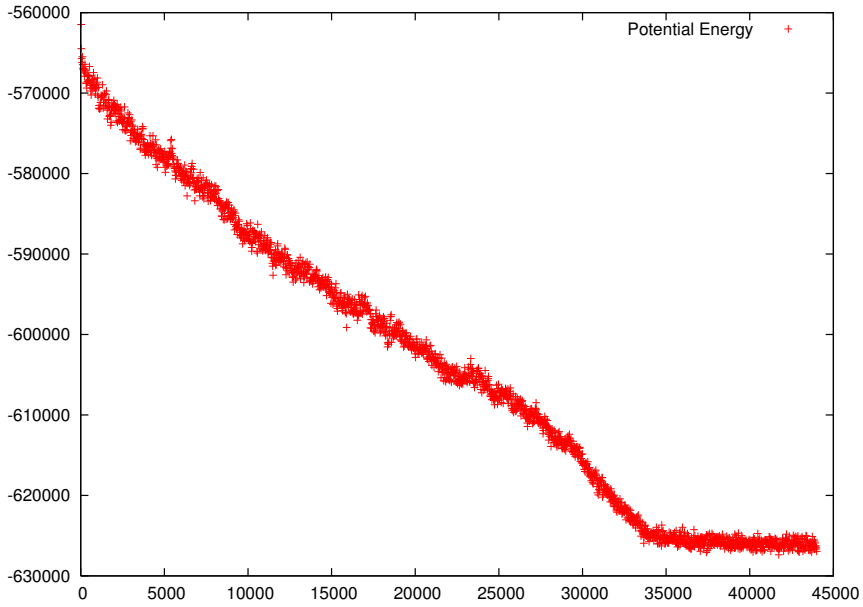


Figure 5.3: The potential energy of the system during the crystallization process

5.3.2 The structure of water surrounding disaccharides in the liquid and solid phases

The spatial distribution of oxygen atoms surrounding sugar oxygen atoms is defined as follows:

$$g(r) = \frac{< N(r) >}{4\pi\rho r^2(dr)} \tag{5.1}$$

where r is the shortest distance from any oxygen atom on the disaccharides(the origin) to the oxygen atom on a water molecule of interest. $4\pi r^2(dr)$ is the volume of a thin shell around a sugar oxygen, and $N(r)$ is the number of water oxygens falling within the thin shell. ρ is the number density of oxygen atoms in the system.

Our definition which considers all oxygen atoms within the sugars, differs from that of Debeneditti [67], who only focused on single hydroxyl oxygens.

In Fig. 5.5 we compare the distribution of water in the neighborhood of dimannose and trehalose at different temperatures in the liquid phase. As expected, the higher the temperature is, the less structured water appears to be around both saccharides. While water behaves similarly around each of the two studied saccharides, it is clear that both at 260 K and 300 K the height of the first peak is different. Trehalose is able to more strongly organize water. This can be appreciated from the higher first peak in g(r) at both studied temperatures.

In the study of Debeneditti and coworkers [67], trehalose is also special with a high first peak in g(r). It is important to emphasize that $O_w - O_w$ g(r) is more structured with higher first g(r) peak than $O_w - O_{sugar}$.

In contrast to previous studies, here we focus on crystallization. Fig. 5.6 illustrates radial distribution functions for the two studied disaccharides in the crystalline state. Here also, trehalose appears to be more structured and this is not only seen from the first peak of g(r) but also in subsequent second and third solvation shells. These results are averaged over three simulations containing each four dissacharides after water had completely frozen around them. It is clear that the small differences in the liquid state are amplified upon freezing.

The tetrahedrality of water surrounding a disaccharides can be studied by using Eq.2.35.

As discussed in section 2.5 $< q >= 1$ corresponds to a perfect tetrahedral structure whereas $< q >= 0$ corresponds to the case of an ideal gas, in which the four vectors arrange randomly.

Fig. 5.7 shows the average tetrahedrality of the oxygen network as a function of the shortest distance from a carbohydrate oxygen to the water oxygen corresponding to the central atom of the tetrahedron. The features in Fig. 5.7 should be understood as follows; at short distance water molecules surrounding the carbohydrate are organized by the sugar's oxygen atoms, at long distances water is organized as in bulk water. It is the intermediate distance that is most interesting since it corresponds to the intervening region between two different oxygen networks and therefore is where disorder and glassy behavior is to be expected. As is to be expected, Fig. 5.7 shows that in general, tetrahedrality is higher at lower temperature. In the liquid phase (or supercooled liquid phase) at distances beyond 3.5Åwater surrounding both carbohydrates appears to be very similar. It is only in the first solvation shell that differences between these two carbohydrates are most noticeable. Trehalose is consistently able to organize water in its immediate surroundings much more strongly that dimannose. It should be pointed out that the distance values in Fig. 5.7 are somewhat misleading since they correspond to the distance from a vertex to the center of a tetrahedron, but the organization is for the whole tetrahedron structure. This means that the organization ability of trehalose may extend somewhat beyond the first solvation layer.

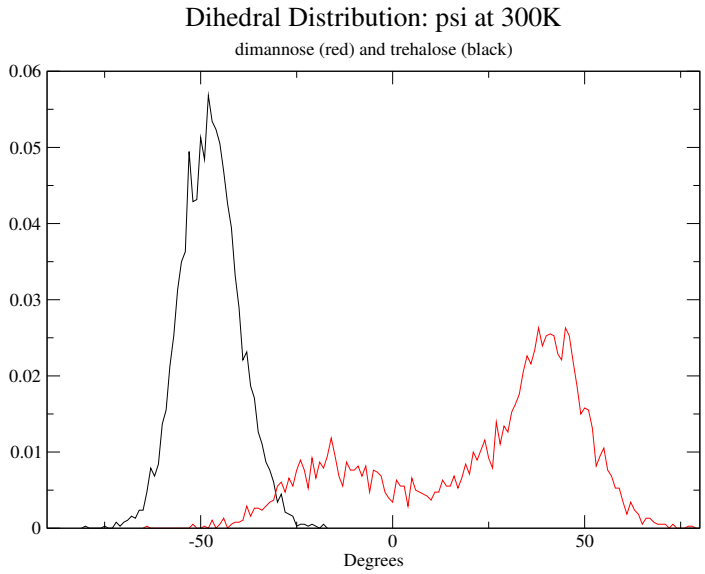
The long distance behavior of the tetrahedrality parameter reflects the intrinsic order of bulk water at the two studied temperatures. It is clear that at least at higher temperatures, tetrahedrality is higher than in the bulk when adjacent to trehalose and also beyond the boundaries of the intervening region in the case of both saccharides. Furthermore, Fig. 5.7 shows that at lower temperatures the intervening region shifts closer towards the saccharides.

Fig. 5.8 shows perhaps the most interesting results in this section. We see from this figure that when these two different carbohydrates are trapped in ice the average of the tetrahedrality parameter q is significantly different in the case of the two carbohydrates. For the most part, beyond the immediately adjacent water layer trehalose appears to disrupt ice even at significantly large distances corresponding to several water layers. It is not surprising then that trehalose is experimentally known to be a good cryoprotectant. It is interesting that while dimannose appears to be worse at properly fitting within the tetrahedral structure of ice at close contact, the effect is quickly diminished when compared to the case of trehalose.

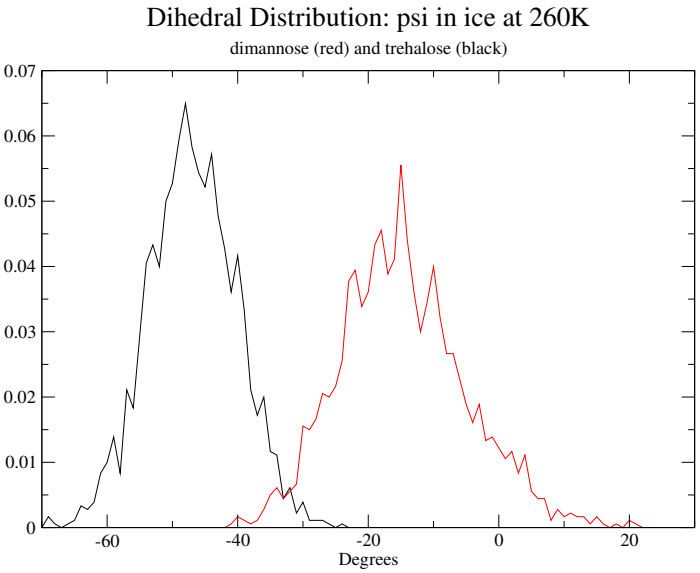
Compared to di-mannose, in our simulations trehalose shows a more disruptive effect on ice growth in the range 4.4-8 Å. This supports the adsorption inhibition mechanism. Finally, we show in Fig. 5.9 that when the distance between the surfaces of different disaccharides is shorter than about 8 Å, water between these appears to be amorphous even within ice. It is therefore not surprising that at experimental concentrations (much higher than in our studies, sometimes between 40% and 60% wt) water is in a glassy state.

5.4 Conclusion

In this chapter, we investigated structural properties of water and ice adjacent to individual disaccharides or a group of disaccharides corresponding to solutions of 1% by wt. Whereas in the liquid phase trehalose appears to more strongly organize water in its surroundings, in the ice phase it appears to disrupt the crystal phase the most. Water adjacent to these saccharides is glassy and this glassy structure extends further away in the case of trehalose. When the sugar concentration is larger, the intervening space between different sugars is not crystalline even if this distance is on the order of a nanometer.



(a) psi dihedral distribution for dimannose and trehalose in the liquid phase



(b) Same as 5.4(a) but in the ice phase

Figure 5.4: In the liquid phase the psi distribution in the case of dimannose is bimodal and is unimodal in the case of trehalose. Upon freezing both distributions become unimodal.

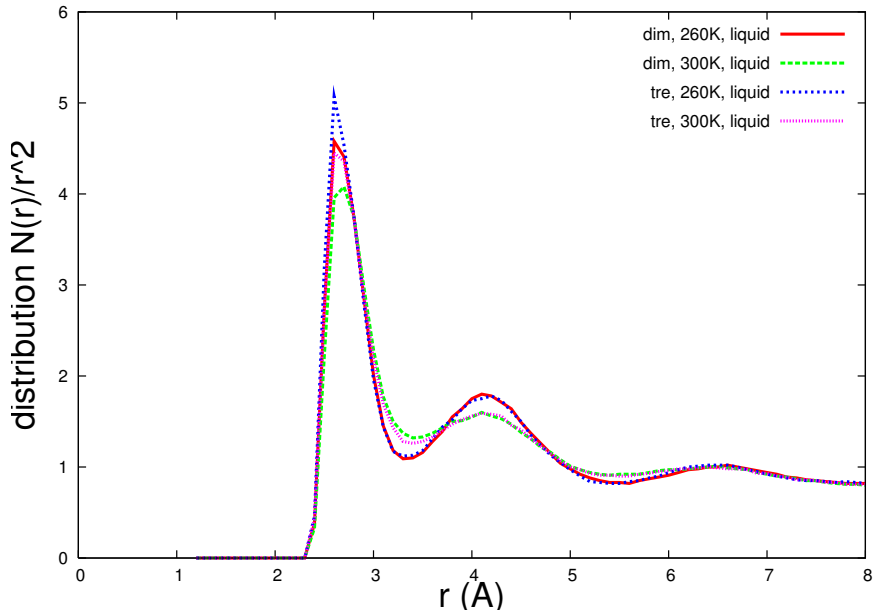


Figure 5.5: The radial distribution function of oxygen on water in liquid state respect to the oxygen on carbohydrates at different temperatures

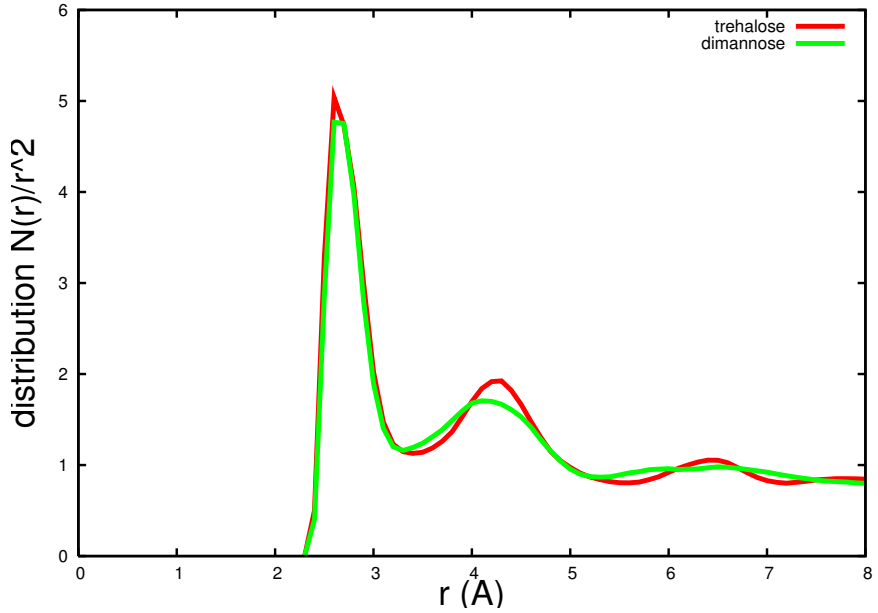


Figure 5.6: The radial distribution function of oxygen on water in crystalized state respect to the oxygen on carbohydrates at 260K

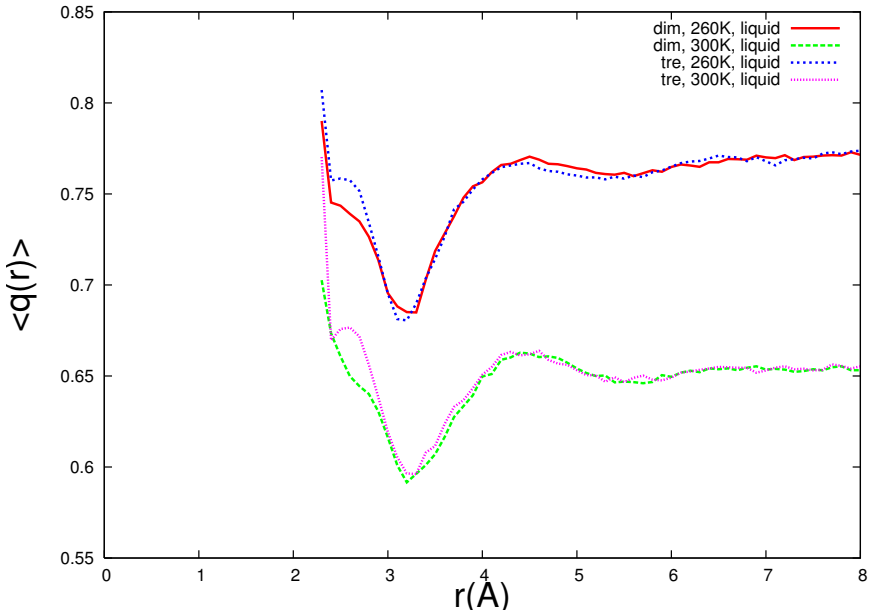


Figure 5.7: The tetrahedrality q of liquid water surrounding various carbohydrates at different temperatures.

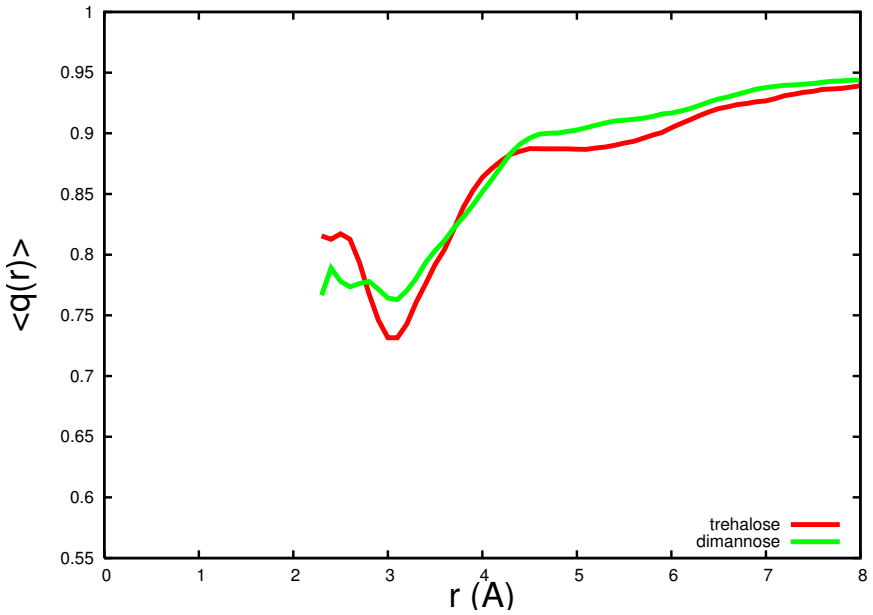


Figure 5.8: The average tetrahedrality q of crystalized water surrounding various carbohydrates at 260K

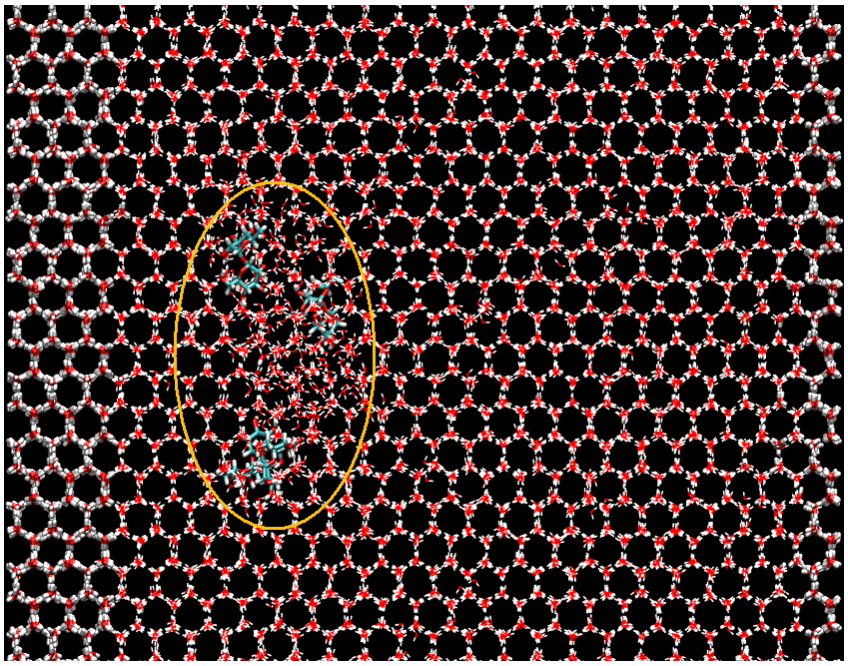


Figure 5.9: Even within ice, if saccharides are separated by distances that are smaller than 1nm the intervening space will likely look like a glass and not a crystal. We highlight this with a yellow ellipsoid.

CHAPTER 6
CONCLUSION AND FUTURE WORK

The non-trivial structural and dynamical bio-relevance of water has recently attracted considerable attention across scientific communities. In this thesis, I have investigated water structure and dynamics within the context of different biomolecules in the liquid, glassy and crystalline phases. In addition, and to accomplish these goals we have developed new theoretical methods and computational algorithms to better understand diffusivity patterns and residence times in complex environments where inherent simulation statistics are by nature poor.

Our results indicate that the cumulative probability distribution of first passage times is less error prone than other methodologies previously used to study residence times of molecules on different surfaces. This is due to the better treatment of molecules with long times of residence for which MD studies do not provide good statistics. Our proposed algorithm presents a better way to estimate the times of residence of a molecule on a rough molecular surface and is supported by a clear analytical link between the current formulation and previously proposed survival time correlation functions.

In the study of water behavior in the vicinity of oligosaccharides, we find the secondary structures, sizes, branching and other geometric factors strongly affect the behavior of water in contact with carbohydrates. Branched sugars and carbohydrates with wide helices are the most prone to water trapping. Longer water residence times as well as slower rotational and translational diffusion are observed when in contact

with wide helices and branched sugars. In contrast, water in contact with extended helices and smaller oligosaccharides, appears to display faster and less hindered dynamics. Clearly branching, the type of linkage between monomers and the anomeric configuration are all key elements that determine the structural and dynamical features of water in the immediate vicinity of oligosaccharides.

When studying the structure of water and dissacharides upon freezing we find that trehalose can act as an organizing center in the liquid phase and a strongly disorganizing object in the ice phase. A saccharide concentration increase leads to glassy water in the intervening space between them surrounded by ice. This glassy layer is not small and distortion of the ice lattice is observed even when the separation between saccharides is on the nm scale. This project is ongoing and we expect to be able to discuss not only initial (liquid) and final (ice) states but also structural changes during the process of freezing in a future publication.

Because of the enormous computational expense of running large freezing simulations, we have not yet investigated the case of other relevant sugars. A more thorough comparison across concentrations and antifreeze species should be part of future directions. For example sucrose is also known to be a good anti-freeze molecule and the comparison between this molecule and trehalose is of potential interest. It is possible that the rigidity of the dihedral angle in disaccharides may play a role in determining its antifreezing ability. In my simulations trehalose apears to be more rigid than dimannose. This observation should be explored before further conclusions can be made and could also be part of the future directions of this project. It is worth

mentioning that experimentally, there is nearly no difference in the antifreezing abilities of trehalose, sucrose, and glucose at very low concentrations. Differences are only observed at higher concentrations when glassy structure and dynamics is likely to be prevalent.

REFERENCES

[1] Jose L. F Abascal, Ramon Garcia Fernandez, Carlos Vega, and Marcelo A Carignano. The melting temperature of the six site potential model of water. *J Chem Phys*, 125(16):166101, Jan 2006.

[2] P Adhikari, K Bachhawat-Sikder, C J Thomas, R Ravishankar, A A Jeyaprakash, V Sharma, M Vijayan, and A Surolia. Mutational analysis at asn-41 in peanut agglutinin. a residue critical for the binding of the tumor-associated thomsen-friedenreich antigen. *J Biol Chem*, 276(44):40734–9, Nov 2001.

[3] N Akiya and PE Savage. Effect of water density on hydrogen peroxide dissociation in supercritical water. 1. reaction equilibrium. *J Phys Chem A*, 104(19):4433–4440, Jan 2000.

[4] N Akiya and PE Savage. Effect of water density on hydrogen peroxide dissociation in supercritical water. 2. reaction kinetics. *J Phys Chem A*, 104(19):4441–4448, Jan 2000.

[5] M.P. Allen and D.J. Tildesley. *Computer Simulation of Molecular Liquids*. Clarendon Press, Oxford, 1987.

[6] A Almond. Towards understanding the interaction between oligosaccharides and water molecules. *Carbohyd Res*, 340(5):907–920, Jan 2005.

[7] A Almond, A Brass, and JK Sheehan. Oligosaccharides as model systems for understanding water-biopolymer interaction: Hydrated dynamics of a hyaluronan decamer. *J Phys Chem B*, 104(23):5634–5640, Jan 2000.

[8] A Almond and JK Sheehan. Predicting the molecular shape of polysaccharides from dynamic interactions with water. *Glycobiology*, 13(4):255–264, Jan 2003.

[9] C. Austen Angell. Insights into phases of liquid water from study of its unusual glass-forming properties. *Science*, 319(5863):582–587, Jan 2008.

[10] James N Arnold, Mark R Wormald, Robert B Sim, Pauline M Rudd, and Raymond A Dwek. The impact of glycosylation on the biological function and structure of human immunoglobulins. *Annu Rev Immunol*, 25:21–50, Jan 2007.

[11] TY Aw. Molecular and cellular responses to oxidative stress and changes in oxidation-reduction imbalance in the intestine. *Am J Clin Nutr*, 70(4):557–565, Jan 1999.

[12] B Bagchi. Water solvation dynamics in the bulk and in the hydration layer of proteins and self-assemblies. *Annu. Rep. Prog. Chem., Sect. C*, 99:127–175, 2003.

[13] B Bagchi. Water dynamics in the hydration layer around proteins and micelles. *Chem Rev*, 105(9):3197–3219, Jan 2005.

[14] Eti Baruch and Yitzhak Mastai. Antifreeze properties of polyglycidol block copolymers. *MACROMOLECULAR RAPID COMMUNICATIONS*, 28(23):2256–2261, Dec 2007.

[15] FC Bawden, NW Pirie, JD Bernal, and I Fankuchen. Liquid crystalline substances from virusinfected plants. *Nature*, 138:1051–1052, Jan 1936.

[16] David A C Beck, Darwin O V Alonso, and Valerie Daggett. A microscopic view of peptide and protein solvation. *Biophys Chem*, 100(1-3):221–37, Jan 2003.

[17] AR Bizzarri and S Cannistraro. Molecular dynamics simulation evidence of anomalous diffusion of protein hydration water. *Phys Rev E*, 53(4):R3040–R3043, Jan 1996.

[18] AR Bizzarri and S Cannistraro. Molecular dynamics of water at the protein-solvent interface. *J Phys Chem B*, 106(26):6617–6633, Jan 2002.

[19] TL Bluhm and A Sarko. Triple helical structure of lentinan, a linear beta-(1-]3)-d-glucan. *Can J Chem*, 55(2):293–299, Jan 1977.

[20] CL Brooks and M Karplus. Solvent effects on protein motion and protein effects on solvent motion - dynamics of the active-site region of lysozyme. *J Mol Biol*, 208(1):159–181, Jan 1989.

[21] F Bruge, E Parisi, and SL Fornili. Effects of simple model solutes on water dynamics: Residence time analysis. *Chemical Physics Letters*, 250(5-6):443–449, Jan 1996.

[22] Andrey V Brukhno, Jamshed Anwar, Ruslan Davidchack, and Richard Handel. Challenges in molecular simulation of homogeneous ice nucleation. *J Phys-Condens Mat*, 20(49):494243, Jan 2008.

[23] Sandhya S Buchanan, Sherilyn A Gross, Jason P Acker, Mehmet Toner, John F Carpenter, and David W Pyatt. Cryopreservation of stem cells using trehalose: evaluation of the method using a human hematopoietic cell line. *Stem Cells Dev*, 13(3):295–305, Jun 2004.

[24] Garry R Buettner, Chin F Ng, Min Wang, V G J Rodgers, and Freya Q Schafer. A new paradigm: manganese superoxide dismutase influences the production of h2o2 in cells and thereby their biological state. *Free Radic Biol Med*, 41(8):1338–50, Oct 2006.

[25] MA Carignano, E Baskaran, PB Shepson, and I Szleifer. Molecular dynamics simulation of ice growth from supercooled pure water and from salt solution. *Annals of Glaciology*, 44:113, 2006.

[26] D Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, Jan 2005.

[27] PL CHAU and AJ HARDWICK. A new order parameter for tetrahedral configurations. *Mol Phys*, 93(3):511–518, Jan 1998.

[28] DS Chekmarev, T Ishida, and RM Levy. Long-time conformational transitions of alanine dipeptide in aqueous solution: Continuous and discrete-state kinetic models. *J. Phys. Chem. B*, 108(1):19487–19495, Jan 2004.

[29] YK Cheng and PJ Rossky. Surface topography dependence of biomolecular hydrophobic hydration. *Nature*, 392(6677):696–699, Jan 1998.

[30] Y Chung, J Xia, and C Margulis. Diffusion and residence time of hydrogen peroxide and water in crowded protein environments. *J. Phys. Chem. B*, Jan 2007.

[31] YH Chung. *The Diffusion of Small Molecules in a Crowded EnVironment: The Case of Water and Hydrogen Peroxide*. Thesis, University of Iowa, IA, 2006.

[32] J S Clegg. Cryptobiosis–a peculiar state of biological organization. *Comp Biochem Physiol B, Biochem Mol Biol*, 128(4):613–24, Apr 2001.

[33] L Cordone, G Cottone, and S Giuffrida. Role of residual water hydrogen bonding in sugar/water/biomolecule systems: a possible explanation for trehalose peculiarity. *J Phys-Condens Mat*, 19(20):205110, Jan 2007.

[34] T Darden, D YORK, and L PEDERSEN. Particle mesh ewald - an n.log(n) method for ewald sums in large systems. *J Chem Phys*, 98(12):10089–10092, Jan 1993.

[35] Jennifer L Dashnau, Kim A Sharp, and Jane M Vanderkooi. Carbohydrate intramolecular hydrogen bonding cooperativity and its effect on water structure. *J Phys Chem B*, 109(50):24152–9, Dec 2005.

[36] PG Debenedetti. Supercooled and glassy water. *J Phys-Condens Mat*, 15(45):R1669–R1726, Jan 2003.

[37] PG Debenedetti and HE Stanley. Supercooled and glassy water. *Phys Today*, 56(6):40–46, Jan 2003.

[38] JP Declercq, C Evrard, A Clippe, D Vander Stricht, A Bernard, and B Knoops. Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 angstrom resolution. *J Mol Biol*, 311(4):751–759, Jan 2001.

[39] Mari L DeMarco and Robert J Woods. Structural glycobiology: a game of snakes and ladders. *Glycobiology*, 18(6):426–40, Jun 2008.

[40] Danielle H Dube and Carolyn R Bertozzi. Glycans in cancer and inflammation–potential for therapeutics and diagnostics. *Nat Rev Drug Discov*, 4(6):477–88, Jun 2005.

[41] Raymond A Dwek. Glycobiology: Toward understanding the function of sugars. *Chem Rev*, 96(2):683–720, Mar 1996.

[42] SB Engelsen, C eline Monteiro, CH DUPENHOAT, and S Perez. The diluted aqueous solvation of carbohydrates as inferred from molecular dynamics simulations and nmr spectroscopy. *Biophys. Chem.*, 93:103–127, Jun 2001.

[43] SB Engelsen and S Perez. The hydration of sucrose. *Carbohyd Res*, 292:21–38, Jan 1996.

[44] JR Errington and PG Debenedetti. Relationship between structural order and the anomalies of liquid water. *Nature*, 409(6818):318–321, Jan 2001.

[45] M Falconi, M Brunelli, A Pesce, M Ferrario, M Bolognesi, and A Desideri. Static and dynamic water molecules in cu,zn superoxide dismutase. *Proteins*, 51(4):607–615, Jan 2003.

[46] DM Ferguson. Parameterization and evaluation of a flexible water model. *J. Comp. Chem.*, 16(4):501–511, Jan 1995.

[47] RG Fernandez, JLF Abascal, and C Vega. The melting point of ice i-h for common water models calculated from direct coexistence of the solid-liquid interface. *J Chem Phys*, 124(14):144506, Jan 2006.

[48] F Franks, J Darlington, T Schenz, SF Mathias, L Slade, and H Levine. Antifreeze activity of antarctic fish glycoprotein and a synthetic-polymer. *Nature*, 325(6100):146–147, Jan 1987.

[49] AD French and VG Murphy. Effects of changes in ring geometry on computer models of amylose. *Carbohyd Res*, 27(2):391–406, Jan 1973.

[50] D.A. French. Conformational accessibility of some simple polyglucosides. *ACS Symp. Ser.*, 260:43–59, 1984.

[51] AE Garcia and L Stiller. Computation of the mean residence time of water in the hydration shells of biomolecules. *J Comput Chem*, 14(11):1396–1406, Jan 1993.

[52] B Halle. Cross-relaxation between macromolecular and solvent spins: The role of long-range dipole couplings. *J Chem Phys*, 119(23):12372–12385, Jan 2003.

[53] M Heyden, E Bruendermann, U Heugen, G Niehues, D. M Leitner, and M Havenith. Long-range influence of carbohydrates on the solvation dynamics of water-answers from terahertz absorption measurements and molecular modeling simulations. *J Am Chem Soc*, 130(17):5773–5779, Jan 2008.

[54] L Hua, XH Huang, RH Zhou, and BJ Berne. Dynamics of water confined in the interdomain region of a multidomain protein. *J Phys Chem B*, 110(8):3704–3711, Jan 2006.

[55] Arthur T. Hubbard. *Encyclopedia of Surface and Colloid Science*. CRC Press, 2004.

[56] A Imberty and S Perez. Structure, conformation, and dynamics of bioactive oligosaccharides: Theoretical approaches and experimental validations. *Chem. Rev.*, 100:4567–4588, Jan 2000.

[57] RW Impey, PA Madden, and IR McDonald. Hydration and mobility of ions in solution. *J Phys Chem-Us*, 87(25):5071–5083, Jan 1983.

[58] WL Jorgensen, J Chandrasekhar, JD Madura, RW Impey, and ML Klein. Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, 79(2):926–935, Jan 1983.

[59] WL Jorgensen, DS Maxwell, and J TiradoRives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J Am Chem Soc*, 118(45):11225–11236, Jan 1996.

[60] Jun Soo Kim, Srinivasan Damodaran, and Arun Yethiraj. Retardation of ice crystallization by short peptides. *J Phys Chem A*, 113(16):4403–4407, Jan 2009.

[61] Jun Soo Kim and Arun Yethiraj. The effect of salt on the melting of ice: A molecular dynamics simulation study. *J Chem Phys*, 129(12):124504, Jan 2008.

[62] K N Kirschner and R J Woods. Solvent interactions determine carbohydrate conformation. *P Natl Acad Sci Usa*, 98(19):10541–5, Sep 2001.

[63] CA Knight and AL Devries. Melting inhibition and superheating of ice by an antifreeze glycopeptide. *Science*, 245(4917):505–507, Jan 1989.

[64] D Case.; T Darden; TE Cheatham III; C Simmerling; J Wang; R Duke; R Luo; K Merz; D Pearlman; M Crowley; R Walker; W Zhang; B Wang; S Hayik; A Roitberg; G Seabra; K Wong; F Paesani; X Wu; S Brozell; V Tsui; H Gohlke; L Yang; C Tan; J Mongan; V Hornak; G Cui; P Beroza; D Mathews; C Schafmeister; W Ross; P Kollman. *AMBER 9*. University of California, San Fransisco, San Fransisco, CA, 2006.

[65] J Koput, S Carter, and NC Handy. Potential energy surface and vibrational-rotational energy levels of hydrogen peroxide. *J Phys Chem A*, 102(31):6325–6330, Jan 1998.

[66] Pradeep Kumar, Giancarlo Franzese, and H. Eugene Stanley. Dynamics and thermodynamics of water. *J Phys-Condens Mat*, 20(24):244114, Jan 2008.

[67] SL Lee, PG Debenedetti, and JR Errington. A computational study of hydration, solution structure, and dynamics in dilute carbohydrate solutions. *J Chem Phys*, 122(20):204511, Jan 2005.

[68] B Leroux, H Bizot, JW Brady, and V Tran. Water structuring around complex solutes: Theoretical modeling of alpha-d-glucopyranose. *Chem Phys*, 216(3):349–363, Jan 1997.

[69] M Levitt and R Sharon. Accurate simulation of protein dynamics in solution. *P Natl Acad Sci Usa*, 85(20):7557–61, Oct 1988.

[70] Q Liu, RK Schmidt, B Teo, PA Karplus, and JW Brady. Molecular dynamics studies of the hydration of alpha,alpha-trehalose. *J Am Chem Soc*, 119(33):7851–7862, Jan 1997.

[71] V Lounnas and BM Pettitt. A connected cluster of hydration around myoglobin - correlation between molecular-dynamics simulations and experiment. *Proteins*, 18(2):133–147, Jan 1994.

[72] A Luise, M Falconi, and A Desideri. Molecular dynamics simulation of solvated azurin: Correlation between surface solvent accessibility and water residence times. *Proteins*, 39(1):56–67, Jan 2000.

[73] RM Lyndenbell, JS Vanduijneveldt, and D Frenkel. Free-energy changes on freezing and melting ductile metals. *Mol Phys*, 80(4):801–814, Jan 1993.

[74] VA Makarov, BK Andrews, PE Smith, and BM Pettitt. Residence times of water molecules in the hydration sites of myoglobin. *Biophys J*, 79(6):2966–2974, Jan 2000.

[75] VA Makarov, M Feig, BK Andrews, and BM Pettitt. Diffusion of solvent around biomolecular solutes: A molecular dynamics simulation study. *Biophys J*, 75(1):150–158, Jan 1998.

[76] C J Margulis. Computational study of the dynamics of mannose disaccharides free in solution and bound to the potent anti-hiv virucidal protein cyanovirin. *J Phys Chem B*, 109(8):3639–47, Mar 2005.

[77] Marilia T. C Martins-Costa and Manuel F Ruiz-Lopez. Molecular dynamics of hydrogen peroxide in liquid water using a combined quantum/classical force field. *Chem Phys*, 332(2-3):341–347, Jan 2007.

[78] F Massi and JE Straub. Structural and dynamical analysis of the hydration of the alzheimer's beta-amyloid peptide. *J Comput Chem*, 24(2):143–153, Jan 2003.

[79] Yitzhak Mastai, Jan Rudloff, Helmut Cölfen, and Markus Antonietti. Control over the structure of ice and water by block copolymer additives. *Chemphyschem*, 3(1):119–23, Jan 2002.

[80] M Matsumoto, S Saito, and I Ohmine. Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature*, 416(6879):409–413, Jan 2002.

[81] Marco G Mazza, Nicolas Giovambattista, H. Eugene Stanley, and Francis W Starr. Connection of translational and rotational dynamical heterogeneities with the breakdown of the stokes-einstein and stokes-einstein-debye relations in water. *Phys Rev E*, 76(3):031203, Jan 2007.

[82] MG Mazza, N Giovambattista, FW Starr, and HE Stanley. Relation between rotational and translational dynamic heterogeneities in water. *Phys Rev Lett*, 96(5):057803, Jan 2006.

[83] K Modig, E Liepinsh, G Otting, and B Halle. Dynamics of protein and peptide hydration. *J Am Chem Soc*, 126(1):102–114, Jan 2004.

[84] I Muegge and EW Knapp. Residence times and lateral diffusion of water at protein surfaces - application to bpti. *J Phys Chem-Us*, 99(5):1371–1374, Jan 1995.

[85] N Nandi and B Bagchi. Dielectric relaxation of biological water. *J Phys Chem B*, 101(50):10954–10961, Jan 1997.

[86] K Ogawa, K Okamura, and A Sarko. Packing analysis of carbohydrates and polysaccharides .12. molecular and crystal-structure of the regenerated form of (1-]3)-alpha-d-glucan. *Int J Biol Macromol*, 3(1):31–36, Jan 1981.

[87] G Otting, E Liepinsh, and K Wuthrich. Protein hydration in aqueous-solution. *Science*, 254(5034):974–980, Jan 1991.

[88] SK Pal, J Peon, B Bagchi, and AH Zewail. Biological water: Femtosecond dynamics of macromolecular hydration. *J Phys Chem B*, 106(48):12376–12395, Jan 2002.

[89] Marco Paolantoni, Paola Sassi, Assunta Morresi, and Sergio Santini. Hydrogen bond dynamics and water structure in glucose-water solutions by depolarized rayleigh scattering and low-frequency raman spectroscopy. *J Chem Phys*, 127(2):024504, Jan 2007.

[90] S Park and JG Saven. Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins*, 60(3):450–463, Jan 2005.

[91] S Perez, A Imberty, and RP Scaringe. Modeling of interactions of polysaccharide chains. *ACS Symp. Ser.*, 430:281–299, 1990.

[92] EF Pettersen, TD Goddard, CC Huang, GS Couch, DM Greenblatt, EC Meng, and TE Ferrin. Ucsf chimera - a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–1612, Jan 2004.

[93] Thomas D Pfister and Kenneth B Storey. Insect freeze tolerance: Roles of protein phosphatases and protein kinase a. *Insect Biochem Mol Biol*, 36(1):18–24, Jan 2006.

[94] Francesco Pizzitutti, Massimo Marchi, Fabio Sterpone, and Peter J Rossky. How protein surfaces induce anomalous dynamics of hydration water. *J Phys Chem B*, 111(26):7584–7590, Jan 2007.

[95] LR Pratt and A Pohorille. Hydrophobic effects and modeling of biophysical aqueous solution interfaces. *Chem Rev*, 102(8):2671–2691, Jan 2002.

[96] V. S. R. Rao, P. K. Qasba, P.V. Balaji, and R. Chandrashekaran. *Conformation of Carbohydrates*. Harwood Academic Publishers, Amsterdam, The Netherlands, 1998.

[97] J A Raymond and A L DeVries. Freezing behavior of fish blood glycoproteins with antifreeze properties. *Cryobiology*, 9(6):541–7, Dec 1972.

[98] S Raymond, B Henrissat, DT Qui, A Kvick, and H Chanzy. The crystal-structure of methyl beta-cellotrioside monohydrate-0.25 ethanolate and its relationship to cellulose-ii. *Carbohyd Res*, 277(2):209–229, Jan 1995.

[99] S Redner. *A Guide to First Passage Processes*. Cambridge University Press, Cambridge, England, 2001.

[100] D.A. Rees. *Polysaccharide Shapes*. Chapman and Hall, London, 1977.

[101] S G Rhee. Redox signaling: hydrogen peroxide as intracellular messenger. *Exp Mol Med*, 31(2):53–9, Jun 1999.

[102] C Rocchi, AR Bizzarri, and S Cannistraro. Water residence times around copper plastocyanin: A molecular dynamics simulation approach. *Chem Phys*, 214(2-3):261–276, Jan 1997.

[103] C Rocchi, AR Bizzarri, and S Cannistraro. Water dynamical anomalies evidenced by molecular-dynamics simulations at the solvent-protein interface. *Phys Rev E*, 57(3):3315–3325, Jan 1998.

[104] S ROY and B Bagchi. Solvation dynamics in liquid water. a novel interplay between librational and diffusive modes. *J Chem Phys*, 99(2):9938, Jan 1993.

[105] BK Sathyana and VSR Rao. Conformational studies of beta-glucans. *Biopolymers*, 10(9):1605–&, Jan 1971.

[106] BK Sathyana and VSR Rao. Conformational studies of alpha-glucans. *Biopolymers*, 11(7):1379–&, Jan 1972.

[107] F Q Schafer and G R Buettner. Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. *Free Radic Biol Med*, 30(11):1191–212, Jun 2001.

[108] M Settles and W Doster. Anomalous diffusion of adsorbed water: a neutron scattering study of hydrated myoglobin. *Faraday Discuss*, 99:269–279, 1996.

[109] JP Simons, RA Jockusch, P Carcabal, I Hunig, R T Kroemer, NA MacLeod, and LC Snoek. Sugars in the gas phase. spectroscopy, conformation, hydration, co-operativity and selectivity. *Int. Rev. in Phys. Chem.*, 24(3):489–531, Jan 2005.

[110] PJ Steinhardt, DR Nelson, and M Ronchetti. Bond-orientational order in liquids and glasses. *Phys Rev B*, 28(2):784–805, Jan 1983.

[111] K B Storey and J M Storey. Freeze tolerance in animals. *Physiol Rev*, 68(1):27–84, Jan 1988.

[112] Mannargudi S Sujatha, Yellamraju U Sasidhar, and Petety V Balaji. Energetics of galactose- and glucose-aromatic amino acid interactions: implications for binding in galactose-specific proteins. *Protein Sci*, 13(9):2502–14, Sep 2004.

[113] MM Teeter. Water-protein interactions - theory and experiment. *Annu Rev Biophys Bio*, 20:577–600, Jan 1991.

[114] WA Tiller. *The Science of Crystallization: Microscopic Inter-facial Phenomena*. Cambridge University Press: Cambridge, UK, 1991.

[115] EJ Toone. Structure and energetics of protein carbohydrate complexes. *Curr Opin Struc Biol*, 4(5):719–728, Jan 1994.

[116] SM Tschampel and RJ Woods. Quantifying the role of water in protein-carbohydrate interactions. *J Phys Chem A*, 107(43):9175–9181, Jan 2003.

[117] M Umemura, S Hayashi, T Nakagawa, H Urakawa, and K Kajiwara. Structure of water molecules in aqueous solutions of di-and penta-d-glucopyranoses using molecular dynamics simulation. *J Mol Struc-Theochem*, 639:69–86, Jan 2003.

[118] Robert Vacha, Petr Slav cek, M Mucha, Barbara J Finlayson-Pitts, and Pavel Jungwirth. Adsorption of atmospherically relevant gases at the air/water interface: Free energy profiles of aqueous solvation of n2, o2, o3, oh, h2o, ho2, and h2o2. *J. Phys. Chem. A*, 108:11573–11579, Jan 2004.

[119] J R Vandenheede, A I Ahmed, and R E Feeney. Structure and role of carbohydrate in freezing point-depressing glycoproteins from an antarctic fish. *J Biol Chem*, 247(24):7885–9, Dec 1972.

[120] A Varki. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, 3(2):97–130, Apr 1993.

[121] C Vega, J L F Abascal, M M Conde, and J L Aragones. What ice can teach us about water interactions: a critical comparison of the performance of different water models. *Faraday Discuss*, 141:251–76; discussion 309–46, Jan 2009.

[122] Kasinadar Veluraja and Claudio J Margulis. Conformational dynamics of sialyl lewisx in aqueous solution and its interaction with selectine. a study by molecular dynamics. *J Biomol Struct Dyn*, 23(1):101–11, Aug 2005.

[123] TG Wang, Y Gotoh, MH Jennings, CA Rhoads, and TY Aw. Lipid hydroperoxide-induced apoptosis in human colonic caco-2 cells is associated with an early loss of cellular redox balance. *Faseb J*, 14(11):1567–1576, Jan 2000.

[124] YX Wang, DI Freedberg, S Grzesiek, DA Torchia, PT Wingfield, JD Kaufman, SJ Stahl, CH Chang, and CN Hodge. Mapping hydration water molecules in the hiv-1 protease/dmp323 complex in solution by nmr spectroscopy. *Biochemistry*, 35(39):12694–12704, Jan 1996.

[125] YX Wang, DI Freedberg, PT Wingfield, SJ Stahl, JD Kaufman, Y Kiso, TN Bhat, JW Erickson, and DA Torchia. Bound water molecules at the interface between the hiv-1 protease and a potent inhibitor, kni-272, determined by nmr. *J Am Chem Soc*, 118(49):12287–12290, Jan 1996.

[126] PW Wilson and JP Leader. Stabilization of supercooled fluids by thermal hysteresis proteins. *Biophys J*, 68(5):2098–2107, Jan 1995.

[127] RJ Woods, RA Dwek, CJ Edge, and B Fraserreid. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides .1. glycam-93 parameter development. *J Phys Chem-Us*, 99(11):3832–3846, Jan 1995.

[128] Junchao Xia, Ryan P Daly, Feng-Chuan Chuang, Laura Parker, Jan H Jensen, and Claudio J Margulis. Sugar folding: A novel structural prediction tool for oligosaccharides and polysaccharides. *J Chem Theory Comput*, 3(4):1629–1643, Jan 2007.

[129] Junchao Xia, Ryan P Daly, Feng-Chuan Chuang, Laura Parker, Jan H Jensen, and Claudio J Margulis. Sugar folding: A novel structural prediction tool for oligosaccharides and polysaccharides. *J Chem Theory Comput*, 3(4):1620–1628, Jan 2007.

[130] Junchao Xia and Claudio Margulis. A tool for the prediction of structures of complex sugars. *J Biomol NMR*, 42(4):241–56, Dec 2008.

[131] Y Yeh and RE Feeney. Antifreeze proteins: Structures and mechanisms of function. *Chem Rev*, 96(2):601–617, Jan 1996.

[132] T YUI, K Ogawa, and A Sarko. Packing analysis of carbohydrates and polysaccharides .18. molecular and crystal-structure of konjac glucomannan in the mannan-ii polymorphic form. *Carbohyd Res*, 229(1):41–55, Jan 1992.

[133] RH Zhou, XH Huang, CJ Margulis, and BJ Berne. Hydrophobic collapse in multidomain protein folding. *Science*, 305(5690):1605–1609, Jan 2004.

[134] Y Zhou, KH Kok, ACS Chun, CM Wong, HW Wu, MCM Lin, PCW Fung, HF Kung, and DY Jin. Mouse peroxiredoxin v is a thioredoxin peroxidase that inhibits p53-induced apoptosis. *Biochem Bioph Res Co*, 268(3):921–927, Jan 2000.