Calculating Properties of Bulk Water with Molecular Dynamics

Laura J. Kinnaman\*† and Carrie S. Miller‡

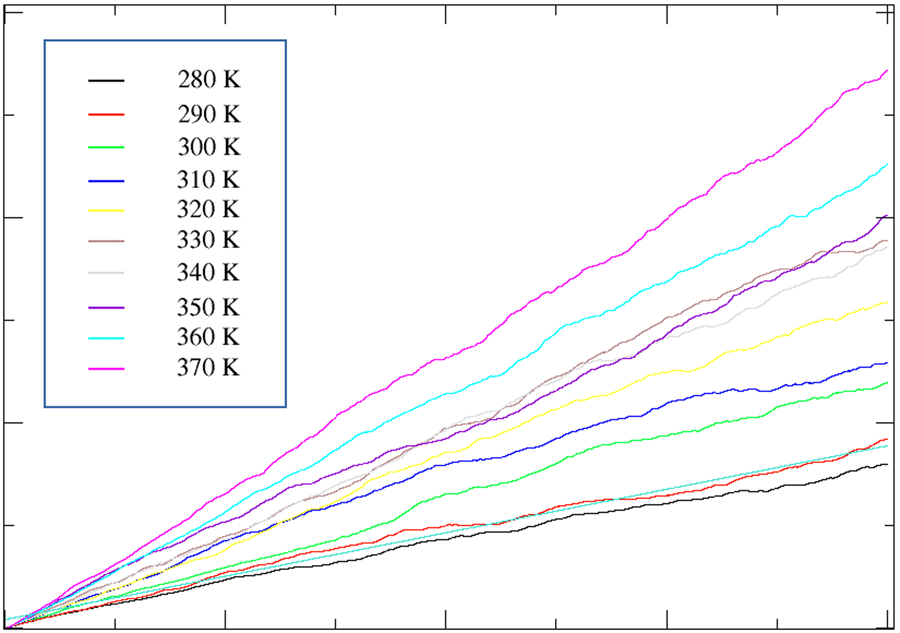
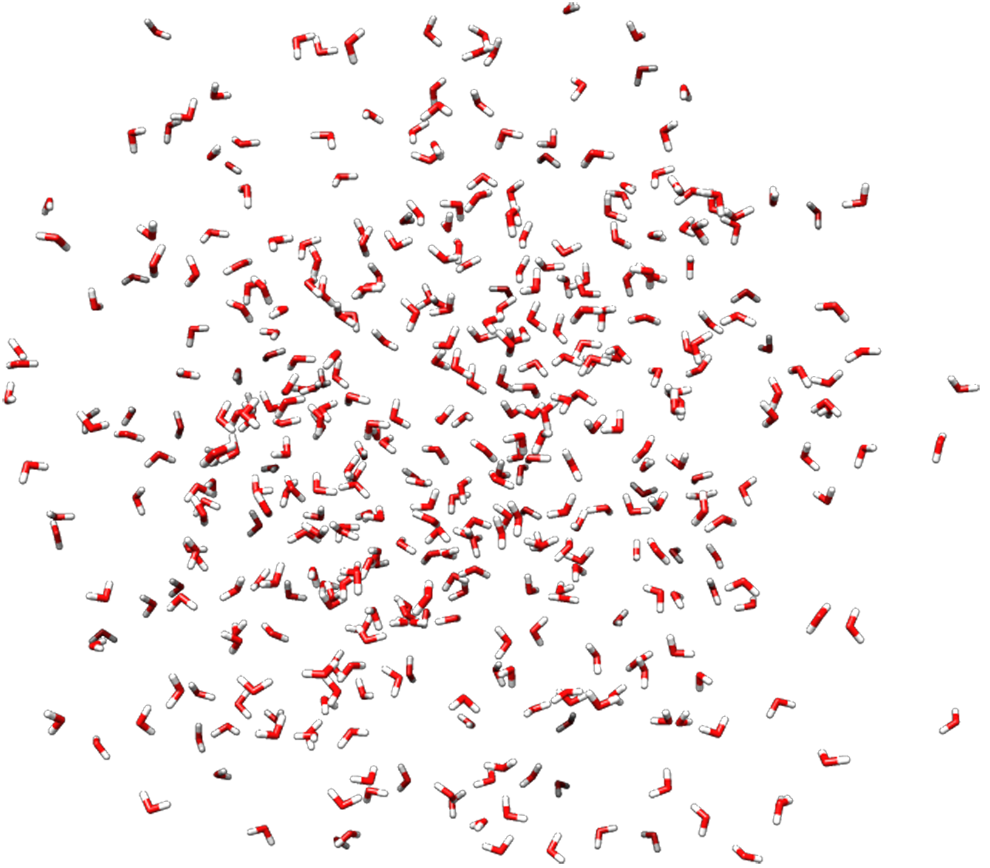
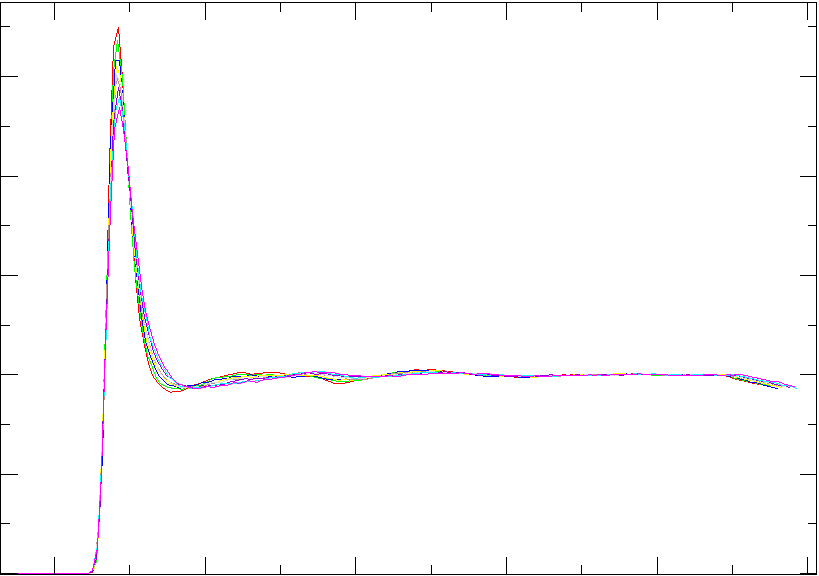
†Department of Mathematical Sciences, Morningside College, 1501 Morningside Ave. Sioux City, Iowa 51106, United States

‡Department of Biology and Chemistry, Azusa Pacific University, 901 E. Alosta Ave. Azusa, California 91702, United States

Abstract

A computational chemistry exercise for the undergraduate physical chemistry laboratory is described. In this exercise students use molecular dynamics to generate trajectories of bulk liquid water. Using a combination of Fortran and Perl scripts to process the trajectory, students calculate structural (radial distribution functions) and dynamic (diffusion coefficients) properties of water that they compare to experimental values. Based on the comparison to experiment, students are also able to draw conclusions regarding the efficacy of TIP3P water at modeling properties of real water.

Graphical Abstract



Keywords

Upper-division Undergraduate, Physical Chemistry, Computer-Based Learning, Computational Chemistry, Molecular Mechanics/Dynamics, Molecular Properties, Water/Water Chemistry

background

Undergraduate students are often exposed to chemical properties through either individual molecules (e.g. dipoles and geometry) or through bulk properties (e.g. boiling point or density) and may have difficulty seeing how the molecular and bulk properties relate to each other. Molecular dynamics (MD) represents a unique opportunity for students to connect these two ideas and see how the behavior of many molecules together, following simple physics laws, can give rise to the bulk properties that are measured.

Because real molecules are quantum mechanical, it is necessary to solve the Schrödinger equation to fully capture their behavior. By using a few approximations, *ab initio* models do so—however, these calculations are limited to very small systems of just a few molecules. In order to actually model systems with enough molecules to reproduce the bulk state, molecules and their dynamics must be treated classically, using Newtonian mechanics. Classical models cannot capture certain effects which rely on electron behavior, such as the breaking of bonds, but they are highly efficient and can produce long simulations with many molecules in a reasonable time frame.

Water is an especially interesting molecule to study, as it has a deceptively simple molecular structure, is vital to many chemical reactions, and demonstrates a hydrogen-bonded network. Because different properties of water are most relevant for different chemical problems, a vast array of computational methods have been developed to achieve different results. Our focus is on the commonly used TIP3P,1 an example of empirical water models, which are parameterized to reproduce given experimental results. Choices must be made when a model is developed, depending on what properties are important for the model to reliably match. A model might, for example, get structural properties such as distances between molecules and orientation of molecules correct, but it might then be unable to reproduce dynamic properties such as diffusion. The accuracy of a model can be then be characterized by comparing its structural, dynamical, and even spectroscopic properties to experimental data or quantum mechanical calculations. Being able to determine when a model is most effective is vital for scientists deciding which water model to use when doing simulations of complex processes such as protein folding,2-6 and interactions of water with semiconductor7-9 or biochemical surfaces.10,11

In this experiment, students will equilibrate bulk liquid water to various temperatures, analysing an example structural property (the radial distribution function) and dynamic property (the diffusion coefficient). By examining these properties under different conditions, students gain insight into the bulk behavior of water as it emerges from the dynamics of many individual molecules. By comparing to experimental results, students gain insight into how well the TIP3P water model works under different conditions.

 Radial Distribution Function

The radial distribution function (RDF) describes the average number of specific atoms in the volume of a sphere with radius *r* from a given molecule’s atom, compared to the liquid’s average density.12 As such, there is an oxygen-oxygen RDF, oxygen-hydrogen RDF, and hydrogen-hydrogen RDF for water. The intensity of peaks in an RDF indicate by how much the local density (for a given solvation shell) exceeds the bulk density. Qualitatively, the RDF can also show how organized the structure is: solids have regular intervals between molecules and hence repetitive peaks in the RDF while gases have random distributions of molecules and no peaks. Liquids have local structure and global disorder, so the RDF shows a few peaks (corresponding to solvation shells) before flattening. Within the liquid phase, the RDF’s first peak shows variation when under high pressure, which allows a significant temperature range; at high temperatures, the minimum between the first and second peaks flattens considerably.13 In order for students to better connect with the states of water being considered, in this lab they work with 1 atmosphere of pressure and temperatures between freezing and boiling.

Diffusion Constant

Water molecules wander from their original positions because of thermal fluctuations, which can cause them to move both away and towards that original position. Eventually, however, at a rate characterized by *D*, the diffusion coefficient, the water molecule will be far from where it started. The diffusion coefficient is a proportionality constant which relates the flux of particles to the concentration gradient. Effectively, particles can move faster through a less dense liquid and in the direction where the liquid is becoming less dense. For long timescales, the diffusion constant can be found using Einstein’s relationship:12

where *r* is the position of the *ith* particle at time *t*. The quantity ri(t) - ri(0) is the difference between the position of molecule *i* at some later time t and that molecule’s original position at time zero. The angle brackets < and > indicate an average over all of the molecules in the simulation at time *t*. Diffusion coefficients also have a simple, proportional relationship with absolute temperature, as predicted by the Stokes-Einstein equation.14

Student learning outcomes

The following are the learning goals for the experiment; by the end of the laboratory students will be able to...

1. Produce MD simulation data of equilibrated water
2. Calculate radial distribution functions of water at different temperatures from MD simulation data
3. Calculate diffusion coefficients of water at different temperatures from MD simulation data
4. Interpret the physical meaning of RDFs and diffusion coefficients
5. Evaluate the ability of a classical water model to recreate experimental properties of water under different temperature conditions

experiment

This experiment consists of three parts: 1) students make their hypotheses for the research questions, based on their reading of the background information in the lab 2) students obtain MD trajectory data for water at their assigned temperatures and 3) students process the MD data in order to calculate the radial distribution functions and diffusion coefficients of liquid water over a range of temperatures.

Research Questions

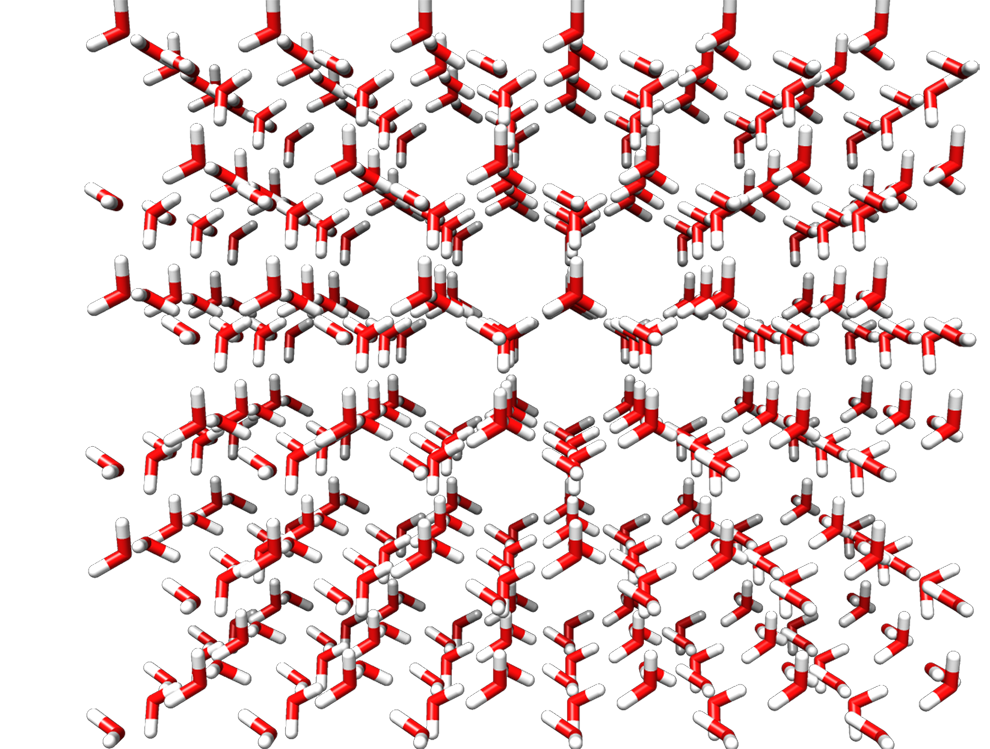
The research questions for this lab are:

1. What is the structure of liquid water?
2. How do the dynamics of liquid water change with temperature?
3. How well does your model of liquid water (and set of assumptions) match experimental results for dynamics and structure?

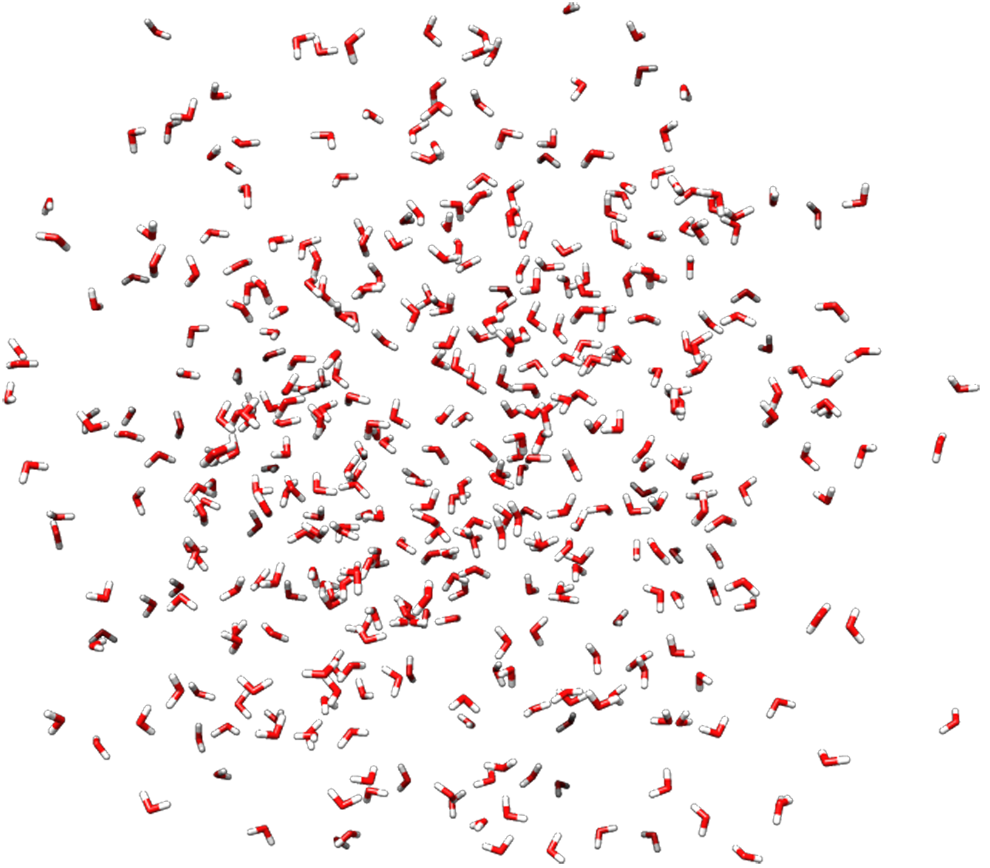
Students answer the first question based on their calculated radial distribution functions (a structural property). They answer the second question based on their calculated diffusion coefficients (a dynamical property). The comparison of both results to experiment will allow them to evaluate the water model in terms of its ability to represent either structure or dynamics of water accurately.

 Process

The molecular dynamics package Amber15 is used to construct a box of TIP3P water, defrost the system to the student’s target temperature (assigned between 280 - 370 K), equilibrate the system, and to obtain the trajectory from which the bulk properties can be calculated. The calculations together have about a 20 minute run time. Optionally, the free molecular visualization program Chimera16 can be used by students to qualitatively see the difference between the original ice structure and their equilibrated system (see Figure 1).



a)



b)

Figure 1. a) Ice-h polymorph of water, showing its hexagonal structure. b) equilibrated box of water at 300 K

The third part of the experiment is the processing of the MD trajectory data to produce the final results. Students use a combination of Perl scripts and Fortran code (provided in the Supplementary Information) to pull the coordinates of the atoms from the Amber trajectory files and calculate the position data needed to find radial distribution functions and the diffusion coefficients. Students determine the position, of the first peak, intensity of the first peak, and position of the first minimum of the oxygen-oxygen RDF at each temperature and determine the diffusion coefficient from the slope of the graph of displacement versus time. They can then compare their results over the whole range of temperatures to see the general trends in structural and dynamic changes. Finally, these values can be compared to experimental results at 300 K.

This laboratory experiment has been performed in a physical chemistry lab for two years and as part of an independent study once. Students worked in pairs to perform the MD simulations for two assigned temperatures. The groups shared their data with their peers, and then wrote full lab reports individually. The lab takes about 2 hours per pair (for two temperatures) on a dual processor PC. If a longer lab session or more computers are available, each pair can perform the simulations for more temperatures.

Several components of the software used in this laboratory can be replaced with alternates as needed: the simulations can be performed with any molecular dynamics package (e.g. CHARMM17 or NAMD18). If an alternative MD package is used, the instructor will need to adapt the provided Perl script (parse\_mdcrd.pl) to be able to process the format of the output of their package of choice. Visualization of the PDB files can be done with many different software packages (e.g. Chimera or VMD19). Graphing the results and the RDFs can likewise be done in a variety of software (e.g. Grace20 or Excel).

 hazards

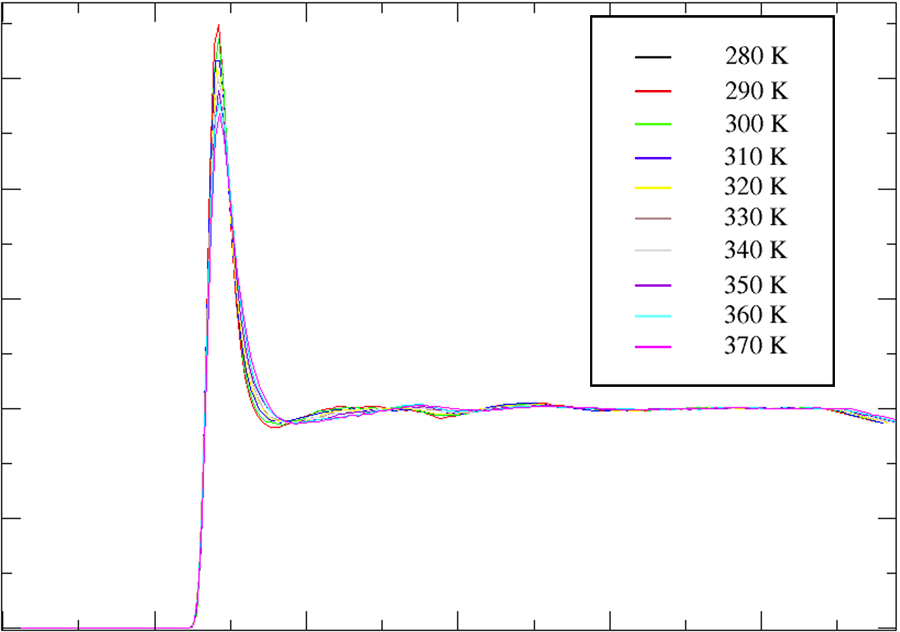
There are no hazards involved in this experiment.

results and discussion

The oxygen-oxygen radial distribution functions for TIP3P water, calculated at temperatures between 280 and 370 K, are shown in Figure 2. Qualitatively, all of the RDFs exhibit the expected intense first peak. The second peak is much more flat, which is characteristic of TIP3P. At 300 K, the position of the first peak is 3% larger than reported values for TIP3P while the intensity and position of the first minimum match reported values.21 TIP3P underestimates the intensity and position of the first peak while overestimating the position of the first minimum, compared to neutron diffraction data.13

Because of differences in pressure conditions, direct comparisons to neutron diffraction data are difficult over our range of temperatures. However, Soper shows that at increasing temperature (and pressure), the local minimum between the first two peaks somewhat flattens.13  Liquid water becomes somewhat less ordered as temperature increases, which is consistent with a lower intensity and broader first peak. Compared to the bulk density, fewer molecules will reside at the particular distance, associated with the first solvation shell, from any given water molecule.

g(r)OO



r (Å)

4

8

2.5

1.0

0

0

0.5

1.5

2.0

Figure 2. Oxygen-oxygen radial distribution function (g(r)OO) for TIP3P liquid water over a range of temperatures.

 Table 1. The First Solvation Shell’s Position and Intensity

|  |  |  |  |
| --- | --- | --- | --- |
| Temperature (K) | Position of First Maximum (Å) | Intensity of First Maximum | Position of First Minimum (Å) |
| 280 | 2.84 | 2.68 | 3.53 |
| 290 | 2.85 | 2.75 | 3.69 |
| 300 | 2.84 | 2.68 | 3.67 |
| 310 | 2.84 | 2.56 | 3.67 |
| 320 | 2.82 | 2.55 | 3.81 |
| 330 | 2.82 | 2.50 | 3.77 |
| 340 | 2.82 | 2.50 | 3.77 |
| 350 | 2.85 | 2.44 | 3.81 |
| 360 | 2.87 | 2.40 | 3.85 |
| 370 | 2.86 | 2.34 | 3.86 |

The calculated self-diffusion coefficients for TIP3P liquid water are shown in Table 2. The slopes are from linear least squares fits of the mean square displacement values (for all temperatures, R2 > 0.99), with graphs shown in Figure 3. As expected, the coefficients generally increase as the temperature increases. The exception to the trend is the coefficient for T = 330 K, which has a higher than expected *D* value. This result is likely a consequence of the relatively short time duration of the MD simulation; Mark and Nilsson show that the slope of the mean square displacement vs time can vary slightly in different regions of a long (200 ps) simulation.21 While the calculated value for the self-diffusion coefficient at 300 K is not consistent with the experimental value of 0.23 Å2/ps,22 it is consistent with other work on TIP3P, which reports a value of 0.567 Å2/ps at 301 K.21 The discrepancy between the calculated and experimental values are an indication of the TIP3P model’s inability to perfectly capture dynamic properties of bulk water. In addition, it’s important to note that the results from the simulation are outputted in terms of timesteps and must be converted to time units. The Amber input files (available in the Supplementary Information) use a timestep of 1 fs.

time (ps)

0

10

20

5

15

<|ri(t) – ri(0)|2> (Å2)

200

100

0

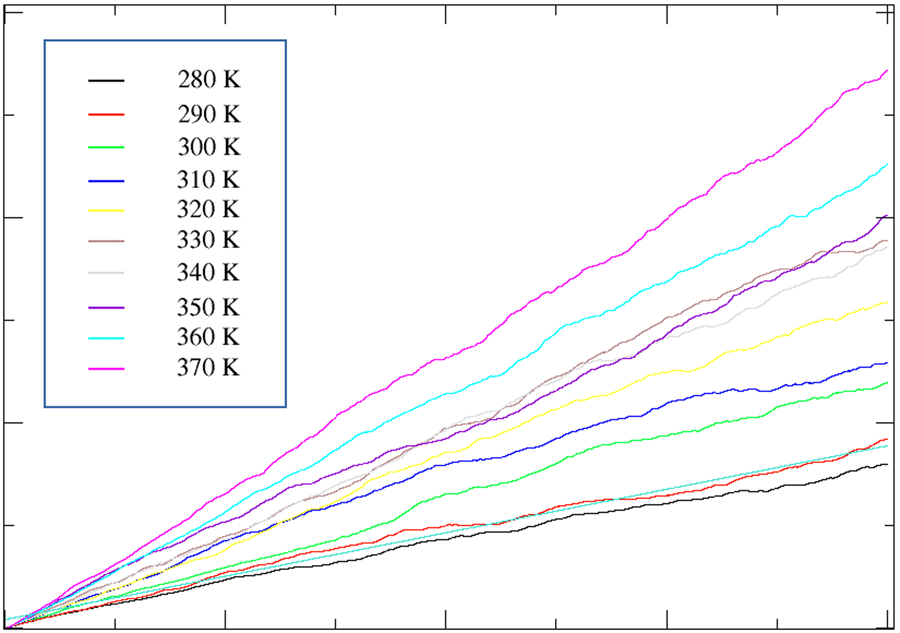


Figure 3. Mean square displacement vs time for 20 ps simulation of 432 waters over a range of temperatures.

 Table 2. Self-Diffusion Coefficients of TIP3P water

|  |  |  |
| --- | --- | --- |
| Temperature (K) | Slope (Å2/timestep) | D (Å2/ps) |
| 280 | 0.00192218 | 0.32 |
| 290 | 0.00211839 | 0.35 |
| 300 | 0.00309296 | 0.52 |
| 310 | 0.0032765 | 0.55 |
| 320 | 0.00409393 | 0.68 |
| 330 | 0.00500756 | 0.83 |
| 340 | 0.00470088 | 0.78 |
| 350 | 0.00475364 | 0.79 |
| 360 | 0.00560399 | 0.93 |
| 370 | 0.00675295 | 1.13 |

CONCLUSIONS

We have developed a computational chemistry laboratory that is appropriate for a physical chemistry course. Students prepare, equilibrate, and run production simulations of bulk liquid water using Amber, then analyze a structural quantity (the oxygen-oxygen radial distribution function) and a dynamical quantity (the diffusion coefficient) over a range of temperatures. These quantities can be compared both to each other, to see trends as temperatures changes, and to experimental values, to evaluate the accuracy of the water model itself.

This laboratory has been successfully implemented for two years in a physical chemistry lab course. During a four hour laboratory, student pairs have been able to complete simulations of two different temperatures. Depending on available time and inclination of the instructor, this lab can be easily modified to produce analysis of either the RDF or diffusion coefficient, or for a smaller range of temperatures. Ambitious students can be encouraged to modify the RDF code to produce analysis of the HH and OH radial distribution functions, which give additional insight into water’s hydrogen bonding.

Associated content

Supporting Information

Lab\_Handout.doc (handout for the lab, including background and procedure)

1\_min.in (input file for the minimization step of the MD run)

2\_defrost.in (input file for the defrost step of the MD run)

3\_equil.in (input file for the equilibrium step of the MD run)

4\_md.in (input file for the production step of the MD run)

diffusion.f90 (Fortran code for the diffusion calculation)

grOO.f90 (Fortran code for the radial distribution function calculation)

ice.pdb (PDB of initial ice structure)

input\_commands.txt (list of commands used in the lab)

parse\_mdcrd.pl (Perl script that reformats the trajectory file from the MD run)

parse\_mdxyz.f90 (Fortran code that extracts the oxygen coordinates from the reformated trajectory)

AUTHOR INFORMATION

Corresponding Author

\*E-mail: cmiller@apu.edu

Acknowledgments

The authors thank Melanie Luallen for her help confirming the results and Rachel Roller for correcting a bug in the parsing script.

REFERENCES

1. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*(2), 926; DOI:10.1063/1.445869
2. Kauzmann, W. *In the Mechanism of Enzyme Action*. The Johns Hopkins Press: Baltimore, Maryland, 1954
3. Jackson, S. How do small single-domain proteins fold? *Folding and Design*, **1998**, *3*(4), R81-91; DOI: 10.1016/S1359-0278(98)00033-9
4. Southall, N.; Dill, K.; Haymet, A. A View of the Hydrophobic Effect. *J. Phys. Chem. B.* **2002**, *106*(3), 521-533; DOI: 10.1021/jp015514e
5. Prabhu, N.; Sharp, K. Protein-Solvent Interactions. *Chem. Rev.* **2006**, *106*(5), 1616-1623; DOI:10.1021/cr040437f
6. Sklenar, H.; Eisenhaber, F.; Poncin, M.; Lavery, R. *Theoretical Biochemistry and Molecular Biophysics*. Adenine Press: New York, 1990
7. Fujishima, A. and Honda; K. Electrochemical Photolysis of Water at a Semiconductor Electrode. *Nature*, **1972**, *238*, 37-38; DOI:10.1038/238037a0
8. Fujishima, A.; Rao T.; Tryk, D. Titanium dioxide photocatalysis. *J. Photochem. and Photobio*. *C*, **2000**, *1*, 1-21; DOI: 10.1016/S1389-5567(00)00002-2
9. Tryk, D.; Fujishima A.; Honda. K. Recent topics in photoelectrochemistry: achievements and future prospects. *Electrochimica Acta*, **2000**, *45*(15), 2363-2376; DOI: 10.1016/S0013-4686(00)00337-6
10. Furse, K. and Corcelli, S. The dynamics of water at DNA interfaces: computational studies of Hoechst 33258 bound to DNA. *J. Am. Chem. Soc*., **2008**, *130*(39), 13103-13109; DOI: 10.1021/ja803728g
11. Nadassy, K.; Wodak, S.; Janin. J. Structural features of protein-nucleic acid recognition sites. *Biochemistry*, **1999**, 38(7), 1999-2017; DOI:10.1021/bi982362d
12. Allen, M. and Tildesley, D. *Computer Simulations of Liquids*. Oxford University Press, Inc.: New York, 1987
13. Soper, A. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 Mpa. *J. Chem. Phys.,* **2000**, *258*, 121-137; DOI: 10.1016/S0301-0104(00)00179-8
14. Einstein. A. Zur Elektrodynamik bewegter Körper *Annalen der Physik* **1905** 322(8), 891-921; DOI: 10.1002/andp.19053221004
15. Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Götz, I. Kolossváry, R. C.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, Q.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. AMBER 12, University of California, San Francisco, 2012
16. Pettersen, E.; Goddard, T.; Huang, C.; Couch, G.; Greenblatt, D.; Meng, E.; Ferrin, T. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comp. Chem.* **2004**, *25*(13), 1605-1612; DOI: 10.1002/jcc.20084
17. Brooks, B.; Brooks III, C. L.; Mackerell Jr., A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: the biomolecular simulation program. *J. Comp. Chem*. **2009**, *30*(10), 1545-1614; DOI: 10.1002/jcc.21287
18. Phillips, J.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comp. Chem.* **2005**, *26*, 802; DOI: 10.1002/jcc.20289
19. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Molec. Graphics*, **1996**, *14*, 33-38; DOI: 10.1016/0263-7855(96)00018-5
20. Grace http://plasma-gate.weizmann.ac.il/Grace/ (accessed May 2017)
21. Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A.* **2001**, *105*, 9954-9960; DOI: 10.1021/jp003020w
22. Mills, R. Self-diffusion in normal and heavy water in the range 1-45.deg. *J. Phys. Chem*. **1973**, *77*, 685-688; DOI: 10.1021/j100624a025