

MOLECULAR CLUSTERING CHARACTERISTICS IN TERNARY  
TREHALOSE AND CHOLINE DIHYDROGEN PHOSPHATE SOLUTIONS

by

Nikita Nikulsin

A thesis submitted to the faculty of  
The University of North Carolina at Charlotte  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Applied Physics

Charlotte

2018

Approved by:

---

Dr. Donald Jacobs

---

Dr. Gloria Elliott

---

Dr. Susan Trammell

© 2018

Nikita Nikulsin

Made available under Creative Commons Attribution-ShareAlike 4.0 License.

## ABSTRACT

NIKITA NIKULSIN. Molecular clustering characteristics in ternary trehalose and choline dihydrogen phosphate solutions. (Under the direction of DR. DONALD JACOBS)

Spatial and temporal characteristics of molecular structure in ternary solutions of trehalose and choline dihydrogen phosphate (CDHP) are studied using molecular dynamics simulations at 300 K for a range of solute concentrations with a 2:1 stoichiometric ratio of trehalose to CDHP. For a given molecular configuration, water molecules are classified as interior (only neighboring other waters) or interfacial (at least one solute neighbor). As a tagged water molecule diffuses, it dynamically exchanges between interior and interfacial type as its local environment changes, with differences in hydrogen bond strength between different molecular species creating a persistent preference for interfacial water. At high solute concentrations, interfacial and interior water have similar diffusivity, which allows for water to collectively act as a plasticizer. The percolation threshold for water was found to be between 81.5% and 83%, which is slightly under the liquid-glass transition, estimated to be near 84.5% solute concentration based on the onset of a volume hysteresis effect. This region of concentrations was further studied using Markov matrices compiled from the transition probabilities for a water molecule to move from a cluster of size  $n$  in one frame to a cluster of size  $m$  in the next frame. The probability distribution functions of the magnitudes of the eigenvalues of these matrices showed a clear signature of the dynamics of the system slowing down starting at 81.5% solute concentration. In general, the structure of the systems were observed to be highly inhomogeneous, with interlaced percolating networks of water and solute coexisting at intermediate concentrations. The density of interior water was found to decrease with increasing solute concentration, creating low-density regions within the matrix.

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Donald Jacobs, for all his help and guidance over the course of this project. I would also like to thank Dr. Azhagiya Singam for guiding me through the initial stages of this project and preparing some of the figures used in this thesis, Dr. Gloria Elliott for collaborating with us and for being on my thesis committee, and Dr. Susan Trammell for being on my committee. Finally, I thank Dr. Shima Ziaei for collaborating with us, Jenny Farmer for creating the probability density estimator program and helping me use it, and Austin Bloom for helping me run molecular dynamics simulations.

## TABLE OF CONTENTS

LIST OF FIGURES	vii
LIST OF TABLES	ix
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: MOLECULAR DYNAMICS AND METHODS	4
2.1. Molecular dynamics	4
2.1.1. Basic principles and the OPLS force field	4
2.1.2. The TIP4P water model	7
2.1.3. Temperature coupling	8
2.1.4. Pressure coupling	9
2.2. Analysis methods	11
2.2.1. Radial distribution functions	11
2.2.2. Diffusion coefficients	13
2.2.3. Reduced second moment	15
2.2.4. Free volume theory	15
CHAPTER 3: SIMULATION PROCEDURE	17
CHAPTER 4: NEAREST NEIGHBORS AND CLUSTERING	20
4.1. Radial distribution functions and nearest neighbors	20
4.2. Interior water definition and properties	23
4.3. Molecular clustering analysis	28
CHAPTER 5: THE GLASS TRANSITION	31
5.1. Volume hysteresis of a glass	31

	vi
5.2. An analogy to free volume theory	32
CHAPTER 6: AN ANALYSIS OF THE TRANSITION REGION USING MARKOV CHAINS	34
6.1. Markov chain model for the dynamics of cluster size statistics	34
6.1.1. Overview	34
6.1.2. Calculation of the transition probabilities	36
6.2. Eigenvalues of the transition matrices	37
6.2.1. The eigenvalue spectrum of a stochastic matrix	37
6.2.2. Analysis of the spectra	38
CHAPTER 7: CONCLUSIONS	41
REFERENCES	43

## LIST OF FIGURES

FIGURE 2.1: Illustration of (a) $r_i$ , (b) $\theta_i$ and (c) $\phi_i$ .	6
FIGURE 2.2: The TIP4P water molecule.	8
FIGURE 4.1: Radial distribution functions for the (a) trehalose-trehalose (b) trehalose-water (c) water-water (d) water-choline (e) water-phosphate and (f) choline-phosphate pairs.	21
FIGURE 4.2: Close-up molecular views of the solution at four different concentrations of solute: (a) 50%, (b) 77%, (c) 83% and (d) 95%. Trehalose molecules and choline cations are shown as thick sticks, waters as thin sticks and phosphate anions as spheres. Choline cations are easily identifiable by the nitrogen atom, which was colored blue in these figures.	22
FIGURE 4.3: Probability distribution functions $P(f)$ of fractions of time $f$ that a particular water molecule spends being interior.	24
FIGURE 4.4: Probability distribution functions $P(N)$ of the numbers of times $N$ that a particular water molecule switches its category from interior to interfacial and vice versa.	24
FIGURE 4.5: The coordination number for interior and interfacial water, i.e. the number of nearest neighbors of an interior or interfacial water averaged over all interior or interfacial waters in all frames. The error bars represent standard deviation.	26
FIGURE 4.6: The diffusivities of interior and interfacial water, calculated by tracking a particular interior or interfacial water molecule while it remains interior or interfacial, respectively, is shown in (a). For reference, the diffusivities of the various components in the system are shown in (b).	27
FIGURE 4.7: The probability $P(n)$ for a water molecule to be a member of a cluster of size $n$ , calculated by dividing the number of water molecules in clusters of a particular size by the total number of water molecules in the system. Concentrations of solute: (a) 77%, (b) 81.5% and (c) 83%. Each inset shows a close-up of the interesting part of the corresponding graph.	30
FIGURE 5.1: Volumes before and after each round of NPT equilibration.	31

- FIGURE 6.1: Snapshots of the evolution of an initial condition consisting entirely of solo waters after (a) 1 frame, (b) 5 frames, (c) 10 frames and (d) 20 frames. 35
- FIGURE 6.2: Plots of the eigenvalues on the complex plain for (a) the 80% solute concentration system and (b) the 89% solute concentration system. 38
- FIGURE 6.3: The maximum time constant (a), and the average time constant (b), both plotted as functions of solute concentration. 39
- FIGURE 6.4: Probability distribution functions  $P(|\lambda|)$  of the magnitudes of the eigenvalues  $|\lambda|$  from various concentration systems. 39

## LIST OF TABLES

TABLE 3.1: Numbers of molecules and ions in the simulated systems with CDHP.	18
--	----

## CHAPTER 1: INTRODUCTION

Anhydrous preservation of biologics may be a less expensive alternative to cryopreservation. This approach consists of adding a glass-forming preservative and removing water from a biologic in solution. Sugar solutions have been used to prevent proteins from being denatured [1, 2]. Trehalose, like other sugars, is capable of stabilizing lipids and proteins by forming a glassy structure around them [3].

As a general empirical trend, the glass transition temperature,  $T_g$ , increases as the water content in the mixture decreases. At low enough moisture contents, room temperature storage of biologics can be enabled. Understanding the effect of water content is an important practical concern in applications to protein formulation because the water content is largely affected by the humidity of the surrounding environment. Salts can be added to the sugar-water mixture as a buffer or to increase  $T_g$  [4, 5, 6]. However, a high  $T_g$  formulation is not the sole design criteria for anhydrous preservation. In some cases, it is favorable to decrease the  $T_g$  in exchange for a lower fragility index and suppression of fast dynamics [7]. In the context of a multi-component mixture, different combinations of various sugars and salts mixed in different proportions with different water contents can be envisioned. Ternary trehalose-salt solutions are notoriously difficult to model using just experimental data [5]. Molecular dynamics (MD) simulations allow one to examine the microscopic-level behavior of the simulated system, providing insight into experimental trends.

The mixture of trehalose with choline dihydrogen phosphate (CDHP), a biocompatible organic salt, was previously investigated by Weng et al [4] for preservation applications. However, that study did not consider the effect that water may have on such a mixture. Here water was introduced as a third component. Due to the

complexity of the resulting ternary mixture, the main focus will be on the behavior of the water and its interactions with the other components in the system. As discussed in previous work [4], the relative strength of inter-component interactions versus self-associations plays an important role in the  $T_g$  behavior of a mixture, as evidenced by the Kwei equation. In this work, information on hydrogen bond lifetimes obtained from the literature is referenced and several qualitative trends observed in the ternary system that are consistent with prior studies are discussed.

As reported in previous work [4], the glass transition temperature of a dry mixture with a 2:1 stoichiometric ratio of trehalose to CDHP is much higher than 300 K, thus at 100% solute concentration the system will be a glass, and it is the plasticizing effect of water that will make it a liquid at lower solute concentrations. As such, it is interesting to see not only how a water molecule interacts with its local environment, but also the emergent global behavior of all the water in the system. The percolation behavior of water is one such global property that has only been briefly touched upon in previous literature [8]. Here, more attention is devoted to the percolation of water in general and various subcategories of water, such as interior and interfacial water. The percolation of solute and trehalose-only clusters is also briefly considered. The existence of percolating networks may have implications for sorption/desorption characteristics of multi-component glasses.

In particular, it may be interesting to use cluster size distributions to quantify the inhomogeneity of the molecular system as water content is varied. The information on clustering properties that is extracted from the simulations is relevant on protein length scales. It has been suggested that protein side chains could position into pockets of void space, which would allow for high levels of local mobility, and this could in turn potentially cause faster rates of degradation at these positions [9, 7]. Even variations in water density may be sufficient to create unwanted gradients in strain. Therefore, the main objective is to investigate the clustering properties of

water in detail, especially because water normally acts as a plasticizer in organic salt solutions.

As water content increases, one can expect the glass transition temperature to decrease below room temperature. Thus, as water content increases, there will be a point for which the system will transition from a glassy state to a liquid state. The glass-liquid transition will be investigated at a fixed temperature of 300 K, while increasing water content, and subsequently performing cluster analysis along the lines of percolation theory. Here, both the spatial and temporal properties of the water clusters will be investigated, and the interactions of various molecular species with each other will be characterized locally through radial distribution functions. Finally, it is important to remember that in an experimental setting, the protocol used to prepare a system in the glass state can have a significant impact on the properties of the system.

Here, a systematic MD study is conducted on trehalose-CDHP solutions holding the stoichiometric ratio of trehalose to CDHP fixed at 2:1 as water content is varied. The 2:1 trehalose to CDHP ratio provides a formulation that has markedly good characteristics for allowing high water content without rapid crystallization [10, 4]. All MD simulations were carried out at 300 K with 30 ns production runs. While 30 ns is incredibly short by many orders of magnitude compared to the timescales that distinguish the difference between liquid and a molecular glass [11], the percolation features for various constituents already show distinctive characteristics as water content varies.

## CHAPTER 2: MOLECULAR DYNAMICS AND METHODS

### 2.1 Molecular dynamics

A molecular dynamics simulation consists of numerically solving Newton's equations for all of the atoms in the simulated system. Usually, the initial coordinates of the atoms are provided in an initial configuration file, and the initial velocities are either randomly generated from a Maxwell-Boltzmann distribution corresponding to the temperature at which the simulation is started, or can be extracted from a previous simulation of the same system. The forces on each atom at each time step are calculated from the positions of all of the other atoms by taking the gradient of the potential energy function, more often called the force field in the context of molecular modeling [12].

#### 2.1.1 Basic principles and the OPLS force field

The Optimized Potential for Liquid Simulations all atom force field (OPLS-AA) was used for the simulations performed in the course of this study. All-atom force fields consider hydrogen as an individual atom, in contrast to united-atom force fields, which treat carbon atoms and any bound hydrogens as one atom for increased computation speed. The OPLS-AA force field has the following functional form [13]:

$$E(r^N) = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{torsion}} + E_{\text{nonbond}}, \quad (1)$$

where

$$E_{\text{bonds}} = \sum_i k_{r,i} (r_i - r_{0,i})^2, \quad (2)$$

$$E_{\text{angles}} = \sum_i k_{\theta,i} (\theta_i - \theta_{0,i})^2, \quad (3)$$

$$E_{\text{torsion}} = \sum_i \left[ \frac{V_{1,i}}{2}(1 + \cos \phi_i) + \frac{V_{2,i}}{2}(1 + \cos 2\phi_i) + \frac{V_{3,i}}{2}(1 + \cos 3\phi_i) + \frac{V_{4,i}}{2}(1 + \cos 4\phi_i) \right], \quad (4)$$

$$E_{\text{nonbond}} = \sum_i \sum_{j>i} f_{ij} \left( \frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right). \quad (5)$$

Here,  $r^N$  is the set of coordinates of all  $N$  atoms in the system,  $r_i$  is the distance between two bound atoms,  $\theta_i$  is the angle formed by three atoms bound together and  $\phi_i$  is the dihedral angle formed by four atoms bound together (Figure 2.1). Strictly speaking, the coordinates of atoms which comprise water molecules should not be included in the set  $r^N$  (see next subsection). The constants  $k_{r,i}$ ,  $k_{\theta,i}$  and the  $V_i$ 's depend on the types of atoms participating in the bonds. Also,  $r_{0,i}$  and  $\theta_{0,i}$  are the equilibrium bond length and angle, respectively. The sums in equations (2), (3) and (4) are taken over all bonds, angles and dihedrals, respectively. In equation (5), the first term is the standard Coulomb interaction, with  $q_i$  and  $q_j$  being the partial charges on atoms  $i$  and  $j$ , whereas  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the Lennard-Jones parameters. The Lennard-Jones parameters need only be defined between similar atom types, those between different atom types can be calculated from combination rules:  $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$  and  $\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}}$ . The sum in equation (5) is taken over all pairs of atoms, and  $f_{ij} = 0$  if atoms  $i$  and  $j$  are separated by one or two bonds,  $f_{ij} = 0.5$  if the atoms are separated by three bonds and  $f_{ij} = 1$  if they are separated by four or more bonds or if they are in different molecules [13].

The force field parameters can be determined from experimental data or quantum mechanical calculations. For example, the equilibrium bond length and angles can be determined from crystallographic data. An approach employed in determining the OPLS force field parameters consisted of making plausible guesses and then carrying out simulations on several liquid systems to see if thermodynamic properties of the liquid, such as density and heat of vaporization, derived from simulation matched experimental data. The parameters were then adjusted for better agreement with

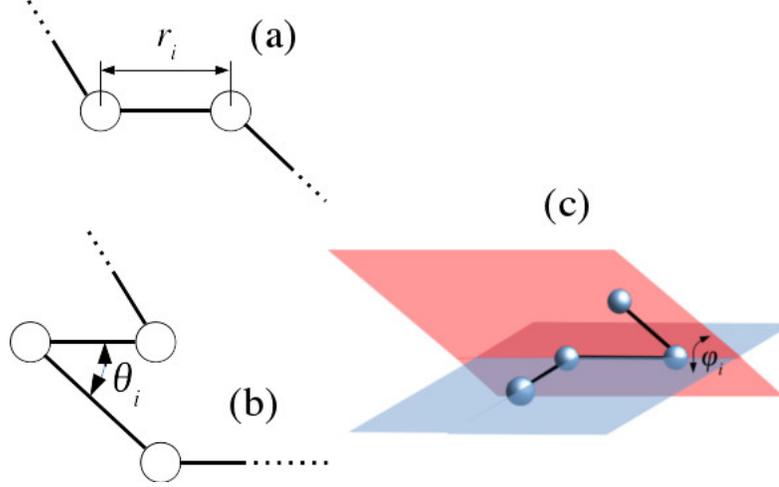


Figure 2.1: Illustration of (a)  $r_i$ , (b)  $\theta_i$  and (c)  $\phi_i$ .

experiment. This procedure was repeated until the thermodynamic properties of the simulated liquid accurately reproduced the experimental values [14].

Once the parameters are known, the motion of atom  $n$  can be calculated from Newton's second law:

$$m \frac{d^2 \vec{r}_n}{dt^2} = \nabla_n E(r^N). \quad (6)$$

These equations were numerically integrated using the leap-frog algorithm. For a sufficiently small time step  $\Delta t$ , the position and velocity of atom  $n$  can be approximated as follows:

$$\begin{aligned} \vec{v}_n \left( t + \frac{\Delta t}{2} \right) &= \vec{v}_n \left( t - \frac{\Delta t}{2} \right) + \frac{\Delta t}{m} \nabla_n E(r^N(t)), \\ \vec{r}_n(t + \Delta t) &= \vec{r}_n(t) + \Delta t \vec{v}_n \left( t + \frac{\Delta t}{2} \right). \end{aligned} \quad (7)$$

If the initial conditions at some time  $t_0$  are known, Euler's method should first be used to calculate  $\vec{v}_n(t_0 + \Delta t/2)$  for all atoms, after which the leap-frog algorithm can be started. The positions and velocities of all of the atoms are saved to a trajectory file in specified intervals of time steps [12].

Finally, it should be noted that most commonly in molecular dynamics, periodic boundary conditions are imposed on the simulated systems. This means that the

particles are contained in a simulation box of a specified finite volume, which is surrounded by translated copies of itself. Thus, if an atom passes through a box boundary, it will reappear on the opposite side of the simulation box. This also means that the shortest distance between two atoms might not be the one that is given by the distance formula. When calculating short-range (non-electrostatic) non-bonded interactions, the minimum image convention is employed, i.e. contributions only from the nearest image of each atom (or the actual atom, if it is closer to the reference atom than any of its images) are considered. For electrostatic interactions, ignoring farther images will lead to significant error, so the Particle Mesh Ewald (PME) method is used to sum over infinitely many images in Fourier space [12].

### 2.1.2 The TIP4P water model

Water parameters are usually not included in force fields, and OPLS-AA is no exception. In order to simulate the effects of water, a special force field, known as a water model, should be used. For the purposes of this study, the Transferable Intermolecular Potential 4-point (TIP4P) water model was employed. In this model, the water molecule is a rigid structure, so the potential energy function consists of only non-bonded terms [15]:

$$E(r^{Nw}) = \sum_m \sum_{n>m} E_{mn},$$

$$E_{mn} = \sum_i^{\text{mol } m} \sum_j^{\text{mol } n} \frac{q_i q_j e^2}{r_{ij}} + \frac{A}{r_{O_m O_n}^{12}} - \frac{C}{r_{O_m O_n}^6}, \quad (8)$$

where  $r^{Nw}$  is the set of coordinates of all oxygens in the water molecules along with the angles specifying the spatial orientation of the molecules, the variables in the Coulomb term are defined similarly to those in equation (5), and  $A$  and  $B$  are the Lennard-Jones parameters, with  $r_{O_m O_n}$  being the distance between the oxygen of molecule  $m$  and that of molecule  $n$ . The indices  $i$  and  $j$  enumerate the atoms within

molecules  $m$  and  $n$ , respectively [15].

Since the water molecule is considered to be rigid, interactions are only considered between atoms in distinct molecules. Interactions between water and solute are accounted for by including the corresponding terms in the non-bonded interaction potential (5) and using combination rules to get the appropriate Lennard-Jones parameters. Lennard-Jones interactions involving water hydrogens are not considered.

In the TIP4P water molecule, the negative partial charge is moved off of the oxygen atom to a fourth interaction site (hence the name), which is in the direction of the hydrogens along the bisector of the HOH angle (Figure 2.2) [15].

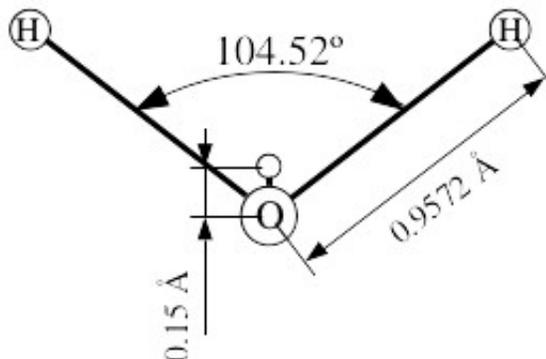


Figure 2.2: The TIP4P water molecule.

### 2.1.3 Temperature coupling

The usage of molecular dynamics as described above will produce a microcanonical ensemble (constant number of particles, constant volume, constant energy). To simulate more realistic conditions, the system should be coupled to an external thermal reservoir. A simple way to achieve this is to use a Berendsen thermostat [12].

A Berendsen thermostat aims to gradually make the temperature of the system  $T$  match the temperature of the reservoir  $T_0$  by imposing the following condition [16]:

$$\frac{dT}{dt} = \frac{1}{\tau}(T_0 - T), \quad (9)$$

where  $\tau$  is a coupling constant. This condition is enforced by scaling the velocities of

all atoms by a factor  $\lambda$ . When time is discretized, equation (9) becomes:

$$\Delta T = \frac{\Delta t}{\tau}(T_0 - T). \quad (10)$$

Since  $\frac{N_f}{2}NkT = \sum_i \frac{1}{2}m_i v_i^2$ , where  $N_f$  is the number of degrees of freedom and  $k$  is Boltzmann's constant,

$$\Delta T = \frac{1}{N_f N k} \sum_i m_i (\lambda^2 v_i^2 - v_i^2) = \frac{\lambda^2 - 1}{N_f N k} \sum_i m_i v_i^2 = (\lambda^2 - 1)T. \quad (11)$$

Substituting equation (11) into (10) and solving for  $\lambda$ , one gets [16]:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau} \left( \frac{T_0}{T} - 1 \right)}. \quad (12)$$

A problem with the Berendsen thermostat is that it does not correctly reproduce the canonical ensemble. This can be corrected by adding a stochastic term to equation (9):

$$dT = (T_0 - T) \frac{dt}{\tau} + 2 \sqrt{\frac{TT_0}{N_f \tau}} dW, \quad (13)$$

where  $dW$  is a Wiener process. This modification results in the velocity-rescaling thermostat. When the temperature of the system is far from the temperature of the reservoir, the first term in equation (13) will dominate, causing the thermostat to behave as a Berendsen thermostat, with the  $T$  quickly approaching the  $T_0$ , whereas when the system temperature is close to the reservoir temperature,  $T$  is allowed to fluctuate, more accurately sampling the canonical ensemble [17].

#### 2.1.4 Pressure coupling

In order to simulate a system under constant temperature and pressure, a pressure coupling is also required. A Berendsen barostat provides a simple way to do this.

This time, a condition similar to (9) is imposed on the pressure [16]:

$$\frac{dP}{dt} = \frac{1}{\tau_P}(P_0 - P), \quad (14)$$

where  $\tau_P$  is a coupling constant for pressure. This condition is enforced by scaling the box side lengths and the coordinates of all atoms by a factor  $\mu$ . First, a term corresponding to the change in coordinates due to scaling is added to the velocity:  $\dot{x} = v + \alpha x$ . The volume also changes due to the scaling of the box side lengths:  $\dot{V} = 3\alpha V$ . Using the relation:

$$\frac{dP}{dt} = \frac{-1}{\beta V} \frac{dV}{dt} = \frac{-3\alpha}{\beta}, \quad (15)$$

where  $\beta$  is the isothermal compressibility, along with equation (14), one can obtain the following expression for  $\alpha$ :

$$\alpha = \frac{-\beta}{3\tau_P}(P_0 - P). \quad (16)$$

Using the fact that  $\Delta V = \dot{V}\Delta t = 3\alpha V\Delta t$  and that  $\Delta V = \mu^3 V - V$ , one can solve for  $\mu$  [16]:

$$\mu = \sqrt[3]{1 - \frac{\beta\Delta t}{\tau_P}(P_0 - P)}. \quad (17)$$

Similarly to the thermostat, the Berendsen barostat, even when used with the velocity rescaling thermostat, does not correctly reproduce the isothermal-isobaric ensemble. However, it will suffice if it is only used for system equilibration and then turned off during a data collection run, as was done in this study [12].

## 2.2 Analysis methods

### 2.2.1 Radial distribution functions

Let  $r^N$  be the set of all coordinates of all particles in a conservative system,  $p^N$  be the set of all momenta, and  $f(r^N, p^N)$  be the probability density of finding the particles with coordinates  $r^N$  and momenta  $p^N$ . Then the probability density will be the ratio of the corresponding Boltzmann factor to the partition function:

$$f(r^N, p^N) = \frac{\exp[-\beta H(r^N, p^N)]}{\int \int \exp[-\beta H(r^N, p^N)] dr^N dp^N}, \quad (18)$$

where  $\beta = 1/(kT)$  and  $H(r^N, p^N)$  is the Hamiltonian of the system. Since the Hamiltonian is the sum of the kinetic energy, which only depends on momenta, and the potential energy, which only depends on coordinates, the phase space distribution function factors into the configuration space distribution and the momentum space distribution:  $f(r^N, p^N) = P(r^N)\Phi(p^N)$  [18].

The probability density of finding particle 1 at  $\vec{r}_1$  and particle 2 at  $\vec{r}_2$  can be found by integrating out the coordinates of all other particles:

$$P^{(2/N)}(\vec{r}_1, \vec{r}_2) = \int \dots \int P(r^N) d^3\vec{r}_3 \dots d^3\vec{r}_N, \quad (19)$$

where the  $(2/N)$  superscript refers to the fact that two particles are being tracked in a system of  $N$  particles. Since it does not make much sense to care about a *specific* particle being at some location, it would be more useful to define the distribution function for finding *any* particle at  $\vec{r}_1$  and any other particle at  $\vec{r}_2$ . There are  $N$  ways to select the first particle and  $N - 1$  ways to select the second, thus:

$$\rho^{(2/N)}(\vec{r}_1, \vec{r}_2) = N(N - 1)P^{(2/N)}(\vec{r}_1, \vec{r}_2). \quad (20)$$

One could also define the distribution function for finding any one particle at point

$\vec{r}_1$ :

$$\rho^{(1/N)}(\vec{r}_1) = N \int \dots \int P(r^N) d^3\vec{r}_2 \dots d^3\vec{r}_N. \quad (21)$$

For an isotropic system, this should just be the number density of particles in the system:  $\rho^{(1/N)}(\vec{r}_1) = \rho = N/V$ . If the positions of different particles are uncorrelated, then one can write  $\rho^{(2/N)}(\vec{r}_1, \vec{r}_2) = N(N-1)/V^2 = N^2(1 - \frac{1}{N})/V^2 \approx \rho^2$ . It would then make sense to define a fractional deviation of the true distribution function from the uncorrelated one:

$$g(\vec{r}_1, \vec{r}_2) = \rho^{(2/N)}(\vec{r}_1, \vec{r}_2)/\rho^2. \quad (22)$$

For an isotropic system, this function will only depend on  $|\vec{r}_1 - \vec{r}_2| = r$ :  $g(\vec{r}_1, \vec{r}_2) = g(r)$ . The function  $g(r)$  is called the radial distribution function [18].

The conditional distribution function for finding a particle at  $\vec{r}$  given that another particle is at the origin is  $\rho^{(2/N)}(0, \vec{r})/\rho$ . From equation (22) it follows that  $\rho g(r) = \rho^{(2/N)}(0, \vec{r})/\rho$ , which provides an interpretation for the radial distribution function. In other words,  $\rho g(r)$  is the average number density of particles at a distance  $r$  from some reference particle [18].

In this study, radial distribution functions for particles of type A and B were calculated using the following formula:

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle_A}{\langle \rho_B \rangle_{R,A}}, \quad (23)$$

where  $\rho_B(r)$  is the number density of type B particles at distance  $r$  around an A particle. It is averaged over all A particles in the numerator and over spheres of radius  $R$  around all A particles in the denominator. In all calculations performed,  $R$  was half of the simulation box length [12].

### 2.2.2 Diffusion coefficients

If a concentration gradient exists in a mixture, it is expected that a diffusion flux will appear, with particles moving from areas of higher concentration to those of lower concentration, driving the system towards equilibrium. Fick's first law gives the relationship between diffusion flux and concentration gradient [18]:

$$\vec{j}(\vec{r}, t) = -D\nabla n(\vec{r}, t), \quad (24)$$

where  $\vec{j}(\vec{r}, t)$  is the diffusion flux through  $\vec{r}$  at time  $t$ ,  $D$  is a proportionality constant called the diffusion coefficient and  $n(\vec{r}, t)$  is the concentration at point  $\vec{r}$  and time  $t$ . If there are no chemical reactions in the system, the number of particles is a conserved quantity, and so the concentration will satisfy the continuity equation:

$$\frac{\partial n(\vec{r}, t)}{\partial t} = -\nabla \cdot \vec{j}(\vec{r}, t). \quad (25)$$

Combining the continuity equation (25) with Fick's first law (24) produces Fick's second law:

$$\frac{\partial n(\vec{r}, t)}{\partial t} = D\nabla^2 n(\vec{r}, t). \quad (26)$$

The probability density of finding a particle at point  $\vec{r}$  and time  $t$  is proportional to the concentration of particles of that type at the same point and time, thus:

$$\frac{\partial P(\vec{r}, t)}{\partial t} = D\nabla^2 P(\vec{r}, t), \quad (27)$$

where  $P(\vec{r}, t)$  is the probability density. Next, define the mean squared displacement as  $R^2(t) = \langle |\vec{r}(t) - \vec{r}(0)|^2 \rangle$ , where the average is taken over all diffusing particles. Further, suppose that at  $t = 0$  all particles are at the origin, i.e.  $\vec{r}(0) = 0$  for all particles, then  $R^2(t) = \langle \vec{r}^2(t) \rangle$ . This means that  $R^2(t) = \int \vec{r}^2 P(\vec{r}, t) d^3\vec{r}$ . Taking

the time derivative of the mean squared displacement, using equation (27), and then integrating by parts twice and using the fact that  $P(\vec{r}, t)$  is normalized to one, one obtains:

$$\frac{d}{dt}R^2(t) = \int \vec{r}^2 \frac{\partial}{\partial t} P(\vec{r}, t) d^3\vec{r} = D \int \vec{r}^2 \nabla^2 P(\vec{r}, t) d^3\vec{r} = 6D \int P(\vec{r}, t) d^3\vec{r} = 6D. \quad (28)$$

Thus, the diffusion coefficient can be calculated from the derivative of the mean squared displacement [18].

It should be noted that at short time scales, before a particle encounters its neighbors, it will move inertially, and not diffusively, so equation (28) will not hold. Thus, the proper formula for the diffusion coefficient should be:

$$D = \frac{1}{6} \lim_{t \rightarrow \infty} \frac{d}{dt} \langle |\vec{r}(t) - \vec{r}(0)|^2 \rangle. \quad (29)$$

In practice, this means that the molecular dynamics simulations need to be sufficiently long, so that the mean squared displacement graph will approach a straight line as a function of time [18]. More generally, non-linear time dependence at shorter times will reflect a variety of inter-particle correlations that can be due to persistence in momentum of the particles on very short time scales or due to structural correlations arising from molecular clustering.

Fick's law is a macroscopic view of diffusion, and as a partial differential equation it is an approximation to a more accurate microscopic view that comes from statistical mechanics. Both views are equivalent under the condition that length scales and time scales are long compared to intermolecular correlation of particles in space and time. Fick's law, or the diffusion equation that results, describes Brownian motion of particles in a medium where a diffusion constant describes the rate of spread of particles through the medium. More generally, it is possible to calculate the mean squared displacement of a sample of tagged particles in a system, where the time

dependence of mean squared displacement reflects all particle correlations. At long time scales, even if it was assumed that at  $t = 0$  all particles are at the origin, i.e. their distribution at  $t = 0$  is the delta function, the particles will diffuse out so that the concentration gradient becomes zero. Then, it can be considered that data collection is started at some time  $t_0 > 0$ , and the self-diffusion constant for a molecular species that is captured from microscopic fluctuations in equilibrium is equivalent to the diffusion coefficient from Fick's law, as shown above. The microscopic form of calculating the diffusion constant from self-diffusion is applied.

### 2.2.3 Reduced second moment

The reduced second moment (RSM, also called the reduced average cluster size) was used to determine the percolation thresholds [19]. The RSM was defined as follows:

$$M'_2 = \frac{1}{G} \sum_{n=1}^{n_{max}} i_n n^2 - \frac{n_{max}^2}{G}, \quad (30)$$

where  $G$  is the total number of molecules under consideration (interior water molecules or water molecules in general),  $i_n$  is the number of clusters of size  $n$ , and  $n_{max}$  is the size of the largest cluster. Below the percolation threshold, there are typically several clusters of size near  $n_{max}$ , so subtracting out one of them does not significantly impact the value of  $M'_2$ . On the other hand, above the percolation threshold, the percolating cluster is the largest cluster and contains most of the molecules. All the other clusters have size much smaller than  $n_{max}$ , so subtracting out the percolating cluster causes a significant decrease in the value of  $M'_2$ . This produces a sharp peak in the graph of  $M'_2$  around the percolation threshold.

### 2.2.4 Free volume theory

From the standpoint of free volume theory for the liquid/glass transition [20, 21, 22], a liquid state will exist when the molecular structure supports a percolating network of free volume among the molecular species within a mixture. Conceptually, this

means a pair of neighboring molecules both have enough local free volume available to exchange locations with each other without crossing over a high free energy barrier, and there is negligible change in free energy after the exchange. This high degree of degeneracy in molecular configurations allows molecules to readily move around each other, which is the signature property of a liquid. Importantly, the exchange requires additional free volume to be available from the nearest neighbors of this pair of molecules to enable the two molecules to pass around each other. This is essentially the experience one has walking through a crowded room. Other persons must move a little to make space for a person to pass through, and once passage is made, the space fills back up.

A molecule is referred to as liquid-like if it has free volume available; otherwise it is solid-like. When a cluster of liquid-like molecules percolates, the system is said to be in a liquid phase. Inevitably, the question of how to identify a molecule as liquid-like or solid-like arises because computing free volumes from free energy functions, as prescribed by Cohen and Grest [21, 22] is not straightforward. In some previous studies [23, 24, 25], the shape of the molecule's Voronoi cell was used as a criterion, with more spherical cells being considered liquid-like and elongated or irregular-shaped cells being solid-like. In this work, it is proposed that the percolation of the plasticizer (water) may be a sufficient condition for the system to exhibit liquid behavior.

The simplistic approach adopted by free volume theory may not accurately reflect the reality of multi-component organic glasses, where hydrogen bonds, steric effects and other factors significantly complicate the situation. Thus, at this stage, free volume theory is invoked under the assumption that water can be treated as the dominant liquid-like component. The percolation of water could then provide a lower bound estimate of the glass transition. If water cannot be treated as a liquid-like component, one would expect to find the percolation threshold of water to be above the glass transition.

## CHAPTER 3: SIMULATION PROCEDURE

Simulations were performed on ternary solutions of trehalose and choline dihydrogen phosphate (CDHP), dissolved in water. Twenty-three different concentrations were simulated. All of the systems involved in this study were simulated using the molecular dynamics software GROMACS [26]. The numbers of molecules and ions in each system (Table 3.1) were chosen so that they could fit into a box with a volume of about  $1000 \text{ nm}^3$ . These numbers were calculated from the experimental number densities for each individual compound under the assumption of ideal mixing. In all simulations, the stoichiometric ratio of trehalose to CDHP was 2:1. All simulations used the TIP4P water model, except the 100% compositions, which had no water. The OPLS-AA force field was used for all simulations. The parameters were obtained from [13, 27, 28], except the partial charges on CDHP atoms, which were generated using the R.E.D. server [29, 30, 31] with GAMESS-US software [32].

The initial atomic coordinates were generated using the Packmol software [33], using crystallographic trehalose [34] and CDHP [35] structures as input. Appropriate numbers of molecules and ions were packed into a cubic box with an edge length of 10 nm. These coordinates were then input into GROMACS for potential energy minimization. Energy minimization was stopped once the maximum force in the system became less than  $500 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-1}$ . When this value could not be reached, the packing was repeated with the same box size, or, if that did not work, the edge length of the box was increased by 1 nm and the molecules and ions were repacked. After energy minimization, the systems were coupled to a velocity rescaling thermostat set to 300 K with a coupling constant of 0.1 ps and a Berendsen barostat set to 1 bar with a coupling constant of 1 ps and an isothermal compressibility of  $4.5 \cdot 10^{-5} \text{ bar}^{-1}$ ,

Table 3.1: Numbers of molecules and ions in the simulated systems with CDHP.

Percent solute	Trehalose molecules	CDHP formula units	Water molecules
0	0	0	33427
10	140	70	30975
20	292	146	28713
30	454	227	26042
40	630	315	23231
50	818	409	20109
53	878	439	19141
56	940	470	18157
59	1002	501	17117
62	1068	534	16092
65	1134	567	15011
68	1202	601	13905
71	1272	636	12772
74	1344	672	11609
77	1416	708	10398
80	1492	746	9170
81.5	1532	766	8549
83	1570	785	7905
84.5	1610	805	7260
86	1674	837	6221
89	1734	867	5269
95	1906	953	2466
100	2058	1029	0

and equilibrated for a minimum of 1 ns. During equilibration, the box volume would decrease in all cases due to the initial packing protocol. Convergence was checked by plotting the simulation box volume as a function of time and performing a linear fit on the last 100 ps. If the absolute value of the regression slope was less than  $0.02 \text{ nm}^3 \cdot \text{ps}^{-1}$ , the systems were considered to have converged to their final volumes. If this was not the case, the constant temperature and pressure equilibration was extended. After a system had converged, the barostat was removed, and the system was equilibrated at constant volume and temperature for an additional 3 ns. After the equilibration simulations were complete, a 30 ns data collection simulation was performed. A time step of 2 fs was used in all simulations. Coordinates and velocities

were saved to the trajectory file every 500 steps during the equilibration runs and every 3000 steps during the data collection run.

For the 80% solute by weight system, as well as all systems with 83% solute by weight or higher, the NPT equilibration run was repeated using a different initial box volume. For the 95% and 100% cases, the second NPT runs were each followed by a second NVT run and a second production run.

## CHAPTER 4: NEAREST NEIGHBORS AND CLUSTERING

### 4.1 Radial distribution functions and nearest neighbors

The radial distribution functions between the various species in the system were calculated with the GROMACS tool `gmx rdf`, which implements formula (23). The "A" and "B" particles correspond to the oxygens in a reference species and oxygens within another species, respectively. The plots of the trehalose-trehalose, trehalose-water, water-water, water-choline, water-phosphate and choline-phosphate RDFs obtained from the 20%, 50%, 80% and 95% CDHP systems are shown in Figure 4.1.

All RDFs in all systems have a first peak followed by a minimum around 0.35 nm. This peak corresponds to interactions between nearest neighbors, often in the form of hydrogen bonds [5]. Also, given the van der Waals radii of the atoms [36], an atom belonging to another molecule cannot be located between two atoms that are 0.35 nm apart or less. Thus, two molecules are defined to be nearest neighbors if a heavy atom (i.e. any atom that is not a hydrogen) belonging to one molecule is within the cutoff distance, 0.35 nm, of a heavy atom belonging to the other molecule.

For the trehalose-trehalose RDF, the broad and tall second peak at low solute concentrations indicates that trehalose molecules cluster together instead of being evenly distributed throughout the system. This is illustrated in Figure 4.2. The clustering peaks are still present at higher solute concentrations, but because there is more trehalose packed within the same volume, the trehalose distribution should be more uniform with the clustering less pronounced. Hence, the peaks are lower. The complex set of peaks after the first peak in the trehalose-trehalose RDF is attributed to the trehalose molecule containing eleven oxygens, each of which are contributing to different peaks. Interestingly, the first peak is sharp and well defined but has a

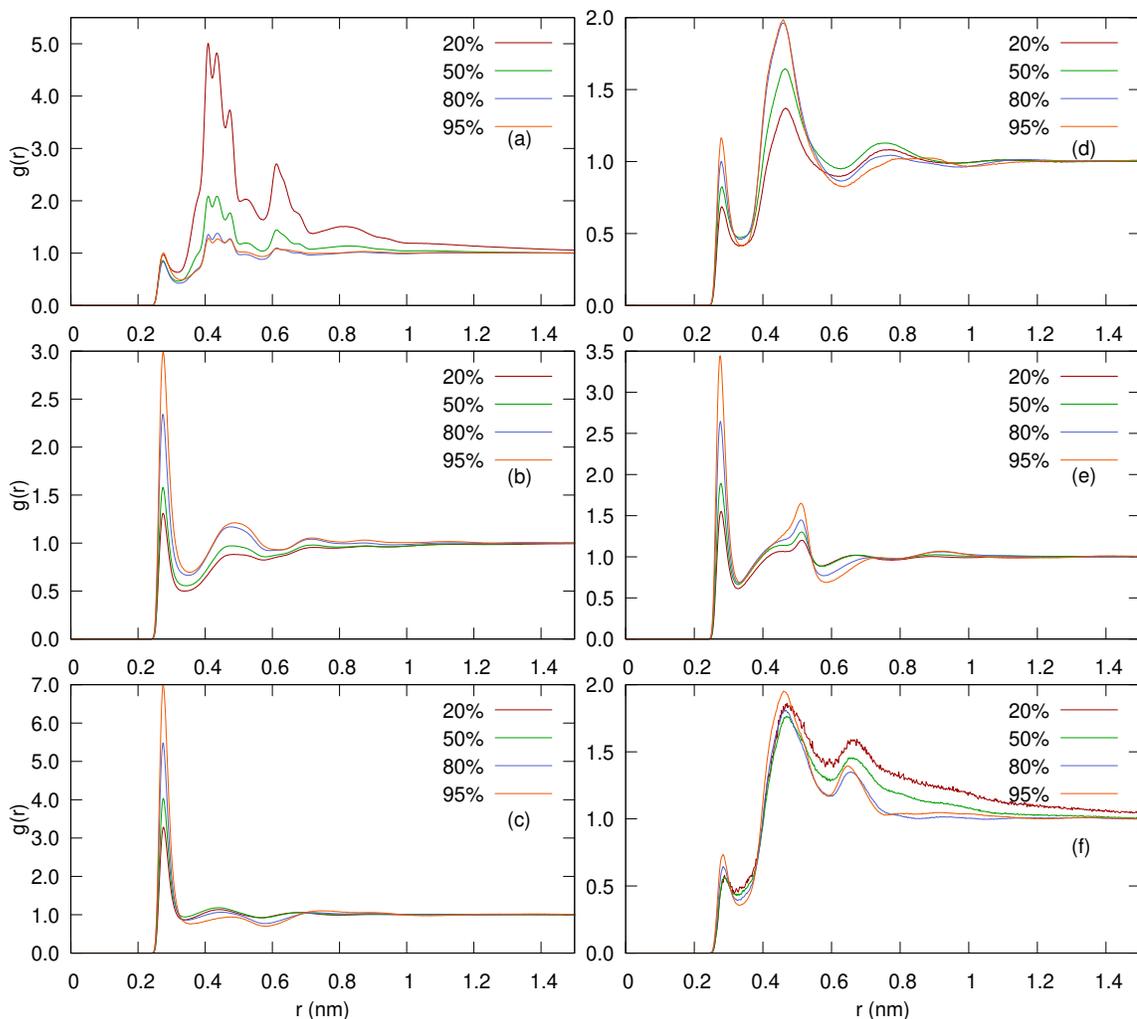


Figure 4.1: Radial distribution functions for the (a) trehalose-trehalose (b) trehalose-water (c) water-water (d) water-choline (e) water-phosphate and (f) choline-phosphate pairs.

very small amplitude. This means that while trehalose tends to cluster on length scales commensurate with its molecular size, there are random orientations between neighboring trehalose molecules. Conversely, the water-water RDF has a strongly pronounced first peak, after which the packing of water molecules quickly assumes bulk behavior. This suggests that water molecules also cluster, however, being a small molecule with only one oxygen, the correlation length scale after which the water assumes bulk behavior is much smaller than that of trehalose. Although a quick transition to bulk water in the RDF is not a sufficient indication that water is a

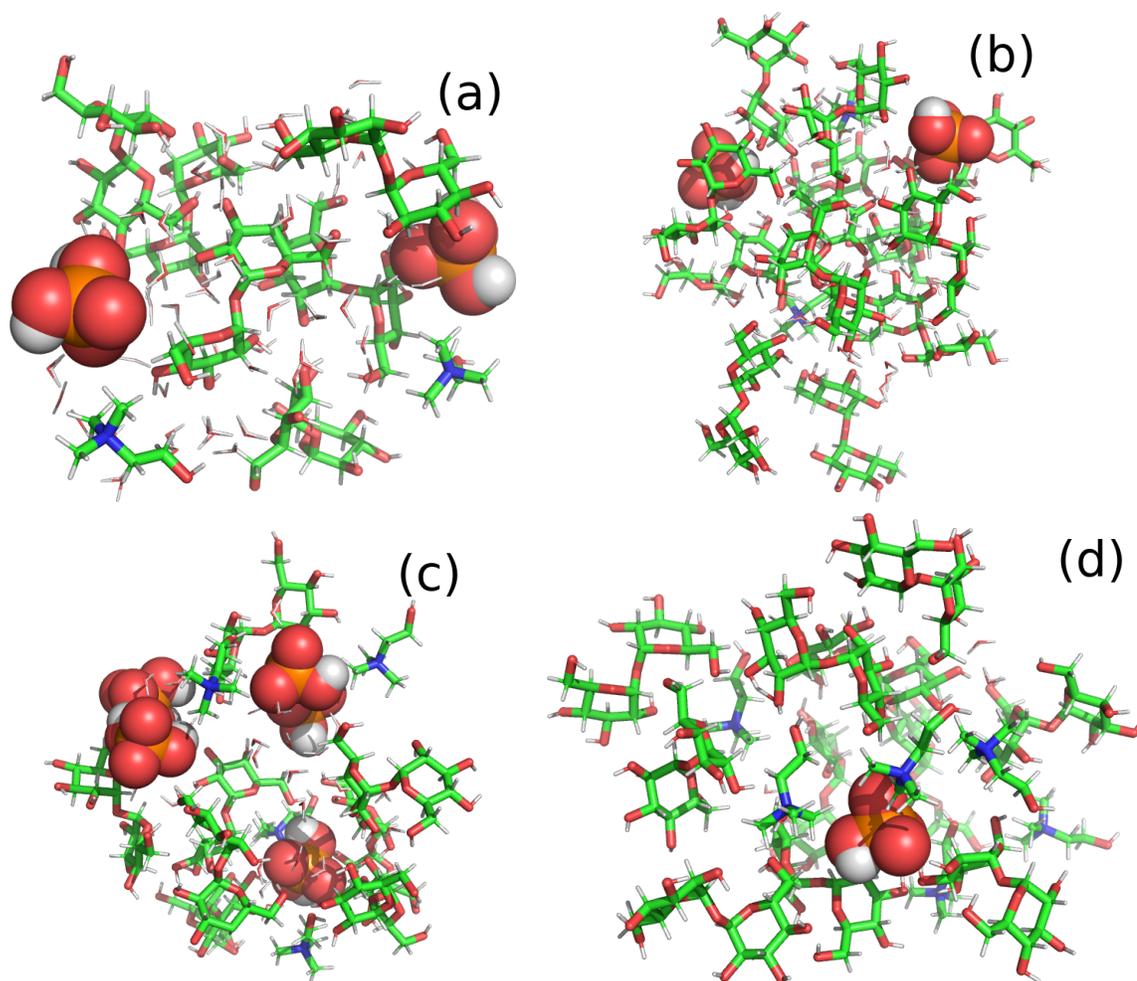


Figure 4.2: Close-up molecular views of the solution at four different concentrations of solute: (a) 50%, (b) 77%, (c) 83% and (d) 95%. Trehalose molecules and choline cations are shown as thick sticks, waters as thin sticks and phosphate anions as spheres. Choline cations are easily identifiable by the nitrogen atom, which was colored blue in these figures.

liquid-like molecule, or that it acts as a plasticizer, this result is consistent with both of these properties. Finally, the second peak in the trehalose-water RDF suggests that water tends to cluster around trehalose more readily than to cluster to itself, even forming a second hydration shell. This effect becomes more pronounced at high concentration of solute, and suggests that water and trehalose will have a high propensity to form an interface at low water content. This is consistent with Figure 4.2, where water does not form large pools, preferring to cluster to solute.

Analysis of the water-choline and water-phosphate RDFs shows significantly lower populations of water associating with choline than with phosphate, as indicated by the peaks in Figure 4.1 d being significantly lower than in Figure 4.1 e. The results also suggest that water mediates hydrogen bonds between choline and phosphate. The peaks around  $r = 0.45$  nm and  $r = 0.7$  nm suggest that water molecules act as a bridge between choline and phosphate ions. This can be seen in Figure 4.2, especially in panels (a) and (c).

## 4.2 Interior water definition and properties

Interior water is defined as the set of all water molecules in a given frame whose only nearest neighbors are other water molecules. Thus, at least in a first order approximation, large clusters of interior water will resemble bulk properties of water. However, narrow channels or small clusters of interior water can form, and these structures do not have sufficient volume to take on bulk properties. Nevertheless, interior waters may have distinctly different characteristics than the water molecules that directly interact with solute. The set of water molecules that are not interior, i.e. water molecules that have at least one solute nearest neighbor, is referred to as interfacial water.

As a tagged water molecule diffuses throughout the system, it is classified as either an interior or interfacial water based on its local environment that is dynamically changing. Any particular interior water molecule can move to be within the cutoff distance of a solute heavy atom, thus becoming interfacial water, just as any interfacial water molecule can become interior water. To quantify how persistent a typical water molecule is in being interior or interfacial, one calculates, for each water molecule, the fraction of time that it spends being interior by dividing the number of frames in which it is interior by the total number of frames. The probability density function of these fractions for all water molecules in the system is then determined for a particular concentration using a robust nonparametric probability density estimation

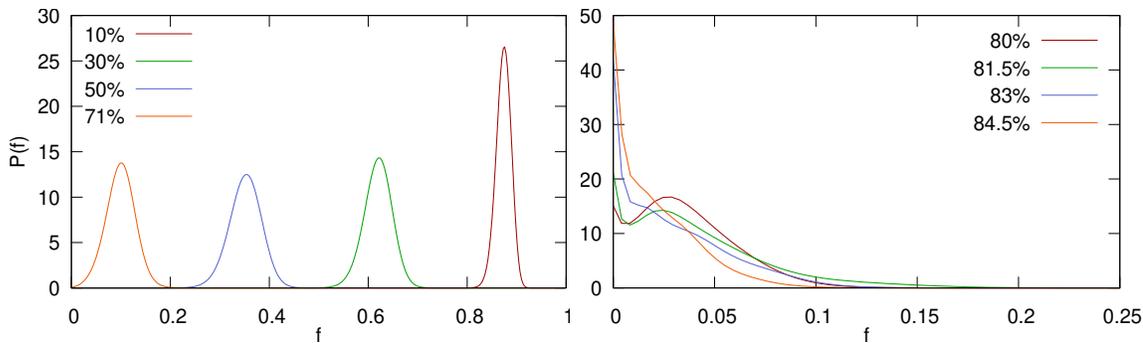


Figure 4.3: Probability distribution functions  $P(f)$  of fractions of time  $f$  that a particular water molecule spends being interior.

algorithm [37]. These distributions are shown in Figure 4.3. As can be observed, the distributions are Gaussian-like for lower solute concentrations, however, as water content decreases, the distributions become sharply peaked at  $f = 0$ , eventually becoming exponential-like.

Also, for each water molecule, the number of times that it switches its category from interior to interfacial and vice versa during the course of a simulation is counted. The probability density functions for these counts are shown in Figure 4.4. Again, for lower solute concentrations, the distributions are Gaussian-like. Due to there being much more water than solute at lower solute concentrations it will, on average, take an interior water molecule longer to encounter a trehalose molecule or ion, so the switching events are less frequent. As the solute concentration increases, encoun-

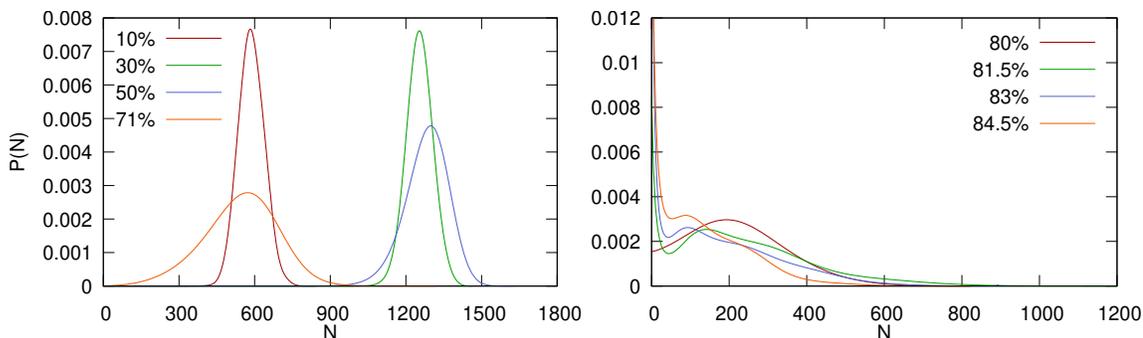


Figure 4.4: Probability distribution functions  $P(N)$  of the numbers of times  $N$  that a particular water molecule switches its category from interior to interfacial and vice versa.

ters between interior water and solute become more frequent, and so the number of switchings increases. At this point, the number of interfacial sites is defined as the maximum number of interfacial water molecules that a given amount of solute can accommodate; it is roughly equivalent to the average water coordination number for a solute heavy atom multiplied by the total number of solute heavy atoms. As solute concentration continues to increase, the ratio of water molecules to interfacial sites decreases, and thus more water molecules can be in the more favorable (as explained in the following paragraph) interfacial state. With the competition for interfacial sites becoming less intense, and with less interior water, the number of switchings decreases. Just as in Figure 4.3, the distributions in Figure 4.4 that correspond to high solute concentrations are sharply peaked at zero, indicating that many water molecules remain interfacial throughout the entire simulation.

This data can be interpreted from the standpoint of hydrogen bond strength. Since in general, water-water hydrogen bonds have a shorter lifetime than water-sugar [38] and water-ion [39] hydrogen bonds, and since hydrogen bond lifetime correlates with strength [5], the water-water hydrogen bonds are weaker. Thus, water prefers the interfacial state. At low solute concentrations, due to the abundance of interior water, whenever water-solute hydrogen bonds are broken, the interfacial water molecule can be readily replaced by an interior water molecule, which only has weak water-water hydrogen bonds. A parallel can be made to the situation in a bad job market, where high-paying jobs are scarce, and if one loses such a job, the person will most probably have to settle for a lower-paying one. It is worth noting that the distribution for the fraction of time a water molecule is interior (Figure 4.3) is not bimodal. A second peak near  $f = 0$  would imply the presence of a local phase separation, meaning at least some interfacial water would rarely exchange into interior water. This behavior may be naively expected due to water-solute hydrogen bonds being stronger than water-water hydrogen bonds. However, at 300 K the hydrogen bonds readily break

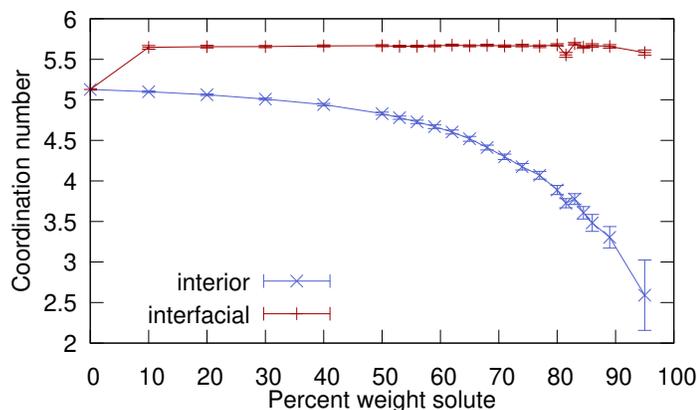


Figure 4.5: The coordination number for interior and interfacial water, i.e. the number of nearest neighbors of an interior or interfacial water averaged over all interior or interfacial waters in all frames. The error bars represent standard deviation.

and reform, allowing water to frequently exchange its category, as can be seen in Figure 4.4. On the other hand, at high solute concentrations, there are much more solute molecules and ions available, so many water molecules can remain interfacial throughout the simulation. Continuing the analogy, this is similar to a situation where there is an abundance of high-paying jobs and not enough people to fill them, and so a qualified individual might never need to work a low-paying job.

The coordination number for both interior and interfacial water is shown in Figure 4.5. This is the average number of nearest neighbor heavy atoms of a water molecule. As can be seen in Figure 4.5, as solute concentration increases, the average number of water oxygens within 0.35 nm of an interior water molecule's oxygen decreases, whereas the average number of nearest neighbors of a interfacial water molecule remains about the same. Once the trehalose matrix begins to stiffen, water will be left to fill any voids. Due to the interfacial state being preferable, water will mainly sit at the interfacial sites, with the remaining water becoming interior and forming a lower-density pool, as opposed to collecting into a bulk-density "droplet" inside the void. These lower density regions could allow for higher local mobility, which could enhance the rate of protein degradation.

In addition, the shape of voids is an important consideration. If the voids are

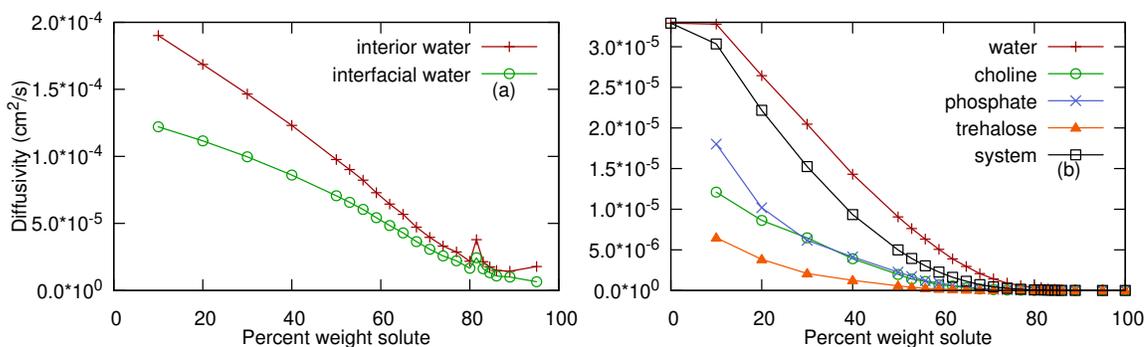


Figure 4.6: The diffusivities of interior and interfacial water, calculated by tracking a particular interior or interfacial water molecule while it remains interior or interfacial, respectively, is shown in (a). For reference, the diffusivities of the various components in the system are shown in (b).

narrow and channel-like, a side chain is more likely to be positioned across it than along it, with most of the side chain being restrained within the glassy matrix. On the other hand, if the void is somewhat spherical, a side chain will have much more freedom to move within it, even if its other end is restrained in the matrix. In this study, the shape of a typical void was found to be highly non-spherical. A 2003 paper studying water percolation in sucrose made a similar observation that "water distribution ... is heterogeneous in the scale of a few angstroms. Water does not form 'pockets' but instead forms scattered clusters with chain-like and star-like portions." [8] Essentially the same clustering geometries are observed in this system.

Finally, as a measure of global mobility, the diffusivities for interior and interfacial water are reported in Figure 4.6 a. A distinction should be made between the previously discussed local mobility, which is a measure of the distance a molecule can travel, e.g. by wiggling around in a void, and global mobility, which relates to transport in the medium. Global mobility implies local mobility, but not vice versa. The diffusivities are calculated for water by tracking a particular interior or interfacial water molecule while it remains interior or interfacial, respectively. Once it switches, the square of its displacement divided by the time elapsed while it was tracked is recorded, and the same molecule is then tracked as the opposite type, with the dis-

placement and time being reset to zero, until it switches again. These records were then averaged for interior and interfacial waters at a particular concentration. The same procedure was repeated for the ions. At low solute concentrations, the interior and interfacial water both have high global mobility, with the interfacial water being less mobile, due to being bound to heavier solute molecules. At higher solute concentrations, as the solute matrix takes up more space, there are less pathways for the interior water to diffuse, decreasing its global mobility and diminishing the difference between interior and interfacial water diffusivities. Nevertheless, interior water always remains more mobile than interfacial water. Note that the interior/interfacial water diffusivities follow the same pattern as the diffusivities in Figure 4.6 b, but are greater by an order of magnitude. None of the diffusivities calculated here are the same as the macroscopic diffusion constants measured experimentally, since simulations are performed on much shorter length and time scales. In particular, the ratio of the diffusion coefficients of any two components should have been a constant, at least in the liquid state, due to the Stokes-Einstein relation, however, this was not the case. A likely reason for this is that the Stokes-Einstein relation is derived under the assumption of a uniform liquid, which is true macroscopically, but on nanometer scales clustering plays a major effect. The interior/interfacial diffusivities are calculated on an even shorter time scales, due to frequent switching, which causes them to differ significantly from the water diffusivity in Figure 4.6 b. Nevertheless, the simulated diffusion coefficient is still useful as a qualitative indicator of relative global mobility.

### 4.3 Molecular clustering analysis

Molecules of a specific type are grouped into distinct clusters by a proximity criterion. Various molecular types, including water molecules in general, interior water, solute in general and trehalose molecules, are considered. Two molecules of the same type are assigned to belong to the same cluster when a path connects them through nearest neighbors. Nearest neighbor connection is made when a heavy atom belong-

ing to one molecule is within 0.35 nm of any heavy atom belonging to the other molecule. All clusters are identified within a given MD frame by considering each specified molecule type of interest, and iteratively searching nearest neighbors of the same type until no new neighbors are added to the search. It should be noted that water clusters are highly dynamic in the sense of having short persistence times relative to the time scale of the simulation. Even though the cluster statistics is the same across frames, clusters of a given size that appear in a given frame are not persistent over time. Rather, the set of water molecules that are a member of a particular cluster will likely be redistributed into several different clusters that can be larger or smaller within a relatively short time ( $\sim 50$  ps).

The RSM for the various molecular clusters was calculated, and the peak in the RSM plot was used to identify the percolation threshold. It was found that the percolation thresholds for interior water and water in general occur around 53% and 81.5% solute concentrations, respectively.

In the cluster analysis, the cluster size distribution for a molecular type of interest was normalized by dividing the number of molecules in clusters of a particular size by the total number of molecules (of a specified type) in the system. This normalization is done per MD frame, and then these distributions are averaged over the entire simulation, which represents the probability distribution for a randomly selected molecule (of a particular type) to be in a cluster of a given size. From Figure 4.7, it can be seen that at high concentrations of solute, where there is an absence of a percolating water cluster, the cluster size distribution resembles an exponential decay. On the other side of the percolation threshold, there is an additional Gaussian-like distribution corresponding to the percolating cluster. From Figure 4.7 and the RSM data, one can see that the percolation threshold for water in general is between 81.5% and 83% solute concentrations.

It was found that the percolation threshold for both general solute clusters and

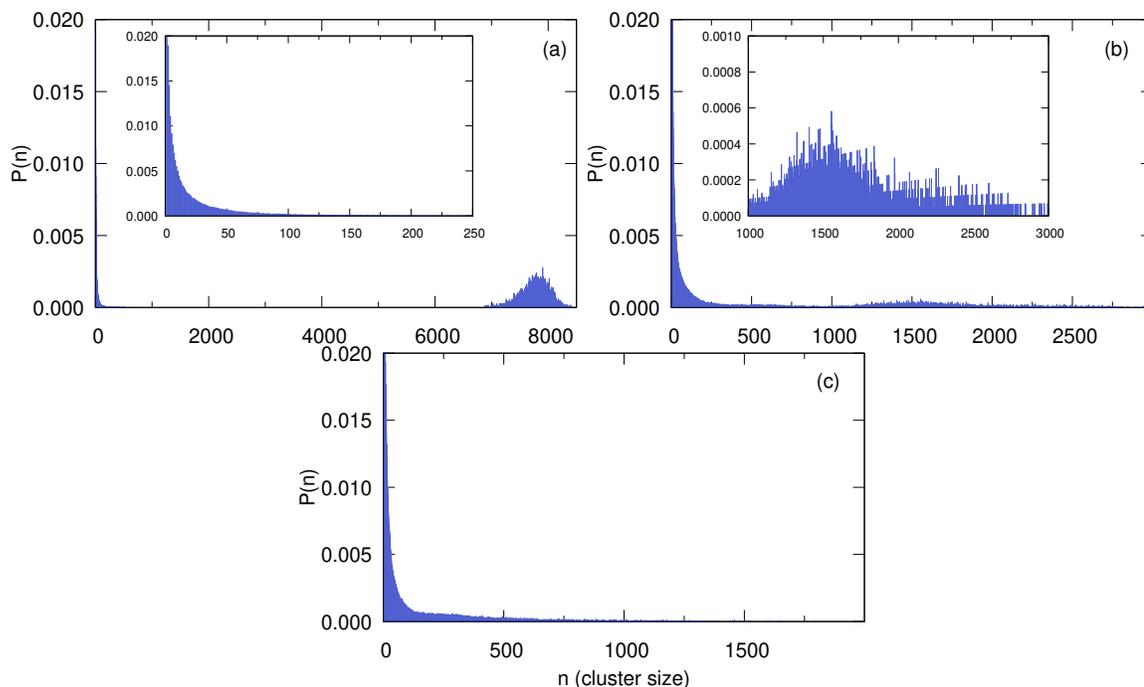


Figure 4.7: The probability  $P(n)$  for a water molecule to be a member of a cluster of size  $n$ , calculated by dividing the number of water molecules in clusters of a particular size by the total number of water molecules in the system. Concentrations of solute: (a) 77%, (b) 81.5% and (c) 83%. Each inset shows a close-up of the interesting part of the corresponding graph.

trehalose-only clusters is between 30% and 40% solute concentrations. Since the volume hysteresis effect (explained in detail in the following chapter) was observed starting around 84.5% solute concentration and the experimental glass transition for trehalose-only (no salt) systems is at  $\sim 90\%$  for temperatures of  $\sim 300$  K [40], it can be safely concluded that neither the solute percolation nor the interior water percolation are relevant to the glass transition. In summary, there are three regimes. First, only water percolates at low solute concentrations, followed by an intermediate phase where solute and water both percolate, and finally the third regime where only solute percolates.

## CHAPTER 5: THE GLASS TRANSITION

### 5.1 Volume hysteresis of a glass

To determine what state a system is in, one can check if the system will release strain. Different amounts of strain can be put into a system by adjusting the initial box size in Packmol. During NPT equilibration, when the volume of the box is allowed to change, the system should be able to release strain.

The final volumes after both rounds of NPT equilibration are shown in Figure 5.1. In the first round, all equilibrations were run for 1 ns, except for 81.5%, 83%, 95% and 100%, which were run for 5 ns, 2 ns, 2 ns and 5 ns, respectively. For all systems, the first round of NPT was followed by NVT equilibration and a production run. For the second round, which only included solute concentrations of 83% and above as well as 80%, new initial coordinates were generated with Packmol, energy minimized, and NPT equilibrated for 5 ns. The initial volume was chosen to be 2744 nm<sup>3</sup>, which was the largest initial volume used in the first round (81.5% system). The second round of NPT runs was followed by NVT equilibration and a production run only for the 95% and 100% systems.

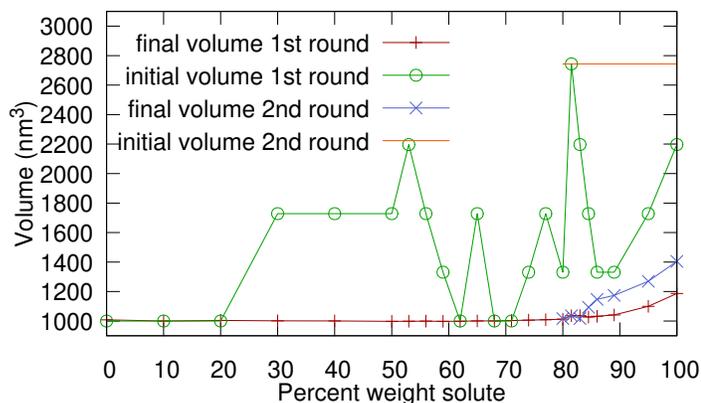


Figure 5.1: Volumes before and after each round of NPT equilibration.

It can be seen that for solute concentrations of 84.5% and above, the final volumes converge to significantly different values when different starting volumes are used. To confirm that this is not due to statistical fluctuation, the NPT equilibration was repeated four more times for the 84.5% system at an initial volume of 1728 nm<sup>3</sup>, with different randomly generated initial coordinates and velocities. The average of all five final volumes was 1038.1 nm<sup>3</sup> and the standard deviation was 10.4 nm<sup>3</sup>.

The observed hysteresis was due to the freezing in of strain into a glass [41, 21, 22]. As an independent confirmation, potential energies were calculated for both production runs of the 95% and 100% systems. The systems with higher volumes showed higher potential energies, which is consistent with the higher volume systems having higher strain. Therefore, one can conclude that the glass transition point is around 84.5% solute by weight, corresponding closely to the water percolation threshold.

## 5.2 An analogy to free volume theory

According to the free volume theory, a system is a liquid when there is a percolating liquid-like cluster. At low solute concentrations this condition is satisfied, but at some point liquid-like clusters will not percolate, since for a dry mixture with a 2:1 stoichiometric ratio of trehalose to CDHP,  $T_g = 358$  K [4], and the simulations were carried out at 300 K. The glass transition, as determined in the previous section, occurs around 84.5%, which is close to the water percolation threshold. Although one case study is insufficient to make definitive conclusions about the effect of water content, the results suggest that the plasticizer behaves as a liquid-like molecule.

Under certain circumstances, however, water can act as an antiplasticizer, creating long-lived hydrogen bonds with the glass-former [42]. In previous studies [43], a similar effect was observed for trehalose, with the absorption of a certain amount of water by a trehalose sample not affecting the glass transition temperature. This is usually explained by water molecules becoming trapped by trehalose and forming a

dihydrate. To verify that all water molecules in the systems act as a plasticizer, the number of water molecules that neighbored the same trehalose molecule for all 30 ns of the simulation was counted. This number was found to be negligibly small even at the highest solute concentrations.

While it makes intuitive sense that the plasticizer acts as a liquid-like molecule, one may ask if there are other liquid-like species in the system. For example, it seems plausible that a fully solvated ion, which is completely surrounded by water molecules and does not interact directly with the solute matrix, would also be liquid-like. If other liquid-like species are indeed present in the system, then the water percolation threshold would serve merely as a lower bound for the glass transition. This is exactly the behavior that was observed in the systems considered. As an additional check, the percolation of clusters consisting of water molecules and fully solvated ions was also considered, but it did not change the percolation threshold. This is not surprising, though, given that free volume theory neglects hydrogen bonds and steric effects, the proximity of the percolation threshold to the glass transition is already more than satisfactory.

## CHAPTER 6: AN ANALYSIS OF THE TRANSITION REGION USING MARKOV CHAINS

### 6.1 Markov chain model for the dynamics of cluster size statistics

#### 6.1.1 Overview

For a system containing  $N$  water molecules, the maximum cluster size cannot exceed  $N$  (all molecules in the same cluster). Thus, the clustering state of the system in a particular frame can be described by an  $N$ -dimensional vector, with the first component being the number of clusters of size 1 (solo waters), the second being the number of clusters of size 2, etc. This vector can then be normalized via division by  $N$ , so that its components are now the probabilities of finding a randomly selected water molecule in a cluster of a particular size.

It seems unlikely that the transition of a tagged water molecule from a cluster of size  $n$  in one frame to a cluster of size  $m$  in the next frame would depend on the sizes of the clusters that the water molecule belonged to in previous frames. Thus, the size of the cluster that the water molecule belongs to in a given frame can be analyzed as a Markov chain [44]. The transition probabilities can be represented as an  $N \times N$  matrix  $\hat{T}$ , which, when successively multiplied on to the probability state vector, will cause the system to evolve and generate the evolution of cluster statistics. For example, one could consider an initial condition that a particular water molecule is isolated, being the sole member of a cluster of size 1. Then, for discrete times (each multiplication), one can track the time evolution of the probability distribution for size of the cluster that the water will belong to. In the long time limit, this distribution will approach the equilibrium [44].

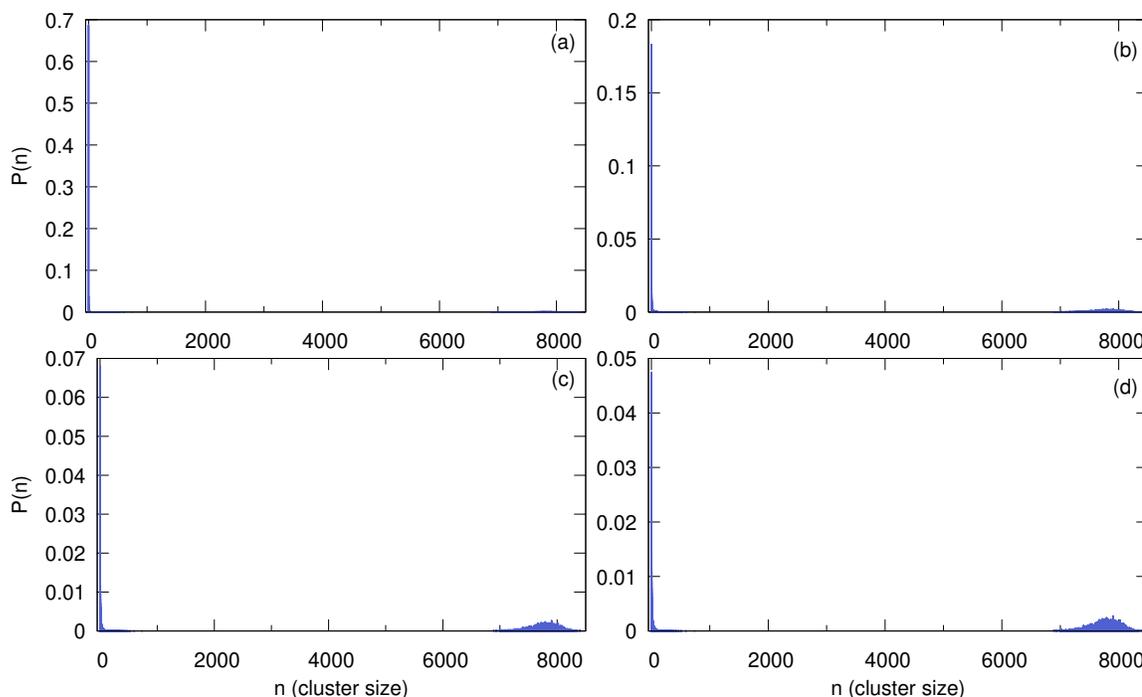


Figure 6.1: Snapshots of the evolution of an initial condition consisting entirely of solo waters after (a) 1 frame, (b) 5 frames, (c) 10 frames and (d) 20 frames.

To demonstrate the evolution of an initial condition under the action of the transition matrix, the transition matrix for the 77% solute system was applied to a state vector with the first component equal to one and all others set to zero. This corresponds to a system where all of the waters are solo waters and no clustering exists. Alternatively, one could also interpret this state vector as it was described in the previous paragraph, corresponding to a particular isolated water molecule, with the future state vectors giving the probability distribution for that same molecule to belong to a cluster of a certain size. Figure 6.1 illustrates this evolution by showing snapshots of the cluster size distribution, as it morphs from delta-function-like to an equilibrium distribution resembling that in Figure 4.7.

One may wonder how fast a system will tend to equilibrium. This question is addressed in the next section through the use of spectral decomposition. The method is similar to normal mode analysis, where the total deviation from equilibrium is the sum of various normal modes, each with its own time constant.

### 6.1.2 Calculation of the transition probabilities

The transition probability for a water in a cluster of size  $n$  to be in a cluster of size  $m$  in the next frame is calculated according to the formula:

$$P(n \rightarrow m) = \left\langle \frac{N_w(m|i)}{n} \right\rangle. \quad (33)$$

The ratio in the angle brackets is evaluated for each cluster of size  $n$  in each individual frame;  $i$  is an index enumerating the clusters of size  $n$  in the current frame and  $N_w(m|i)$  is the number of water molecules in cluster  $i$  in the current frame that will be in clusters of size  $m$  in the next frame. The average is taken over all clusters of size  $n$  in all frames, except for frame 5000, since there is no next frame after that. The formula is derived from the simple definition of the average of a physical quantity over a statistical sample: given a sample of all clusters of size  $n$  in frames 0-4999, one calculates the quantity for each element in the sample and then takes the average over the sample.

Certain cluster sizes, especially in the intermediate range between small clusters and the percolating cluster, may be so rare that they never appear in the 30 ns simulation. Since there is no data for such clusters, one can not average over them, and the corresponding columns in the matrix  $\hat{T}$  will consist entirely of zeros. Although this means that  $\hat{T}$  is technically not a stochastic matrix, since not all of its columns add up to one, it can still be used as such, keeping in mind that the effect of the zero columns would be to produce a zero eigenvalue. Thus, after the matrix is diagonalized, the zero eigenvalues should be discarded as unphysical artifacts.

## 6.2 Eigenvalues of the transition matrices

### 6.2.1 The eigenvalue spectrum of a stochastic matrix

The eigenvalue spectrum of a transition matrix will always contain the eigenvalue 1. The other eigenvalues can, in general, be complex numbers, as the matrix need not be symmetric, but their magnitude must be strictly between 0 and 1 [44]. Due to the asymmetry of the transition matrix, both left and right eigenvectors and eigenvalues can be introduced. Whether left or right eigenvalues will be considered depends on how the matrix was constructed, i.e. whether the rows or columns add up to one. Here, the columns of the matrices will add up to one and right eigenvectors and eigenvalues will be used.

Any state vector can be decomposed over the basis of eigenvectors of  $\hat{T}$ :

$$|\psi\rangle = |1\rangle + \sum_{n=1}^{N-1} c_n |\lambda_n\rangle, \quad (31)$$

where  $|1\rangle$  is the eigenvector corresponding to the eigenvalue 1,  $c_n$  is a decomposition coefficient,  $\lambda_n$  is an eigenvalue not equal to one and  $|\lambda_n\rangle$  is the eigenvector corresponding to that eigenvalue. After applying the matrix  $\hat{T}$  to the vector  $|\psi\rangle$   $M$  times, one gets:

$$\hat{T}^M |\psi\rangle = |1\rangle + \sum_{n=1}^{N-1} \lambda_n^M c_n |\lambda_n\rangle, \quad (32)$$

Thus, since the magnitudes of the  $\lambda_n$ 's are strictly less than one, the eigenvector corresponding to eigenvalue 1 is the equilibrium state, and the system will tend to this state after sufficiently many frames, regardless of the initial condition. To check if the method used to calculate transition probabilities gives reliable results, the equilibrium eigenvectors for the 77%, 81.5% and 83% solute concentration systems were compared to the distributions in Figure 4.7. It was found that the equilibrium eigenvectors closely resemble those distributions.

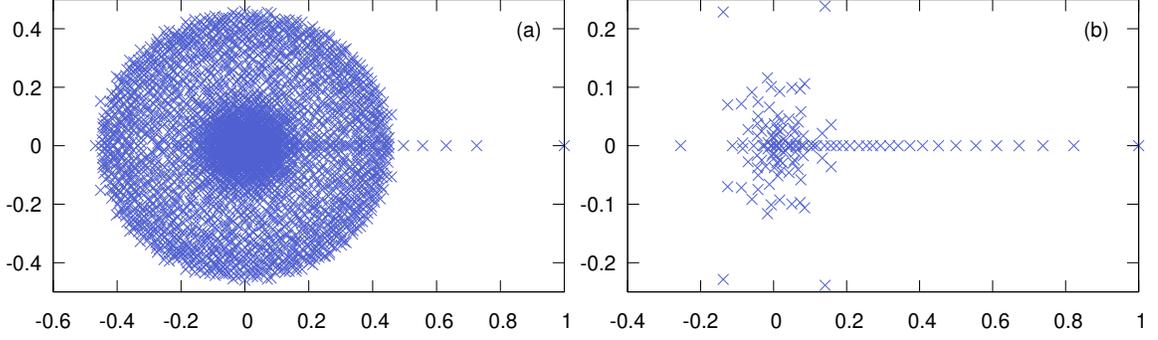


Figure 6.2: Plots of the eigenvalues on the complex plain for (a) the 80% solute concentration system and (b) the 89% solute concentration system.

The eigenvectors corresponding to the other eigenvalues can be interpreted as deviations away from equilibrium. The main interest is in the distribution of magnitudes of these eigenvalues. The existence of a multitude of eigenvalues away from 0 will indicate slow dynamics, otherwise the system will display fast dynamics. Slow dynamics is the hallmark signature of a glass, in contrast to the fast relaxation time characteristic of a liquid. If a clear-cut signature can be observed, this may help in solving the long-standing problem of a robust qualitative criterion for the glass transition [11].

### 6.2.2 Analysis of the spectra

The transition matrices were diagonalized using the SciPy package [45]. The resulting eigenvalues are mostly complex numbers. For reference, Figure 6.2 shows the eigenvalues obtained from the 80% and 89% solute concentration systems.

As was previously mentioned, the magnitude of an eigenvalue is what determines how fast the corresponding eigenvector decays away. To make better sense of this, one can define a time constant of decay,  $\tau_n$ , as follows. Using the identity  $\lambda_n = |\lambda_n|e^{i \arg \lambda_n}$ , one can set  $\lambda_n^M = e^{-t/\tau_n + i\omega_n t}$ , where  $t = M\Delta t$  and  $\Delta t$  is the time interval between consecutive frames, which was 6 ps in this study. It then follows that

$$\tau_n = \frac{-\Delta t}{\ln |\lambda_n|}, \quad (34)$$

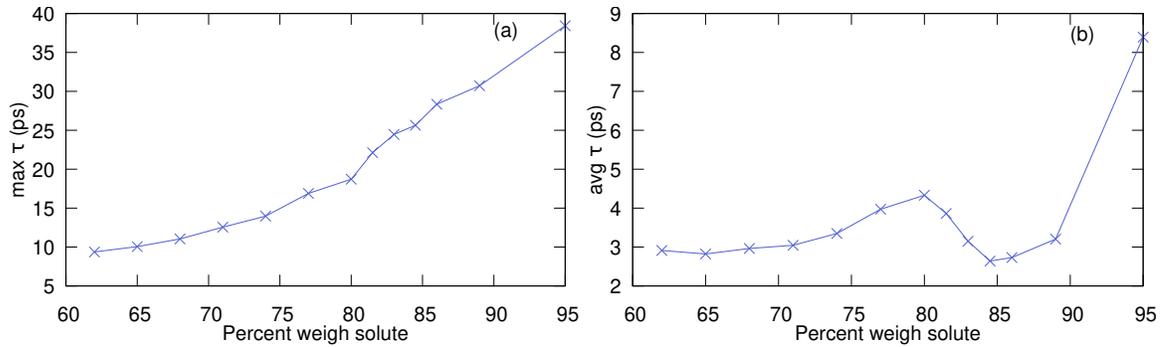


Figure 6.3: The maximum time constant (a), and the average time constant (b), both plotted as functions of solute concentration.

which will be used as the definition of the time constant corresponding to the eigenvalue  $\lambda_n$ . Figure 6.3 a shows the maximum time constant as a function of solute concentration. As one may expect, this quantity increases monotonically with concentration. Thus, at a given concentration, there will always be at least one process that is slower than all processes at any lower concentration. Another interesting metric is the average time constant at a given concentration, which is presented in Figure 6.3 b. Unlike the maximum time constant, the average time constant does not exhibit monotonic behavior. Instead, it gradually increases until it enters the glass transition area, after which it rapidly decreases until it reaches a minimum at 86% and begins to increase again. This interesting behavior may possibly be a signature of the percolation threshold of water or the glass transition, however an explanation of this is not yet available.

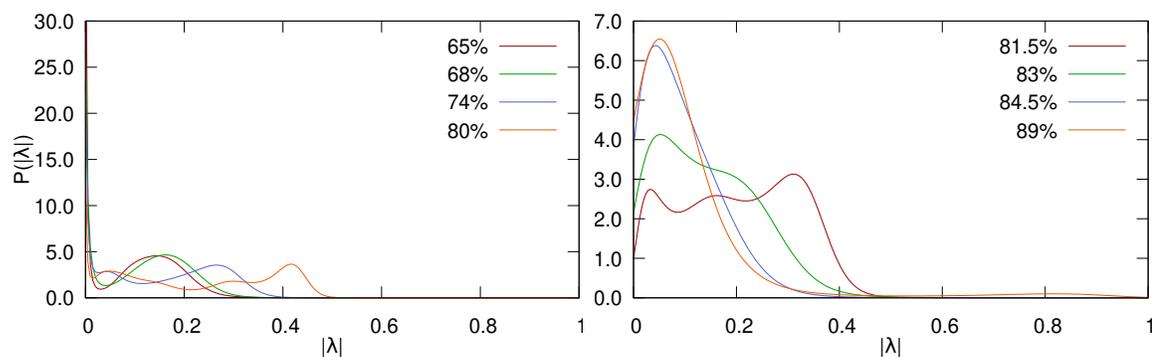


Figure 6.4: Probability distribution functions  $P(|\lambda|)$  of the magnitudes of the eigenvalues  $|\lambda|$  from various concentration systems.

Finally, probability distribution functions of the magnitudes of the eigenvalues have been calculated for various concentrations. As one can see in Figure 6.4, all distributions corresponding to concentrations of 80% solute and below have a prominent sharp peak at zero, whereas for concentrations of 81.5% solute and above, this feature is missing. Moreover, at 89% there is a long tail extending to almost one, which reflects the fact that this system evolves much more slowly than the ones at lower concentrations. The sharp peak at zero, which corresponds to the fast dynamics of a liquid, is exactly the clear-cut signature that was sought after.

## CHAPTER 7: CONCLUSIONS

In this study, molecular dynamics simulations were conducted on a range of different concentrations of ternary mixtures of trehalose and choline dihydrogen phosphate in water. The solute concentration at which the system under consideration becomes a glass was determined to be near 84.5% by finding the point above which a volume hysteresis effect exists. The concept of interior water was introduced, which was defined as the set of water molecules in a particular frame that do not have any non-water neighbors. The water molecules which do have non-water neighbors were referred to as interfacial water. It was observed that water molecules favor the interfacial state, and this was explained in terms of hydrogen bond strength.

The percolation threshold for water was found to be between 81.5% and 83% solute concentrations. In addition to that, it was found that interior water becomes less tightly packed with increasing solute concentration. However, the diffusion coefficients as calculated over persistent times, show little difference between interior and interfacial water at high solute concentration. Therefore, a distinction between the global mobilities of interior and interfacial water becomes blurred around the glass transition, suggesting water in general is behaving as a plasticizer. The proximity of the plasticizer percolation threshold to the glass transition draws a parallel to the percolation of liquid-like clusters in the free volume theory of glass transition, suggesting that the plasticizer molecules can be identified as liquid-like molecules. In the future, this should be checked rigorously by computing the free volumes of water molecules as prescribed by Cohen and Grest [21, 22].

Finally, the concept of Markov chains was applied to the water cluster sizes. Stochastic matrices were obtained from the probabilities of a given water molecule

transitioning from a cluster of size  $n$  in one frame to a cluster of size  $m$  in the next frame. These matrices were diagonalized, and probability distribution functions for the magnitudes of their eigenvalues were calculated. These PDFs show a clear signature of the transition from the fast dynamics of a liquid to slower dynamics characteristic of a glass, in the form of a sharp peak at zero. This peak is present in most liquid PDFs but disappears at 81.5%, which is close to the glass transition as estimated with volume hysteresis. In addition, the time constants for the exponential decay of deviations from equilibrium were considered. While the maximum time constant increased monotonically with solute concentration, the average time constant showed non-monotonic behavior, decreasing around the glass transition. There is not yet a satisfactory explanation for this behavior. This issue is worth considering in future work.

The analysis methods developed in this thesis can be readily applied to other similar systems. For example, most of the analysis programs were written to work automatically with systems in which CDHP is replaced by choline monohydrogen phosphate (CMHP). Work is currently underway to simulate CMHP systems, after which the same analysis programs can be used with no modification to the code. This will provide a vast amount of data for comparison to the data presented in this thesis. Such a comparative study may give valuable insight into the role of the salt in the behavior of the systems.

## REFERENCES

- [1] J. H. Crowe, J. F. Carpenter, and L. M. Crowe, "The role of vitrification in anhydrobiosis," *Annual review of physiology*, vol. 60, no. 1, pp. 73–103, 1998.
- [2] S. Ohtake, Y. Kita, and T. Arakawa, "Interactions of formulation excipients with proteins in solution and in the dried state," *Advanced drug delivery reviews*, vol. 63, no. 13, pp. 1053–1073, 2011.
- [3] W. Rao, H. Huang, H. Wang, S. Zhao, J. Dumbleton, G. Zhao, and X. He, "Nanoparticle-mediated intracellular delivery enables cryopreservation of human adipose-derived stem cells using trehalose as the sole cryoprotectant," *ACS applied materials & interfaces*, vol. 7, no. 8, pp. 5017–5028, 2015.
- [4] L. Weng, R. Vijayaraghavan, D. R. MacFarlane, and G. D. Elliott, "Application of the kwei equation to model the tg behavior of binary blends of sugars and salts," *Cryobiology*, vol. 68, no. 1, pp. 155–158, 2014.
- [5] L. Weng and G. D. Elliott, "Distinctly different glass transition behaviors of trehalose mixed with  $\text{Na}_2\text{HPO}_4$  or  $\text{NaH}_2\text{PO}_4$ : Evidence for its molecular origin," *Pharmaceutical research*, vol. 32, no. 7, pp. 2217–2228, 2015.
- [6] J. Reis, R. Sitaula, and S. Bhowmick, "Water activity and glass transition temperatures of disaccharide based buffers for desiccation preservation of biologics," *Journal of Biomedical Science and Engineering*, vol. 2, no. 08, p. 594, 2009.
- [7] L. Weng and G. D. Elliott, "Local minimum in fragility for trehalose/glycerol mixtures: Implications for biopharmaceutical stabilization," *The Journal of Physical Chemistry B*, vol. 119, no. 22, pp. 6820–6827, 2015.
- [8] V. Molinero, T. ÇaÇşın, and W. A. Goddard Iii, "Sugar, water and free volume networks in concentrated sucrose solutions," *Chemical physics letters*, vol. 377, no. 3-4, pp. 469–474, 2003.
- [9] G. Caliskan, A. Kisiuk, A. Tsai, C. L. Soles, and A. Sokolov, "Influence of solvent on dynamics and stability of a protein," *Journal of non-crystalline solids*, vol. 307, pp. 887–893, 2002.
- [10] G. Elliott and B. Bagheri, 2018. US Pat. 9 930 883 B2.
- [11] C. P. Royall and S. R. Williams, "The role of local structure in dynamical arrest," *Physics Reports*, vol. 560, pp. 1–75, 2015.
- [12] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS User Manual version 5.1.2*. [www.gromacs.com](http://www.gromacs.com), 2016.

- [13] S. V. Sambasivarao and O. Acevedo, “Development of opls-aa force field parameters for 68 unique ionic liquids,” *Journal of chemical theory and computation*, vol. 5, no. 4, pp. 1038–1050, 2009.
- [14] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, “Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids,” *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225–11236, 1996.
- [15] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.
- [16] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak, “Molecular dynamics with coupling to an external bath,” *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [17] G. Bussi, D. Donadio, and M. Parrinello, “Canonical sampling through velocity rescaling,” *The Journal of chemical physics*, vol. 126, no. 1, p. 014101, 2007.
- [18] D. Chandler, *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.
- [19] J. Hoshen and R. Kopelman, “Percolation and cluster distribution. i. cluster multiple labeling technique and critical concentration algorithm,” *Physical Review B*, vol. 14, no. 8, p. 3438, 1976.
- [20] R. Zallen, *The physics of amorphous solids*. John Wiley & Sons, 2008.
- [21] M. H. Cohen and G. Grest, “Liquid-glass transition, a free-volume approach,” *Physical Review B*, vol. 20, no. 3, p. 1077, 1979.
- [22] G. S. Grest and M. H. Cohen, “Liquids, glasses, and the glass transition: A free-volume approach,” *Adv. Chem. Phys.*, vol. 48, pp. 455–525, 1981.
- [23] I. M. Kalogerias and H. E. Hagg Lobland, “The nature of the glassy state: structure and glass transitions,” *Journal of Materials Education*, vol. 34, no. 3, p. 69, 2012.
- [24] J. G. Montoro and J. Abascal, “The voronoi polyhedra as tools for structure determination in simple disordered systems,” *The Journal of Physical Chemistry*, vol. 97, no. 16, pp. 4211–4215, 1993.
- [25] V. Luchnikov, N. Medvedev, Y. I. Naberukhin, and H. Schober, “Voronoi-delaunay analysis of normal modes in a simple model glass,” *Physical Review B*, vol. 62, no. 5, p. 3181, 2000.

- [26] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1, pp. 19–25, 2015.
- [27] W. Damm, A. Frontera, J. Tirado-Rives, and W. L. Jorgensen, "Opls all-atom force field for carbohydrates," *Journal of Computational Chemistry*, vol. 18, no. 16, pp. 1955–1970, 1997.
- [28] T.-J. Lin and H. Heinz, "Accurate force field parameters and ph resolved surface models for hydroxyapatite to understand structure, mechanics, hydration, and biological interfaces," *The Journal of Physical Chemistry C*, vol. 120, no. 9, pp. 4975–4992, 2016.
- [29] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J. C. Delepine, P. Cieplak, and F.-Y. Dupradeau, "Red server: a web service for deriving resp and esp charges and building force field libraries for new molecules and molecular fragments," *Nucleic acids research*, vol. 39, no. suppl\_2, pp. W511–W517, 2011.
- [30] F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski, and P. Cieplak, "The red. tools: Advances in resp and esp charge derivation and force field library building," *Physical Chemistry Chemical Physics*, vol. 12, no. 28, pp. 7821–7839, 2010.
- [31] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model," *The Journal of Physical Chemistry*, vol. 97, no. 40, pp. 10269–10280, 1993.
- [32] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, *et al.*, "General atomic and molecular electronic structure system," *Journal of computational chemistry*, vol. 14, no. 11, pp. 1347–1363, 1993.
- [33] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, "Packmol: a package for building initial configurations for molecular dynamics simulations," *Journal of computational chemistry*, vol. 30, no. 13, pp. 2157–2164, 2009.
- [34] H. Nagase, N. Ogawa, T. Endo, M. Shiro, H. Ueda, and M. Sakurai, "Crystal structure of an anhydrous form of trehalose: structure of water channels of trehalose polymorphism," *The Journal of Physical Chemistry B*, vol. 112, no. 30, pp. 9105–9111, 2008.
- [35] K. Fujita, D. R. MacFarlane, K. Noguchi, and H. Ohno, "Choline dihydrogen phosphate," *Acta Crystallographica Section E: Structure Reports Online*, vol. 65, no. 4, pp. o709–o709, 2009.

- [36] A. Bondi, “van der waals volumes and radii,” *The Journal of physical chemistry*, vol. 68, no. 3, pp. 441–451, 1964.
- [37] J. Farmer and D. Jacobs, “High throughput nonparametric probability density estimation,” *PloS one*, vol. 13, no. 5, p. e0196937, 2018.
- [38] I. J. Van Den Dries, D. Van Dusschoten, M. A. Hemminga, and E. Van Der Linden, “Effects of water content and molecular weight on spin probe and water mobility in malto-oligomer glasses,” *The Journal of Physical Chemistry B*, vol. 104, no. 44, pp. 10126–10132, 2000.
- [39] M. D. Fayer, “Dynamics of water interacting with interfaces, molecules, and ions.,” *Accounts of chemical research*, vol. 45, no. 1, pp. 3–14, 2011.
- [40] A. Aksan and M. Toner, “Isothermal desiccation and vitrification kinetics of trehalose- dextran solutions,” *Langmuir*, vol. 20, no. 13, pp. 5521–5529, 2004.
- [41] B. Wunderlich, “The nature of the glass transition and its determination by thermal analysis,” in *Assignment of the glass transition*, ASTM International, 1994.
- [42] G. N. Ruiz, M. Romanini, A. Hauptmann, T. Loerting, E. Shalaev, J. L. Tamarit, L. C. Pardo, and R. Macovez, “Genuine antiplasticizing effect of water on a glass-former drug,” *Scientific reports*, vol. 7, no. 1, p. 7470, 2017.
- [43] N. Ekdawi-Sever, J. J. de Pablo, E. Feick, and E. von Meerwall, “Diffusion of sucrose and  $\alpha$ ,  $\alpha$ -trehalose in aqueous solutions,” *The Journal of Physical Chemistry A*, vol. 107, no. 6, pp. 936–943, 2003.
- [44] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, vol. 31. Springer Science & Business Media, 2013.
- [45] E. Jones, T. Oliphant, P. Peterson, *et al.*, “SciPy: Open source scientific tools for Python,” 2001–. (Online).