Modeling Water Using Mercedes-Benz Particles

Daniel Gray

CHM 434A – Chemistry Thesis

Advisor - Dr. Matthew Srnec

6 May 2019

Abstract

Unlike the conditions classical biochemical experiments utilized, a cell is almost entirely volume excluded, and this may affect every major cellular process. Atomistic simulations of crowding are not plausible due to the need to simulate many protein crowders at once. While different approaches to course-graining are possible, we started with simplifying the water model by using 2D Mercedes-Benz particles. Each particle has a simple Lennard-Jones term and an orientation-dependent hydrogen-bonding term. NPT Monte Carlo and NVT molecular dynamics implementations of this model are examined for consistency with themselves and with the qualitative structural and thermodynamic features of water. Specifically, we found the Mercedes-Benz particles mimic water's temperature of maximum density, minimum in isothermal compressibility, region of negative thermal expansion, high heat capacity, and qualitative diffusion, enthalpy, and phase transition trends. This indicates that an explicit distinction between hydrogen bond donors and acceptors and longer-range electrostatics are not essential to model the physical characteristics that give rise to the anomalous properties of water.

Introduction

Macromolecular Crowding and Volume Exclusion

Biochemical reactions have traditionally been carried out in test tubes where the total concentration of solutes is less than 1 mg/mL (1). Conversely, biochemical reactions are natively found in cells whose total concentration of macromolecules ranges from 50-400 mg/mL (2), corresponding to a volume occupancy of up to 40% (3). No individual macromolecule is present in significant quantity so nothing in the cell is concentrated; the media of the cell is referred to as crowded. The volume occupied by a given macromolecule cannot be occupied by other macromolecules due to steric repulsion. It has been postulated that macromolecular crowding can significantly affect protein folding stability, conformation, binding of substrates, enzymatic kinetics, protein-protein interactions, and protein aggregation, many of which have been experimentally or theoretically verified to some degree (4). As such, traditional biochemical measurements of enzyme rates, equilibria, and mechanisms may be inaccurate representations of their host organism.

Available volume is defined as the accessible volume in a system to the center of mass of a given species (1), while excluded volume is the volume not available. As such, the available volume depends on the size and shape of the species being considered (**Figure 1**). In cells, the volume available to small molecules (like water) would be close to the entire volume not occupied by macromolecules, but the available volume to macromolecules is significantly less due to their larger size. In a cell with approximately 30% macromolecule volume occupancy, less than 1% of the cell volume is available to a subsequent macromolecule (5). Random diffusion under these conditions could not account for all biological function, and it has been appreciated for some time now that nearly every process in a cell occurs via large complexes instead of individual macromolecules randomly diffusing (6).

Figure 1: The volume in the round hole is excluded to the square peg, but presumably not to gaseous molecules in the air. This demonstrates how volume exclusion is dependent on the size and shape of the species being considered.

Quantification of Crowding

In vivo experiments related to crowding are experimentally challenging to perform accurately due to the complex environment of the cell which includes many interactions not attributable to crowding. The problem becomes increasingly difficult since crowding effects depend on the chemical composition of the crowders and their environment (7). This necessitates a theoretical understanding of the physical effects crowding may have on proteins. However, pure theoretical work suffers from a lack of simulated data and often disagrees with experiment on the basic effects of crowding (8). Furthermore, simulations of crowded environments should validate theoretical understandings and predict experimental results. However, atomistic simulations of large protein complexes in extended crowded environments is untenable using current technology. The biochemistry of macromolecular complexes, however, can be captured by models much courser than atomistic (9), provided the functional attributes of the system relevant to the data being collected remain.

The development of computationally feasible simulations can be done in two ways, both of which are driven by decreasing the degrees of freedom of the system. One can start with an atomistic model and group atoms together into units with properties such as size, charge, and hydrophobicity defined by preset variables. Course-graining in this way is heavily dependent on the choice of atom groupings and how the variables are determined (10). Another approach is to start with an over-simplified version of the system,

characterize it, and then add degrees of freedom until it behaves in the manner desired. The Mercedes-Benz (MB) model of water is an example of this second approach, where two-dimensional hard circles with added properties such as hydrogen-bonding 'arms' are used to model water (11). The advantages of this is that unneeded aspects of the system can be easily omitted, and properties found to be lacking can be added to increase the accuracy of the model.

Monte Carlo (MC) methods are computational algorithms relying on random sampling to obtain numerical results. This can be thought of in terms of integration where numerical integrals can be obtained by repeatedly checking if random points are within the function space or not (**Figure 2**). Markov Chain MC (MCMC) methods use a probability distribution of a known function as its integrand space. The purpose of MCMC is to sample the system over the entire probability distribution. The specific algorithm used in this research is the Metropolis-Hastings algorithm first proposed in 1953 (13). This algorithm makes a random move to the system, then accepts or rejects the move based on an acceptance function which is constructed from the probability distribution. The key here is that instead of weighting each possible configuration against the probability distribution, the Metropolis-Hasting algorithm uses the distribution to choose the configurations then weights them all evenly. Typically, a Boltzmann distribution is used, which relies on knowing the energy of each configuration. When applying this method, one must be able to calculate the energy of the system as a function of configuration alone.

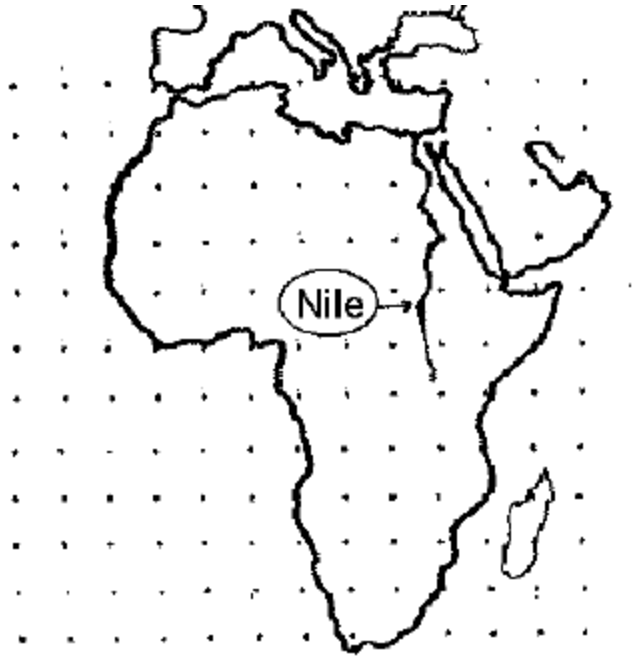

Figure 2: A schematic representation of MC integration. Counting the dots in the space of Africa gives an approximation of Africa's area, with a more accurate numerical approximation if the number of dots is increased. Adapted from (12).

Molecular dynamics (MD) numerically solves Newton's equations of motion to simulate a system with a time scale. The force on each particle is calculated, then the velocity of each particle is changed using Newton's second law. Once the particles are moved, new forces can be calculated. A given time-step must be used and the force on each particle is assumed to be constant for the duration of that time. Therefore, this method will become increasingly accurate as the time-step is lowered. This was first used in chemical applications by simulating liquid argon as spheres which exhibit a classic Lennard-Jones potential (14). To use MD, one must be able to calculate the forces on each particle in the system as a function of position alone. The advantage of MD simulations is that, because of the time scale, more relevant time-resolved data can be obtained such as diffusion coefficients. The limitations of MD are the force fields used to generate the accelerations on each particle. In MC, the energy is a function of position alone, while in MD, the force (which is the derivative of energy) is a function of position. This relationship allows the same model to be simulated using MC and MD programs.

Research Goals and Hypothesis

In this project, the goal was to develop the MB model of water for scaling up in generalized protein systems under crowded environments. We started with water because it is the solvent in all cells and has unusual structural and thermodynamic properties. MB is a relevant model for water since water's defining features are its cage-like structure and high heat capacity. Atomistic representations of water are likely overly-detailed and are not necessarily accurate (15). Furthermore, in these experiments we are not studying water directly, but rather are using the water as a solvent. Thus, all we need is the statistical mechanical averaging, which MB water promises (16). Both MC and MD algorithms were utilized for the simulations. Once the MB water was characterized, the system was scaled up and crowders can eventually be introduced.

The primary hypothesis of this project is that MB water mimics the structure and thermodynamics of liquid water relevant to model volume exclusion. This hypothesis was tested by by calculating the density, heat capacity, isothermal compressibility, thermal expansion coefficient, and diffusion coefficient of MB water under various conditions, while also examining phase transition trends. The secondary hypothesis of this project was that the introduction of crowders would reduce the diffusion coefficient of MB water and any solutes present. The other properties of MB water were hypothesized to change due to increased amounts of interfacial water in crowded environments (17).

Background

Model Description

Mercedes-Benz (MB) particles are represented as circles with three hydrogen-bonding 'arms' (**Figure 3**). In this model potential energy is a function of position alone. MB particles use the Lennard-Jones (LJ) potential (Equation 1) to model the London forces where ε_{LJ} is the depth of the well, σ_{LJ} is the optimum interaction distance, and r_{ij} is the distance between any two particles. MB water also has an orientation-dependent 'hydrogen-bonding' term to model hydrogen bonding (Equation 2), where ε_{HB} gives the hydrogen-bonding strength, r_{HB} is the optimum hydrogen-bonding distance, $\bar{u}_{ij}$ is the unit vector from particle i to particle j, $\cdot$ is the dot product, and the $\hat{\imath}_k$ and $\hat{\jmath}_m$ terms are the kth and mth hydrogen-bonding

arms (represented by a unit vector along the arm) of the ith and jth particle, respectively. G(x) is defined as the Gaussian in Equation 3 where σ_{HB} is a parameter giving the shape of the Gaussian.

$$U_{LJ} = 4 * \varepsilon_{LJ} * \left[\left(\frac{\sigma_{LJ}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{LJ}}{r_{ij}}\right)^{6}\right] \quad [1]$$

$$U_{HB} = \varepsilon_{HB} * G(r_{ij} - r_{HB}) * \sum_{k=1}^{3}\sum_{m=1}^{3}\left[G(\hat{\imath}_k \cdot \hat{u}_{ij} - 1) * G(\hat{\jmath}_m \cdot \hat{u}_{ij} + 1)\right] \quad [2]$$

$$G(x) = e^{\left(\frac{-x^2}{2\sigma_{HB}^2}\right)} \quad [3]$$

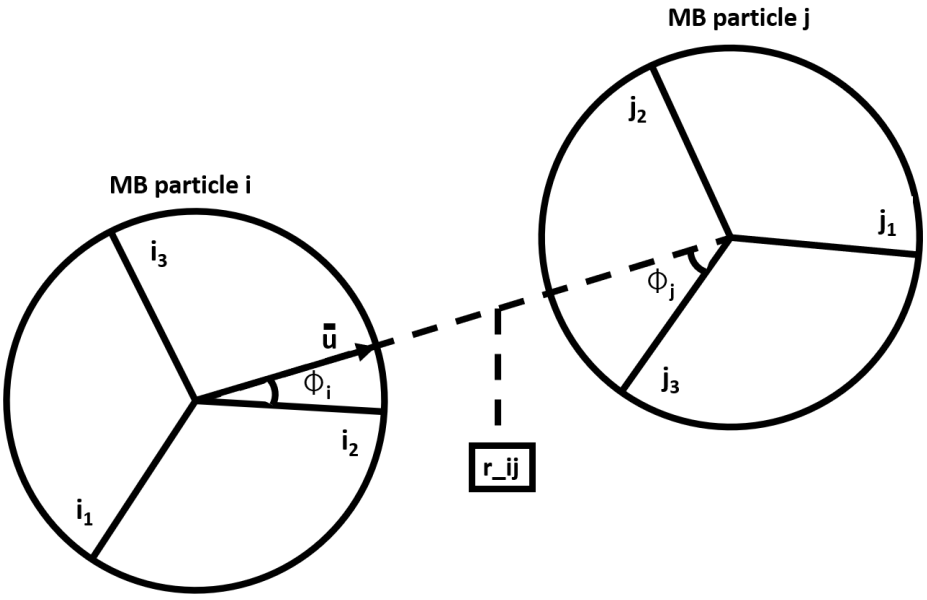

Figure 3: Schematic of two MB particles and the notation used to represent the different values.

Given any two particles, a maximum one pair of hydrogen-bonding arms will give a non-negligible value of U_{HB}, meaning the interaction from all other combinations of arms can be ignored. In the case pictured in **Figure 3**, the only relevant combination is i_2 and j_3. This is determined mathematically using Equations 4 and 5 to find the maximum value of $i_k \cdot \bar{u}_{ij}$ and minimum value of $j_m \cdot \bar{u}_{ij}$, respectively. This allows for a simplified version of U_{HB} given by Equation 6, where i_{max} is always in the interval (0, 1] and j_{min} is always in the interval [-1, 0) due to the geometry of the MB water (**Figure 3**). Finally, the energy of a pair of particles is the sum of U_{LJ} and U_{HB}, and the total energy of the system is simply the sum of all the pairs of particles (Equation 7), where N is the number of particles in the simulation. It is important to note

that Equation 7 implies there are no long-range effects or interactions involving 3 or more particles. The details of this model were adapted from (18), (19), and (20).

$$i_{max} = max\left(i_1 \cdot \hat{u}_{ij}, i_2 \cdot \hat{u}_{ij}, i_3 \cdot \hat{u}_{ij}\right) \quad [4]$$

$$j_{min} = min\left(j_1 \cdot \hat{u}_{ij}, j_2 \cdot \hat{u}_{ij}, j_3 \cdot \hat{u}_{ij}\right) \quad [5]$$

$$U_{HB} = \varepsilon_{HB} * G\left(r_{ij} - r_{HB}\right) * G(i_{max} - 1) * G(j_{min} + 1) \quad [6]$$

$$U_{total} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(U_{LJ,ij} + U_{HB,ij}\right) \quad [7]$$

Water at normal pressure forms ice I_h below its freezing point. If the crystal structure of ice I_h is examined down the (001) crystal face one can see how two-dimensional water would freeze (**Figure 4**). The three hydrogen-bonding arms used in MB water come from this observation of hexagonal ice. At low enough temperatures MB forms crystal-like structures reminiscent of ice I_h especially when looking down the (001) crystal face. As such, MB water has the appropriate 2D geometry of water. In this model there are a total of five parameters (ε_{LJ}, σ_{LJ}, r_{HB}, ε_{HB}, σ_{HB}) and each particle requires three degrees of freedom (x, y, and an orientation variable) to describe. Furthermore, due to the 2-dimenstional structure of MB water, 120 particles correspond to over 1000 3D water molecules[†].

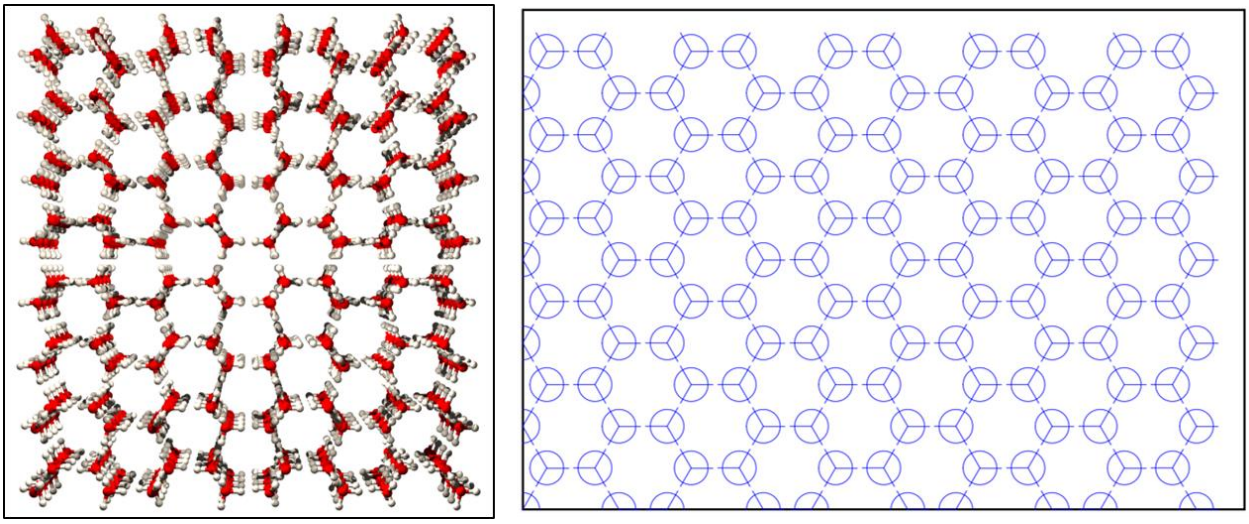

Figure 4: Left: ice I_h crystal structure schematic looking down the (001) face. Right: MB water 'ice' phase. Note the hexagonal structure in both liquid and MB water.

[†] This can be calculated by (120)^3/2 = 1314

Both MC and MD simulations typically operate under periodic boundary conditions and the minimum image convention. This allows for smaller simulations and minimizes surface effects. Periodic boundary conditions operate by removing hard walls and allowing any particle that exits the box to reenter on the opposite side. Furthermore, when calculating particle interactions, the minimum image convention states that the given particle will interact with the nearest image of whatever particle it is interacting with (**Figure 5**).

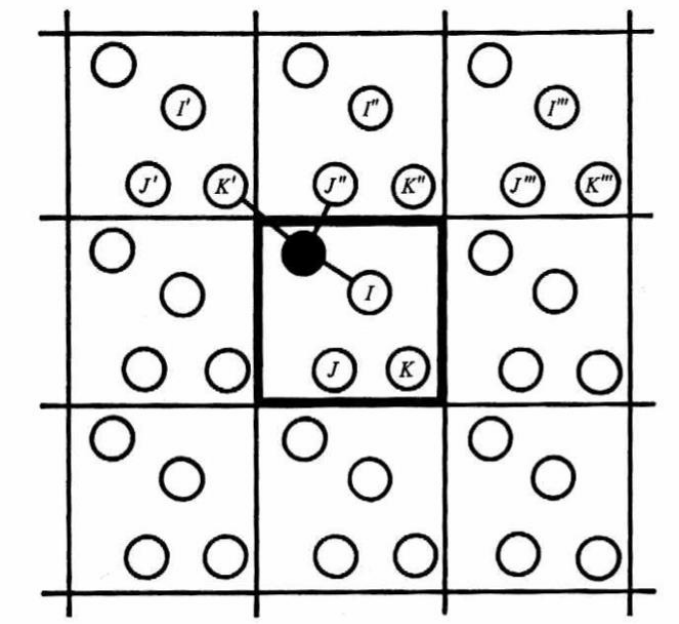

Figure 5: Schematic of periodic boundary conditions; the box in the middle is the system. Note the black circle 'sees' the minimum image of K (K') and J (J'') instead of the K and J in the actual system.

Metropolis MCMC Algorithm

The Metropolis-Hastings algorithm is a MCMC method used to generate a sequence of sample configurations which approximate a probability distribution P(x). Unlike other MC algorithms which take all configurations and weight them to the distribution, the Metropolis algorithm uses the distribution to find the configurations then weights them all evenly. In physical simulations, the relevant distribution is often the Boltzmann distribution. If we take A to be an arbitrary normalization constant, then $P(x) = A*\exp(-E(i)*\beta)$ where E(i) is the energy of configuration i and $\beta = 1/k_BT$. Once an initial configuration is determined, each step of the algorithm works by first making a random move (random change to the

configuration). For the MB water model, this consists of randomly choosing a particle and moving its x, y, or orientation variable (Equation 8), where V is the variable being changed (x, y, or orientation), E is a random number in range (0, 1], and ΔV is a predefined maximum change in V (18).

$$V \rightarrow V + E * \Delta V \quad [8]$$

The acceptance function used gives the probability that the move will be accepted, and is the same for x, y, orientation, or volume changes (Equation 9). Note that for all but the volume moves, $V_f = V_i$ so $\Delta H = \Delta U$. For volume moves, Equation 10 must be used to find the enthalpy change (21).

$$Probability = \begin{cases} 1.0 if \Delta H \leq 0 \\ e^{\left(\frac{-\Delta H}{k_B * T}\right)} if \Delta H > 0 \end{cases} \quad [9]$$

$$\Delta H = \Delta U + P * \Delta V - k_B * T * N * ln\left(\frac{V_f}{V_i}\right) \quad [10]$$

The ensemble that these simulations operate under can vary, but in this project's MC simulations, the isothermal-isobaric (NPT) ensemble was utilized. This means the number of particles (N), the temperature (T), and the pressure were held constant. The pressure was held constant indirectly by incorporating a random volume move (V from Equation 8 would be the volume of the container) every five passes, where a pass is defined as N steps. The volume moves would scale all particle positions and box lengths. Other ensembles are certainly possible, with the canonical (NVT), microcanonical (NVE), and Grand canonical (μVT) being the most useful (22). Typically, these simulations take between 30,000 – 150,000 passes for their energy (and volume in the NPT ensemble) to equilibrate, and after that useful data can be collected. Equations 11, 12, and 13 give the ensemble formulas for the heat capacity, isothermal compressibility, and coefficient of thermal expansion, respectively.

$$C_p = \frac{\langle H^2 \rangle - \langle H \rangle^2}{N * T^2} \quad [11]$$

$$\kappa = \frac{\langle V^2 \rangle - \langle V \rangle^2}{T * \langle V \rangle} \quad [12]$$

$$\alpha = \frac{\langle V * H \rangle - \langle V \rangle * \langle H \rangle}{T^2 * \langle V \rangle} \quad [13]$$

Finally, the coefficient of self-diffusion (D) can be calculated using a mean square deviation (23) derived from the Einstein-Stokes equation first published in (24). Equation 14 states the mean square displacement is proportional to the observation time, with a proportionality constant D (d is the dimensionality of the system). Therefore, if we save the trajectory of the simulation, we can easily plot the mean square displacement vs observation time and find the limiting slope. Note that in these MC simulations our 'time' variable is number of passes, so the self-diffusivity is not found in natural units and is only informative for qualitative comparison among various MC simulations.

$$D = \frac{1}{2*d} * \lim_{t\to\infty} \frac{\langle [r(t_o+t)-r(t_o)]^2 \rangle}{t} \quad [14]$$

MD Algorithm

Molecular mechanical MD simulations use Newton's second law to determine the trajectory of the particles. To accomplish this, one must have force as a function of position. From classical physics we have F = -∇U so to find the force we must differentiate our potential functions. For the Lennard-Jones potential, the energy is independent of orientation, so dU/do = 0 where o is the orientation. Equation 15 gives the total derivative of the Lennard-Jones potential and Equation 16 gives the total derivative of the hydrogen-bonding potential. The details of this setup were derived from (25) and (26) and the derivatives of the individual gaussians in Equation 16 can be found in (25).

$$F_{LJ} = -\nabla U_{LJ} = \frac{-\partial U}{\partial x} dx - \frac{\partial U}{\partial y} dy = -24 * \varepsilon_{LJ} * \sigma_{LJ}^6 * \left(\frac{2*\sigma_{LJ}^6}{r_{ij}^{13}} - \frac{1}{r_{ij}^7}\right) * \left(\frac{\Delta x}{r_{ij}} + \frac{\Delta y}{r_{ij}}\right) \quad [15]$$

$$F_{HB} = -\nabla U_{HB} = -\varepsilon_{HB} * \left[\nabla G(r_{ij} - r_{HB}) * G(i_{max} - 1) * G(j_{min} + 1) + G(r_{ij} - r_{HB}) * \dots\right]$$

$$\dots \left[* \nabla G(i_{max} - 1) * G(j_{min} + 1) + \dots + G(r_{ij} - r_{HB}) * G(i_{max} - 1) * \nabla G(j_{min} + 1)\right] \quad [16]$$

If the standard Taylor series expansion for the coordinates of the particles q(t) around t is taken out to the quadratic term, then Equation 17 is yielded. Remembering that v(t) = dq(t)/dt and a(t) = dv(t)/dt, we can substitute to give Equation 18. Similarly, Equation 19 can be derived from the constant-acceleration Taylor expansion of the velocity. We know from kinematics that these equations are valid for constant acceleration, which we can assume is the case for any discrete dt, provided the time step is small enough. Practically speaking, the time step should be chosen as large as possible without affecting results to save

computation time. After each step, the forces must be recalculated, which is the most time-consuming step in any MD simulation.

$$q(t+\delta t) = q(t) + \frac{\partial q(t)}{\partial t}\big((t+\delta t) - t\big) + \frac{1}{2}\frac{\partial^2 q(t)}{\partial t^2}\big((t+\delta t) - t\big)^2 \quad [17]$$

$$q(t+\delta t) = q(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \quad [18]$$

$$v(t+\delta t) = v(t) + \frac{1}{2}a(t)\delta t \quad [19]$$

The standard implementation of the velocity Verlet algorithm was used in the MD program which starts with a given position q(t). The half-step velocity v(t + 0.5*dt) is calculated using Equation 19 then the new position q(t + dt) is calculated using Equation 18. Then, once the second half-step velocity is calculated, the new forces are calculated from Equations 15 and 16. Several similar MD implementations are possible, but all will give equivalent results when executed properly with a small enough time step. The MD program written used the NVT ensemble and the Berendsen thermostat to maintain the temperature. The Berendsen thermostat works by re-scaling the velocities of the particles to control the temperature of a simulation (27). The Berendsen thermostat was implemented by applying Equation 20 to each particle every 5 steps, where τ is a constant characteristic of the thermostat and T_{curr} is the current temperature calculated from the velocities of the particles. While there exist inaccuracies with the Berendsen thermostat, but it should yield roughly correct results for most properties (28), so was used due to its efficiency.

$$v' = v * \sqrt{1 + \frac{\delta t}{\tau}\left(\frac{T}{T_{curr}} - 1\right)} \quad [20]$$

Parameters and Implementation

Parameters used in the MC and MD simulations are summarized in Table 1. The Monte Carlo simulations were implemented in the NPT ensemble with a 50% acceptance ratio. Every 5*N moves (5 passes) a volume move was made with a 50% acceptance ratio. MB water did not reliably freeze into its native ice structure, so every simulation started with an ice structure to allow for the hydrogen-bonding network to be established. Energy and density were equilibrated for 250,000 passes each simulation and the derivatives were equilibrated for an additional 9 million passes. This allowed for reproducible values for each quantity calculated under each simulation. The MD simulations used a time step of 0.001 and were ran for 1200 time units (1.2 million steps each simulation). The energy and temperature were equilibrated for 200 time units and the derivatives equilibrated for 1000 time units.

Variable	Value	Description	Real Value (29)
ε_{LJ}	0.1	Energy of van der Waals interaction	1.274 kJ/mol
r_{LJ}	0.7	Distance to ideal van der Waals interaction	2.065 Å
ε_{HB}	-1.0	Energy of ideal hydrogen bond	12.74 kJ/mol
r_{HB}	1.00	Ideal hydrogen bond distance	2.95 Å
σ_{HB}	0.085	Gaussian width parameter: determines H-bond spread	0.25 Å
k_B	1.0	Boltzmann constant: relates temperature to kinetic energy	1.38×10^{-23} J/K
mass	1.0	Reduced mass of the MB water particles	2.99×10^{-26} kg
I_{HB}	0.0126	Reduced moment of inertia of MB water particles	Varies by axis
τ	0.20	Constant used in the Berendsen thermostat	NA

Table 1: Variables used in the MC and MD simulation programs written for this project. The mass and moment of inertia were used to convert force to acceleration by F = ma.

Both programs were written in Java. Matplotlib in Python was used to draw snapshots of the simulations (including all the ones in this paper). The simulations were run on a variety of machines over several months and each simulation was done in triplicate to ensure precision. Reduced units were used for all simulations as shown in Table 1. MB water is not meant to quantitatively mimic liquid water, but should demonstrate the important qualitative characteristics of water. Therefore, all subsequent comparisons are done in reduced units and will be denoted by * for a unitless quantity.

Results

The first simulations done were to examine an appropriate simulation size to characterize MB water with the MC program. The goal was to find the smallest sample that did not exhibit so-called finite-size effects, where the smallness of the sample size would affect the data. This was examined by running

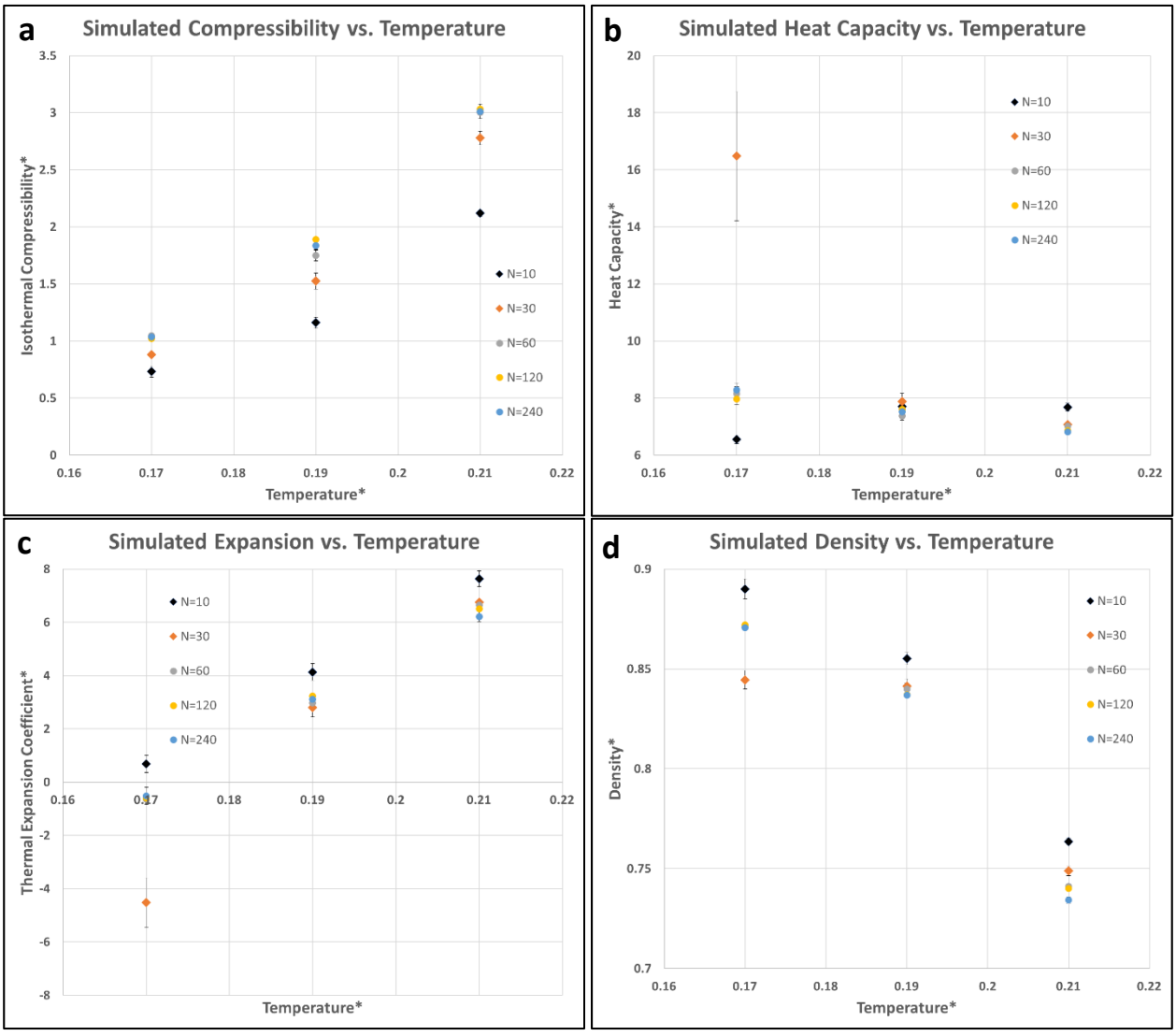

Figure 6: Finite-size effect experiments. **6a** shows the trend in isothermal compressibility while **6b** shows the isobaric heat capacity trend, **6c** is thermal expansion coefficient, **6d** is density.

simulations of 10, 30, 60, 120, and 240 MB particles. **Figure 6** shows the results of the isothermal

compressibility, heat capacity, thermal expansion coefficient, and density trends. This data shows that a system size of less than 60 particles exhibits significant finite size effects. The deviations in compressibility were larger for higher temperatures (**Figure 6a**) and the deviations in thermal expansion were larger for smaller temperatures (**Figure 6b**). This correlated well with the variance in the compressibility and thermal expansion. Practically, this means a particle number that does not deviate significantly for any temperature is warranted. Therefore, using less than 60 particles would result in erroneous data. Original implementation of the MB model in MC simulations used 60 particles (18). However, to ensure accuracy a minimum of 120 particles henceforth were used in all simulations.

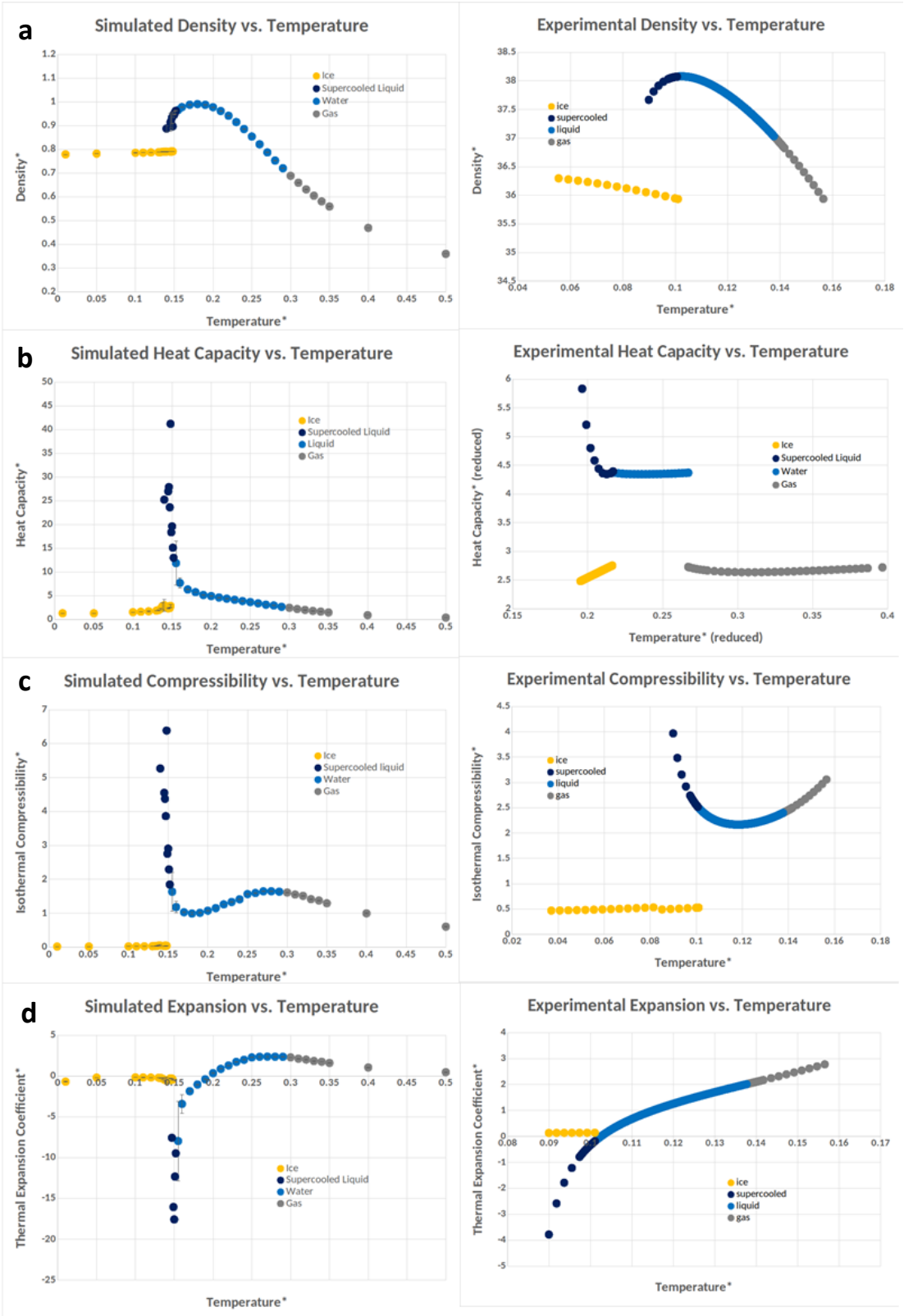

Figure 7: Comparison of MC simulations and experimental data all in reduced units. **7a** shows the density, **7b** the isobaric heat capacity, **7c** the isothermal compressibility, and **7d** the coefficient of thermal expansion. All simulations were done with 120 MB particles.

A reduced pressure of 0.19 was originally used to study MB water; however, it was noted that the pressure was likely in the supercritical region (18). Here, we present results at a reduced pressure of 0.19 with 120 particles which agree with the results already published. **Figure 7** shows the comparison of several quantities with experimental liquid water. Note the qualitative features of water are expressed by the MB model, including a temperature of maximum density (**Figure 7a**), a high liquid heat capacity (**Figure 7b**), a minimum in the isothermal compressibility (**Figure 7c**), and a region of negative thermal expansion coefficient (**Figure 7d**). Note the distinction between the supercooled liquid and the ice. Previous studies of MB water (18, etc.) would report its melting point at T*=0.14. However, at T*=0.14 the MC simulations would sometimes melt and sometimes not, as shown in **Figure 8**. Both simulations had identical starting conditions (both began with identical ice structure coordinates) but simulation **8a** equilibrated at an ice structure and **8b** equilibrated at a liquid structure. This demonstrates that the MC model captures the supercooled liquid behavior of water.

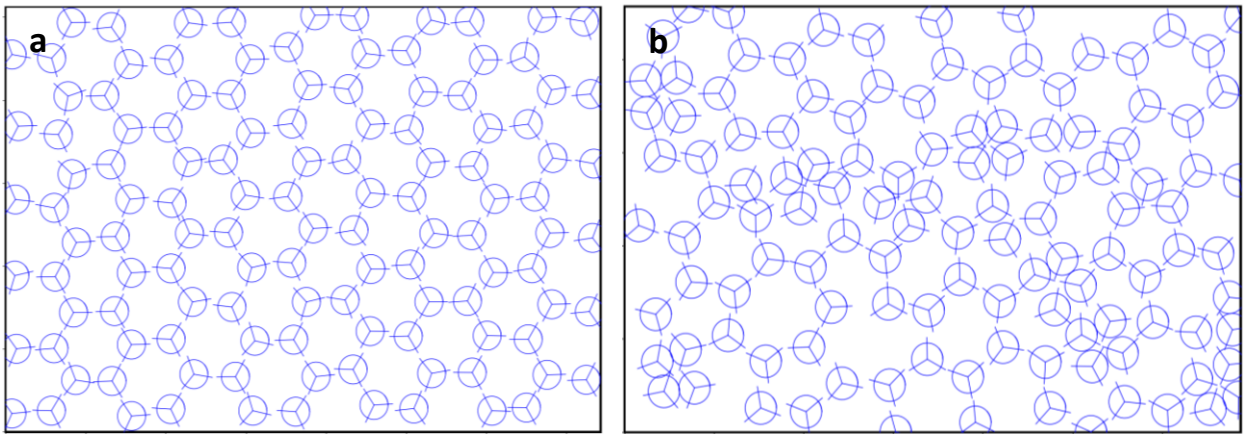

Figure 8: Demonstration that at a given temperature and pressure, the MB model can equilibrate to either a solid or liquid structure. **8a** shows solid MB water at T*=0.14 and **8b** shows supercooled liquid at T*=0.14, just above the spinodal point at P*=0.19.

From **Figure 7**, one can easily estimate the melting point of MB water as 0.16, and further see why previous researchers chose 0.14 (which would instead be a spinodal point (16)). As such, to find the actual melting point, simulations were run at 0.001 temperature intervals starting with ice. The temperature was increased until the system no longer equilibrated to an ice structure, as determined by 8 simulations equilibrating to liquid. This was much more rigorous than any previous studies and gave accurate melting

(and spinodal) points at an uncertainty of T* +/- 0.002. From **Figure 7** one can also notice there exists a distinct phase transition between the solid and liquid, but no clear distinction between the liquid and gas phases. The liquid-gas transition on the graphs were drawn qualitatively by examining snapshots of the simulations. However, this clearly indicates that P*=0.19 is above the critical pressure.

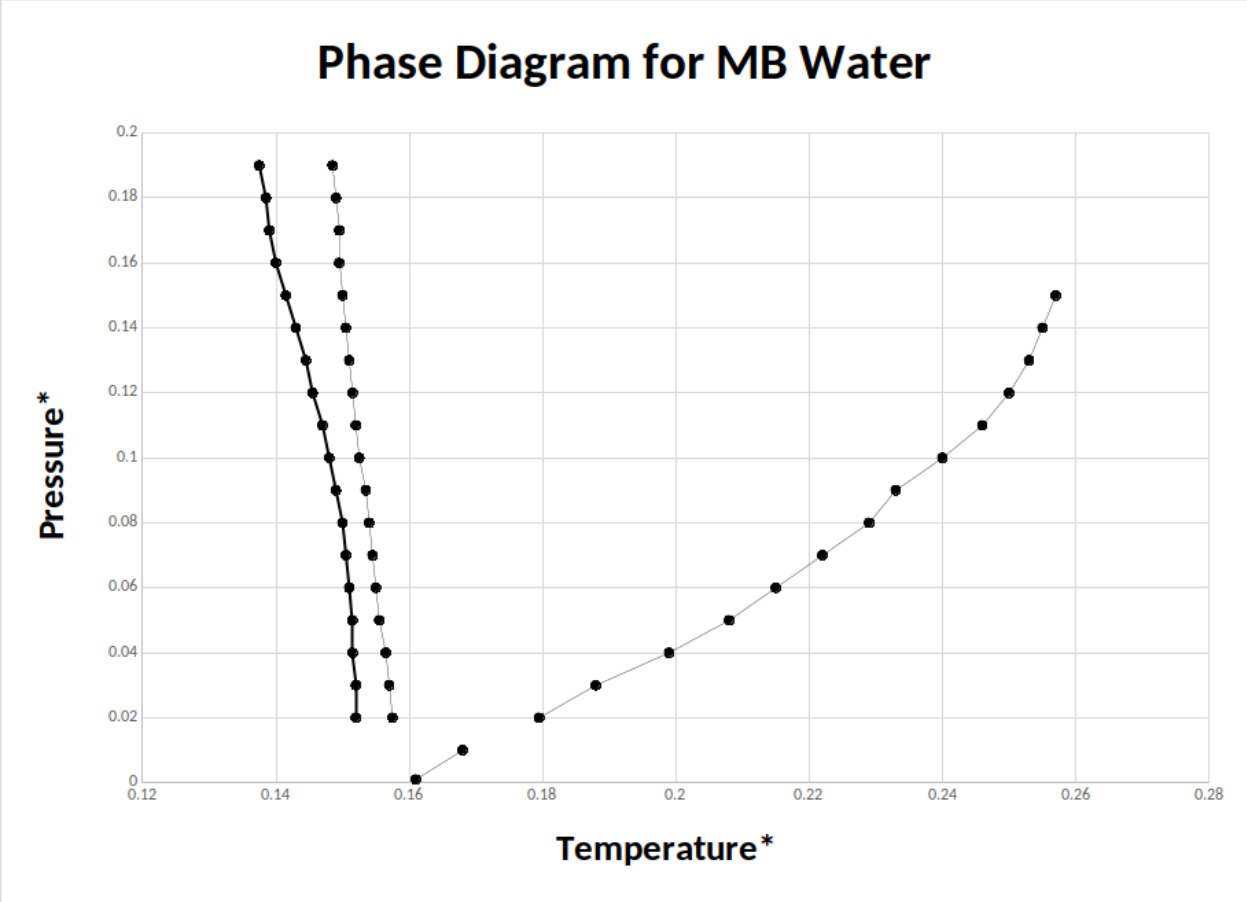

Figure 9: Phase diagram for MB water. The triple point was not found exactly but should be around T*=0.16 and P*=0.001.

To further characterize MB water and to determine an appropriate 'ambient' pressure for the MD (NVT) simulations, the phase diagram of MB water was constructed using the same method used above. The spinodal, melting, and boiling point of MB water was determined for several pressures. The result is enumerated in **Figure 9**, where the spinodal point line correlates with previously reported melting points of MB water. Note the negative slope of the solid-liquid transition line, which is the defining characteristic of the water phase diagram. By interpolation, we can estimate the triple point to be near P*~0.001 and T*~0.16. Further, the critical pressure is likely about P*~0.15, but there was no rigorous determination of

this value. For the rest of the simulations a pressure of P*=0.08 was used and for the MD simulations a volume was chosen such that it matched a pressure of P*=0.08.

Initial MD simulations of 120 particles had energy and temperature convergence and agreed with the MC simulations on energy. The % difference in energy was < 0.1% for each temperature simulated

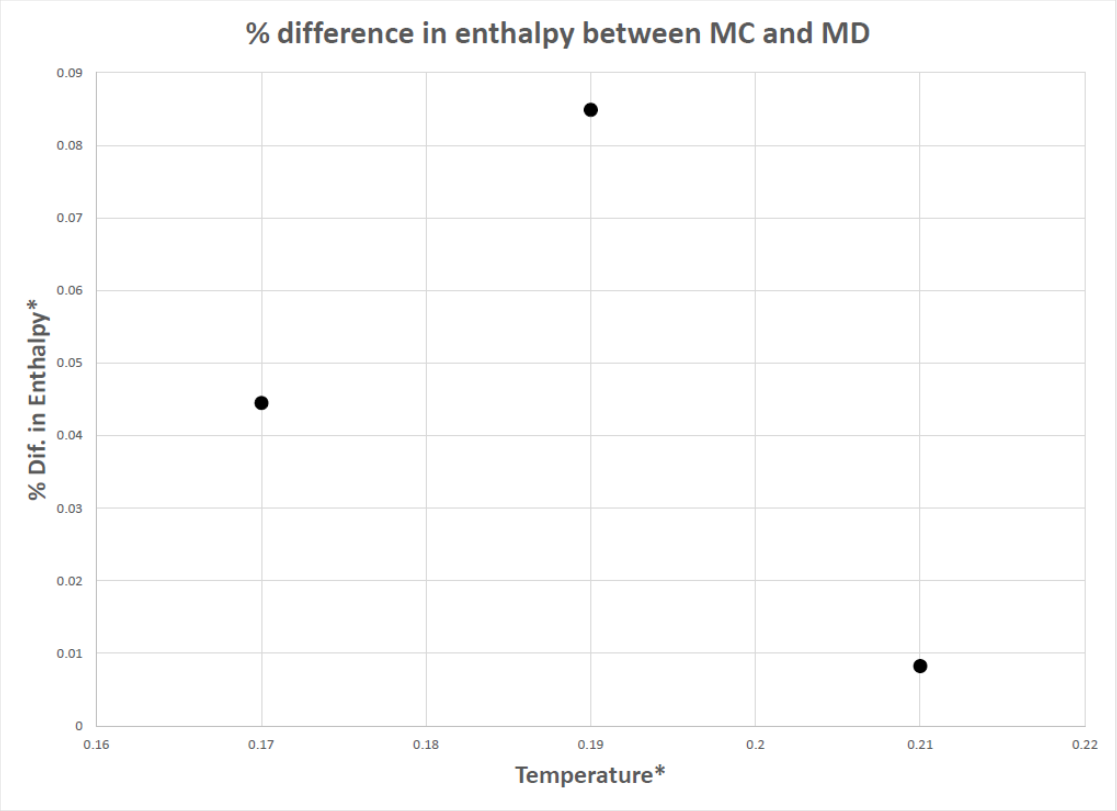

Figure 10: Difference in enthalpy between NPT MC simulations and NVT MD simulations.

(**Figure 10**). The MD converged on pressure that were too high (~0.2 for the densities found by MC at 0.08) at all the temperatures considered (range from 0.17 – 0.21). This may be due to pressure not being calculated the same in the MC code: it was wrapped up in the acceptance function through equation 10 but calculated directly in the MD. The constant-volume heat capacity was less than the constant-pressure heat capacity as expected (30). When the MD simulations were scaled up to 1000 particles, the results agreed with the 120-particle measurement but with about half the variance for each calculated values.

The next experiments done were to introduce solutes into MD simulations of 480 MB particles. The solutes introduced were LJ solutes the same size as the water and could simulate small hydrophobic solutes comparable to the size of water. Physically, these LJ particles could represent non-polar gases dissolved in water. As expected, the solutes disrupted the hydrogen-bonding network in the liquid MB system. As such, the potential energy of the system went down as the number of solutes increased. This correlated with a decrease in the heat capacity as well (**Figure 11**), since the high heat capacity is due to the hydrogen bonding. The diffusion coefficient increased as the number of solutes increased, which is also expected since the water can move more freely when less hydrogen bonding is present. Finally, the kinetic energy stayed the same (**Figure 11**) which is expected since the temperature was held constant.

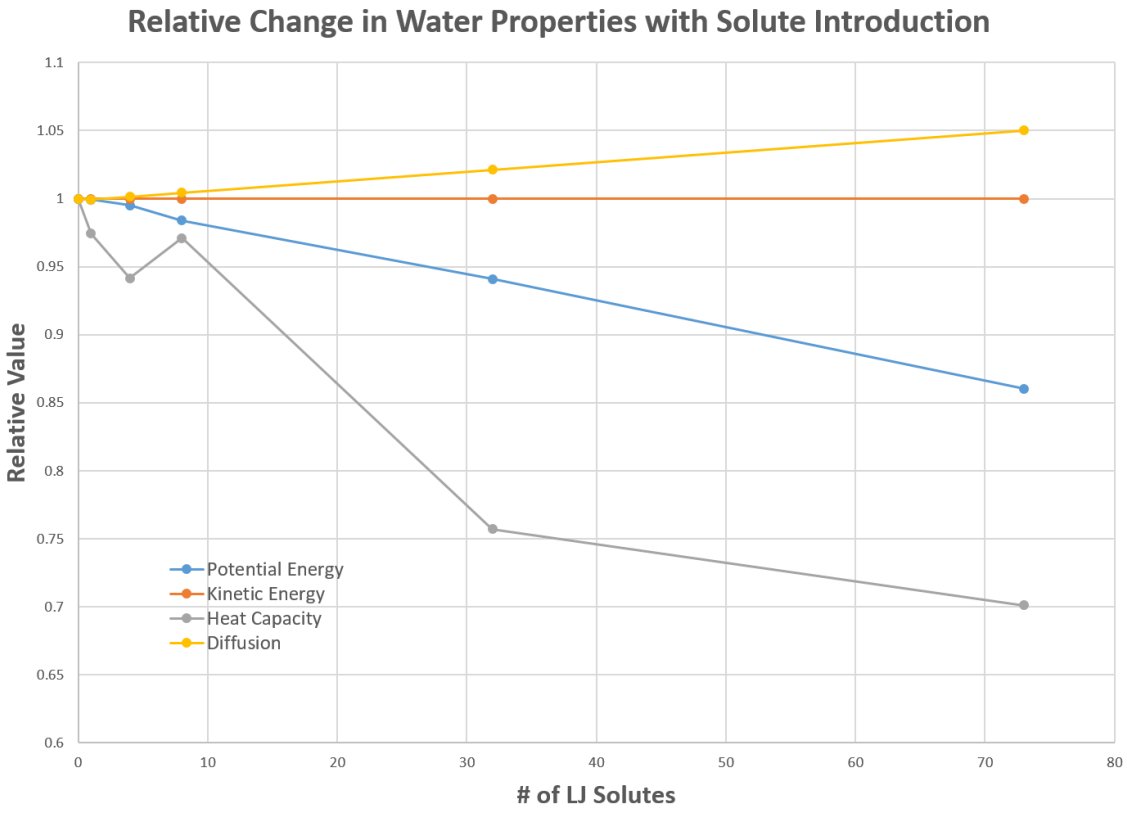

Figure 11: Relative change in MB water properties with introduction of LJ solutes. Note the decrease in heat capacity and potential energy, the increase in diffusion, and the stability of kinetic energy. This data was collected at P*=0.08 and T*=0.19.

Conclusion

These simulations characterized MB water using MC simulations under the NPT ensemble and MD simulations under the NVT ensemble. The qualitative properties of MB water correlated well with liquid water, including reasonable enthalpy and diffusion trends with temperature, a temperature of maximum density, a high heat capacity, a minimum isothermal compressibility, a region of negative coefficient of thermal expansion, a phase diagram comparable to that of water's, a higher pressure-volume heat capacity than constant-volume heat capacity, and finally a reasonable change in properties with introduction of hydrophobic solutes. Scaling up the model gave identical results but with less variance. To use this model on systems with crowding, generalized protein particles of reduced diameter ~200 would need to be used, which means systems of over 50,000 MB particles need to be utilized. Parallelizing the program would be necessary for this task. However, before crowding experiments are done the discrepancy between the MC and MD pressure should be addressed.

Acknowledgment

I'd like to thank everyone who helped me along the way in this project. This includes Dr. Rohde for setting up the internship at Duquesne where I started the project, the chemistry department at Duquesne, Dr. Evanseck's mentoring, Dr. Workman's many answers to stupid questions, Dr. Srnec for setting up his computers for my simulations and being my advisor for this thesis, and the entire chemistry department at Franciscan University for the work they do for us students.

References

1. Minton AP. The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media (2001) *J. Biol. Chem.* **276**, 10577-10580
2. Zimmerman SB, Trach SO. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli (1991) *J. Mol. Biol.* **222**, 599-620
3. Rivas G, Ferrone F, Herzfeld J. Life in a crowded world (2004) *EMBO Rep.* **5**, 23–27
4. Kuznetsova IM, Turoverov KK, Uversky VN. What Macromolecular Crowding Can Do to a Protein (2014) *Int. J. Mol. Sci.* **15**, 23090-23140
5. Ellis RJ, Minton AP. Join the Crowd (2003) *Nature* **425**, 27-28
6. Alberts B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists (1998) *Cell* **92,** 291-294
7. Rivas G, Minton AP. Macromolecular crowding in vitro, in vivo, and in between (2016) *Trends Biochem. Sci.* **41**, 970-981
8. Dix JA, Verkman AS. Crowding Effects on Diffusion in Solutions and Cells (2008) *Annual Review of Biophysics* **37**, 247-263
9. Boninsegna L, Banisch R, Clementi C. A Data-Driven Perspective on the Hierarchical Assembly of Molecular Structures (2018) *J. Chem. Theory Comput.* **14**, 453-460
10. Clementi C. Coarse-grained models of protein folding: toy models or predictive tools? (2008) *Current Opinion in Structural Biology* **18**, 10-15
11. Ben-Naim A. Statistical Mechanics of "Waterlike" Particles in Two Dimensions: Physical Model and Application of the Percus–Yevick Equation (1971) *J. Chem. Phys.* **54**, 3682
12. Frenkel D, Smit B. Understanding Molecular Simulation: From Algorithms to Applications (2001). *Academic Press*, 2nd edition. ISBN: 0122673514
13. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines (1953) *Journal of Chemical Physics* **21**, 1087-1092
14. Rahman A. Correlations in the Motion of Atoms in Liquid Argon (1964) *Phys. Rev.* **136**, A405
15. Florova P, Sklenovsky P, Banas P, Otyepka M. Explicit Water Models Affect the Specific Solvation and Dynamics of Unfolded Peptides While the Conformational Behavior and Flexibility of Folded Peptides Remain Intact (2010) *J. Chem. Theory Comput.* **6**, 3569-3579
16. Brini E, Fennell CJ, Fernandez-Serra M, Hribaa-Lee B, Luksic M, Dill KA. How Water's Properties Are Encoded in Its Molecular Structure and Energies (2017) *Chem. Rev.* **117**, 12385-12414
17. Harada R, Sugita Y, Feig M. Protein Crowding Affects Hydration Structure and Dynamics (2012) *J. Am. Chem. Soc.* **134**, 4842-4849
18. Silverstein KAT, Haymet ADJ, Dill KA. A Simple Model of Water and the Hydrophobic Effect (1998) *J. Am. Chem. Soc.* **120**, 3166-3175
19. Andaloro G, Sperandeo-Mineo RM. Monte Carlo Simulations of Hydrophobic Hydration for Pedagogical Purposes (1990) *European Journal of Physics* **11**, 275-283
20. Southall NT, Dill KA. The Mechanism of Hydrophobic Solvation Depends on Solute Radius (1999) *J. Phys. Chem.* **104**, 1326-1331
21. Vlugt TJH, van der Earden JPJM, Dijkstra M, Smit B, Frenkel D. Introduction to Molecular Simulation and Statistical Thermodynamics (2008) ISBN: 978-90-9024432-7
22. Urbic T, Vlachy V, Dill KA. Confined Water: A Mercedes-Benz Study (2006) *J. Phys. Chem.* **110**, 4963-4970
23. Keffer D. The Working Man's Guide to Obtaining Self Diffusion Coefficients from Molecular Dynamics Simulations (2001) *Department of Chemical Engineering* lecture notes. Accessed 12 March 2019 from http://www.cs.unc.edu/Research/nbody/pubs/external/Keffer/selfD.pdf
24. Einstein A. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen (1905) *Annalen der Physik* **322**, 549-560
25. Nerukh D, Pavlov E, Bardik V, Scukins A. Molecular Dynamics implementation of BN2D or 'Mercedes Benz' water model (2015) *Computer Physics Communications* **190**, 129-138
26. Markesteijn A, Karabasov S, Pavlov E, Nerukh D, Scukins A. Multiscale molecular dynamics/hydrodynamics implementation of two dimensional "Mercedes Benz" water model (2015) *Eur. Phys. J. Special Topics* **224**, 2217-2238
27. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular Dynamics Coupled to an External Bath (1984) *J. Chem. Phys.* **81**, 3684-3690
28. Morishita T. Fluctuation Formulas in Molecular-Dynamics Simulations with the Weak Coupling Heat Bath (2000) *J. Chem. Phys.* **113**, 2976-2982
29. Wagner W, Prub A. The IAPWS Formulation 1995 for the Thermodynamic Properties of Ordinary Water (2002) *J. Phys. Chem. Ref. Data* **31**, No. 2
30. David R. Gaskell. *Introduction to the Thermodynamics of Materials* (2008), 5th edition, Taylor and Francis .