

THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF CHEMICAL ENGINEERING

An Initial Investigation Into the Application of Coarse-Grain Modelling Methods to Solvents in
the Bulk Phase

JOSEPH ZELE
SPRING 2021

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Chemical Engineering
with honors in Chemical Engineering

Reviewed and approved* by the following:

William Noid
Professor of Chemistry
Thesis Supervisor

Wayne Curtis
Professor of Chemical Engineering
Honors Advisor

*Electronic approvals are on file

ABSTRACT

The computational methods which are currently available for performing atomistic molecular dynamics simulations require enormous amounts of computing resources and are thus rendered impractical for application to many systems. Chemical phenomena which occur over large length or time scales are generally beyond the ability of most researchers to simulate with complex systems or systems requiring simulations lasting for more than one hundred nanoseconds taking weeks or even months to complete. An alternative approach of developing coarse-grain models to use in place of the traditional atomistic models is adopted here and investigated for increased efficiency, and accuracy with emphasis on system structure. In particular two homogeneous systems of bulk-phase solvent are investigated using cyclohexane and 1,4-dioxane as the solvents of interest. Conclusions as to the success of the coarse-grain models are developed by comparing the radial distribution functions of the coarse-grain simulations to those of the corresponding atomistic simulations, and by comparing the densities of these simulations to experimental data.

TABLE OF CONTENTS

LIST OF FIGURES	iii
LIST OF TABLES	iv
ACKNOWLEDGEMENTS	v
Chapter 1. Introduction	1
Chapter 2. Theory and Methods	3
2.1 Molecular Dynamics.....	3
2.2 Coarse Grain Models	4
2.3 Methods.....	6
Chapter 3. Simulation Results	8
3.1 Cyclohexane Simulation	8
3.2 1,4-Dioxane Simulation	14
Chapter 4. Conclusion.....	19
Apendix A: Data	21
BIBLIOGRAPHY	23

LIST OF FIGURES

Figure 3.1: Graph of 300 K Cyclohexane Temperature Convergence	8
Figure 3.2: Graph of 300 K Cyclohexane Volume Convergence	9
Figure 3.3: RDF for the 300 K Cyclohexane Simulation	10
Figure 3.4: Potentials Determined for the Cyclohexane Simulations	11
Figure 3.5: Cyclohexane RDFs at all Temperatures	12
Figure 3.6: Potentials Determined for the Dioxane Simulations	15
Figure 3.7: Dioxane RDFs at all Temperatures	16

LIST OF TABLES

Table 3.1: Comparison of Cyclohexane Density Data	13
Table 3.2: Comparison of Cyclohexane Simulation Time	14
Table 3.3: Comparison of Dioxane Density Data	17
Table 3.4: Comparison of Dioxane Simulation Time	18
Table A.1: Temperature Data for Atomistic NVT Equilibration	21
Table A.2: Pressure Data for Atomistic NPT Equilibration	21
Table A.3: Average Temperature and Pressure of Atomistic Production Simulations	21
Table A.4: Coarse-Grain System Volume and Molecular Data	22

ACKNOWLEDGEMENTS

The author would like to specify his gratitude to a number of individuals whose contributions are paramount in the process of writing this paper. The author wishes to thank Dr. Noid both for being an incredible professor, and for being an excellent guide while I was researching and writing this work. Additionally, Dr. Curtis has my gratitude for his enthusiastic assistance at an extremely late point during the thesis writing process, and Dr. Borhan for his assistance throughout my college career. Mr. Ryan Szukalo is due a great deal of gratitude for being patient and providing an incredible amount of assistance during this project whenever it was needed. Lastly, the author would thank all members of the Noid group for their warm welcome and assistance during the short time he was there. The computational results required to complete this work were developed using the programs GROMACS¹⁰, ProDrg¹¹, and BOCS¹² on the Pennsylvania State University's Institute for Computational and Data Sciences' Roar supercomputer.

Chapter 1. Introduction

The capability of scientists to continue advancing fields such as chemistry and biology is closely related to their ability to develop a fundamental understanding of the systems they work with at an increasingly smaller scale. The capability to gather information from systems at a microscopic level is the first step toward understanding these systems. The capability to record this data is not one which many past scientists have had the opportunity to employ. Fortunately, the appearance of the computing age and the advent of computational simulation methods have allowed modern scientists a glimpse into the molecular world. Yet, as of the time of publication for this thesis these methods only allow for a brief view of these systems due to the large amount of time required for even simple simulations over a few hundred nanoseconds (ns). For the atomistic simulations discussed in this paper, 20 ns simulations of 600 to 800 small solvent molecules totaling ~12000 atoms required 3 or more days of computing time on a supercomputing cluster. The systems studied here are generally much smaller than those for which scientists have an interest to simulate.

Larger macromolecules and especially proteins are frequently the objects of interest in computational chemistry of the modern day. A single protein can easily have 1000s of atoms or more. Yet simulations of proteins in isolation provide little useful insight.¹ Most simulations involving proteins also account for their natural environment by surrounding them with solvents, cofactors, and/or substrates which can easily cause the system to develop significant costs in computing power.⁹ Thus any method which would enable simulations to run with less computing power is of value, and could lead to the ability to simulate larger systems without additional advancements in computer capabilities or the dedication of extreme amounts of computing

power. Employing coarse-grained models instead of traditional atomistic models is one such method.

In recent years coarse-grain models have experienced increased popularity as scientists attempt to simulate complicated biological systems. Simulations of loop closing, α -helix formation, and protein folding have used coarse-graining extensively due to the increased computing speed with models such as the UNRES model performing 1000-4000x better than atomistic models during protein folding simulations.⁹ The use of coarse-grain models has even extended to carbohydrates and lipids, and to other processes such as capsid assembly in an HIV-1 system. Even with today's computing infrastructure many of these processes are impractical to simulate with atomistic models.⁹

In the case of this paper single-site coarse-grain models are being investigated for two common solvents, cyclohexane and 1,4-dioxane. These molecules are similar in structure and yet have distinct properties which will serve as a test for the flexibility of the coarse-graining method along with the temperature ranges over which each molecule is being simulated. There are many processes in the chemical production industry where it is common to have dilute solutions in which some sections of the fluid can be approximated to be a homogeneous system of pure solvent. These conditions can arise due to safety concerns when filling storage tanks⁸, the need to maximize mass transfer in membrane separations, or for a variety of other reasons. The ability to accurately represent these molecules in homogeneous systems using coarse-grain models would be the first step toward the application of this method in the simulations of solutions in chemical processes.

Chapter 2. Theory and Methods

2.1 Molecular Dynamics

The theory behind molecular dynamics has existed for over a half-century and yet it is only relatively recent time that this field has experienced widespread application in the analysis of non-simplistic systems. The basic concept is derived from the application of Newton's Laws of motion applied on a molecular scale (See Eqns. 3.1 and 3.2).²

$$\vec{F}_i = m_i \ddot{\vec{r}}_i \quad (3.1)$$

$$\vec{F}_i = - \frac{\partial U(\vec{r}_i)}{\partial \vec{r}_i} \quad (3.2)$$

Where m is the mass of particle i , r is the position of particle i , and F is the force acting of particle i . It also should be noted that Newton's notation for differentiation is employed in Equation 1. Equation 2 specifies that this force acting on particle i is determinable from the interaction potential resulting from the forces being exerted on particle i by the other particles within the system. In this manner the forces exerted by each atom on the other atoms in the system is determined at each time step during the simulation process. This information is then used in order to predict the position and velocities of all of the system's molecules at the next time step. A sufficiently small time step, such as the 1 femtosecond interval commonly employed in molecular dynamics simulations, enables a nearly life-like description of the simulated system with abundant data on how various molecular positional and energetic data vary within to the system. The determination of an accurate interaction potential U is paramount in the success of these simulations.^{1,2}

In order to determine this potential a model known as a force-field must be employed in order to estimate the individual contributions of all other nearby atoms or particles to the force acting on each other. This is done by defining a selection parameter which usually takes the form of a cut-off distance from the atom of interest. This separates the rest of the system into a set of atoms which influence the atom of interest and a set that does not, and is done in order to save computing costs. The force field can then be applied to all atoms within the specified cut-off distance r_c . These force fields are parameterized by developing an empirical model to fit quantum mechanical calculations and frequently, experimental measurements.¹ These models generally take into account the independent contributions of bonds, bond angles, torsional stress, molecular geometry, non-bonded interactions, and electrostatic effects in order to estimate the potentials responsible for the forces acting on each atom (See Eqn. 3.3).⁴

$$\begin{aligned}
 U(r) = & \sum_{bonds} \frac{k_d}{2} (d - d_0)^2 + \sum_{angles} \frac{k_\theta}{2} (\theta - \theta_0)^2 + \sum_{dihedrals} \frac{k_\phi}{2} (1 + \cos(n\phi - \phi_0)) + \\
 & \sum_{impropers} \frac{k_\psi}{2} (\psi - \psi_0)^2 + \sum_{non-bonded\ pairs\ i,j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \\
 & \sum_{non-bonded\ pairs\ i,j} \frac{q_i q_j}{4\epsilon_D r_{ij}} \quad (3.3)
 \end{aligned}$$

The last two elements of Equation 3.3 must then be directly applied for every pair of non-bonded atoms i and j which lay within the cut-off distance, which is usually between 0.8 and 1.4 nm.⁴ This leads to the large computing costs required for these types of simulations with the number of interactions increasing dramatically with any increase in r_c .

2.2 Coarse Grain Models

The idea of coarse-grained models is both direct and elegant in its mimicry of the approach employed elsewhere in classical thermodynamics, and other fields. Using similar

methods to those which lay at the foundation of these fields, coarse-grained models attempt to represent the system using a lower resolution model without incurring significant reductions to the overall accuracy of the results.³ This is analogous to the method of using an approximation of room temperature to do basic thermodynamic calculations. Rather than individually addressing the contributions of each molecule, atom, or sub-atomic particle an average value is instead used to describe the entire area of interest. For coarse-grained models this is done by eliminating atoms as individual objects in the simulation and instead dividing the molecule into sections by grouping them together. A coarse-grain site is then mapped to each of these sections usually by matching the site's center of mass with the center of mass for each section. If done carefully this will result in an object with a similar geometry to the original molecule but fewer individual parts. For a single-site model several elements of equation 3.3 are eliminated with the effective interaction between two coarse-grain sites being approximately represented by equation 3.4.^{3,15}

$$U(r) = \sum_{(i,j)} U_2(r_{ij}) \quad (3.4)$$

Where U_2 is a pair potential dependent on the distance between sites i and j . This greatly simplifies the required calculation and reduces the number of pairs i, j for which the calculation must be done. This then enables an increased ease of simulation for all coarse-grained molecules and thus grants scientists the capability to simulate more complicated systems for longer periods of time.³ Yet, this also introduces another problem. The form and value of U_2 is unknown for coarse-grained models and must be determined using data from atomistic simulations. In this work, the method of multi-scale coarse-graining (MS-CG) will be applied.¹⁵ This method uses data surrounding the average atomistic forces to determine the most optimal potential via the use of a variational principle. This is why an atomistic simulation of the system has to be done first.¹⁵

These models cannot yet be employed for all molecules and the process of applying them remains one which is tedious and can require trial and error. Simulations involving coarse-grained sites are error prone and thus can give results which have little to no value.³ Yet the benefits of employing this methodology are clear and obvious so long as the amount of error accrued is acceptably low, thus prompting the present investigation.

2.3 Methods

The atomistic models which would be the basis of the final coarse grain models had to be developed first. This was done by generating a structural file for a single molecule via ChemDraw which was then imported to ProDrg¹¹ in order to determine the final conformations for the molecules of interest through energy-minimization. These were then put into GROMACS¹⁰ which then generated a positional input file for the simulation. A search of the literature was done to find the needed parameters for the input topology files. The parameters used in this paper for atomistic simulations were taken from Moulton et al.¹³ for cyclohexane and A. O. Yazaydin and R. W. Thompson¹⁴ for 1,4-dioxane. These parameters then allowed for united atom models of both molecules to be run at a set temperature (and pressure in the case of isothermal-isobaric simulations). Once all necessary inputs were generated, topology and positional files were directed to GROMACS¹⁰ to perform the simulation runs. In order to ensure that the system was approximately equilibrated at the desired simulation conditions, two initial 2 ns runs were performed. The first was in the canonical ensemble (NVT) to equilibrate the system temperature, and the second was in the isothermal-isobaric ensemble (NPT) to equilibrate the system pressure. Twenty nanosecond production runs were then performed for each set of conditions.

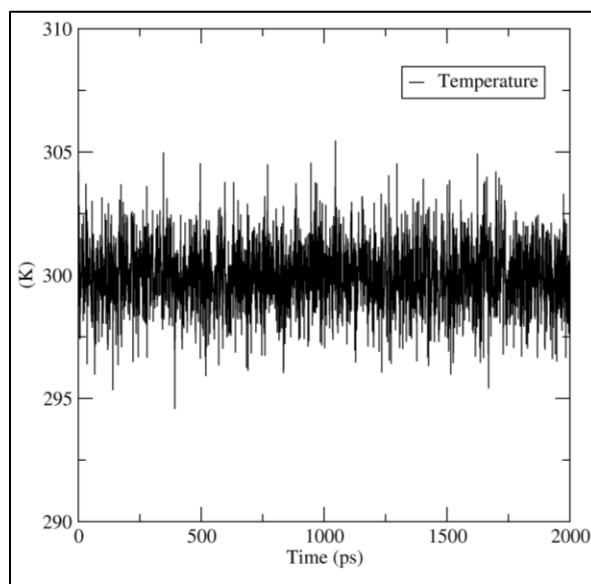
After the completion of these steps for the atomistic system, coarse-grained models were developed for each molecule. This involved taking the position and velocity data from the production runs and mapping all of the atoms in each molecule to a single spherical bead centered at the molecules' center of mass with the correct velocity. The MS-CG method was then employed to determine an average interaction potential for the coarse-grain sites at each set state point. The final step required an adapted version of one of the other GROMACS input files which was generated via a translation program. These steps were all achieved using programs from the BOCS¹² software package. The new coarse-grain models were then simulated in the NVT ensemble at the same state points for which atomistic models were run.

Chapter 3. Simulation Results

3.1 Cyclohexane Simulation

The initial NVT equilibration simulations were completed within 15 hrs with all systems converging to temperatures within .024 K of the desired values (See Table A.1 in Appendix A). A sample graph showing the convergence of the cyclohexane system at 300 K is shown below in Figure 3.1.

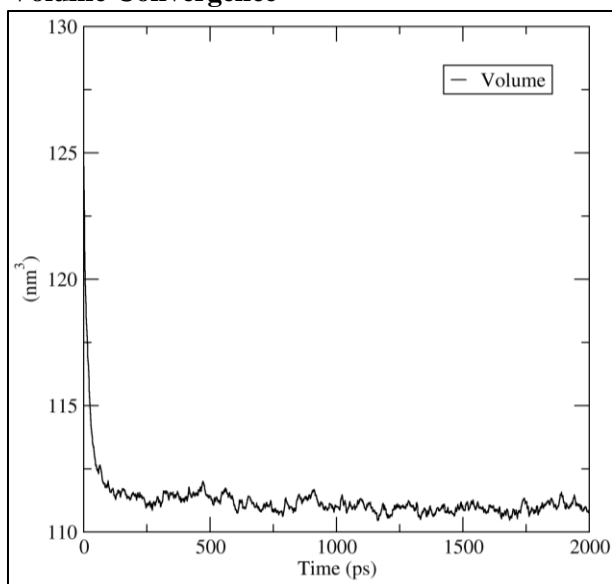
Figure 3.1: Graph of 300 K Cyclohexane Temperature Convergence



The subsequent NPT equilibration simulations also required approximately 15 hrs with all systems converging to within 7 bar of the desired pressure (See Table A.2 in Appendix A). It should be noted that the pressure can deviate greatly during molecular dynamics simulations, and thus either volumetric data or density were preferentially used to confirm the system's convergence to a reasonable state (See Figure 3.2 below). The NPT simulation runs had an

average standard deviation of ~ 400 bar making the equilibration of the system difficult to observe from graphical data.

Figure 3.2: Graph of 300K Cyclohexane Volume Convergence

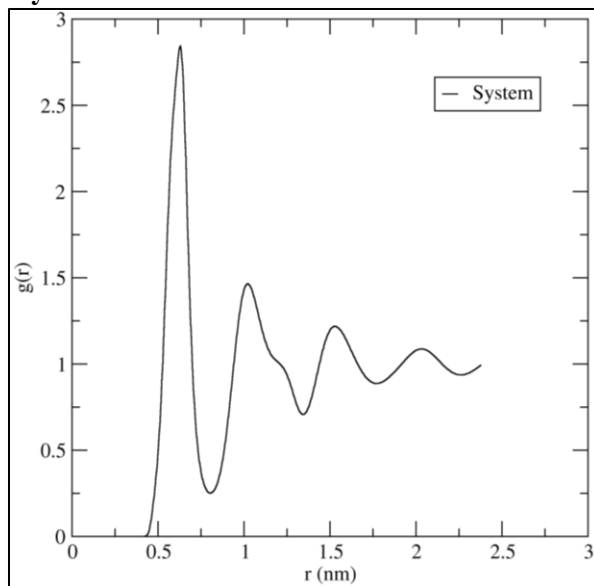


This step is crucial to ensure that the system begins the final simulation at conditions reasonably similar to the specified pressure and temperature. If the system does not it becomes possible for the convergence of the system to the desired state to be disturbed by the presence of a relative minimum among the plethora of conformations the system can inhabit. The worth of such preliminary simulations is in part demonstrated by Figure 3.2. The initial system volume of 125 nm^3 was reduced $\sim 12\%$ to 111 nm^3 . This is a clear indication that the initial system exhibited significantly different properties from the desired simulation conditions.

After all of the cyclohexane simulations at 300, 330, 360, and 390 K finished their equilibration simulations, the resulting trajectory file was then used to perform the final 20 ns production simulation. The completion of this run provided the researcher with the data required to develop and run coarse-grained simulations for each of the previously stated temperatures.

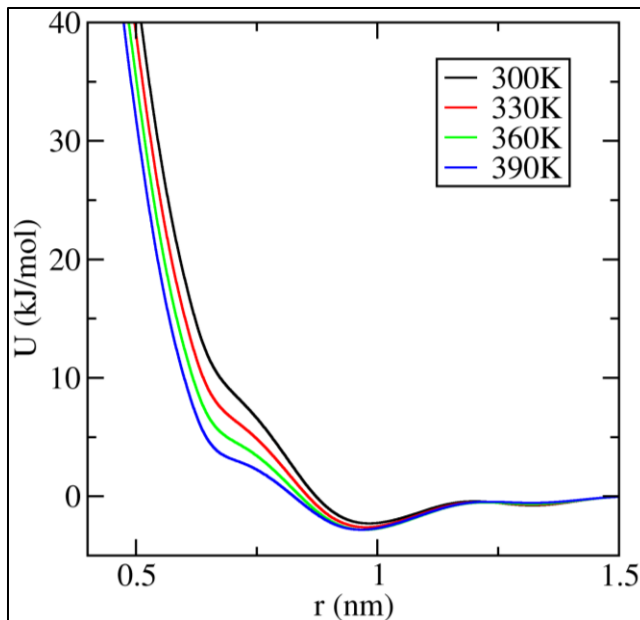
Additionally, this simulation enabled the calculation of the RDF for each of these systems (See Figure 3.3).

Figure 3.3: RDF for the 300 K Atomistic Cyclohexane Simulation



This offers similar information to a probability density function and indicates the likelihood of finding another molecule at a specific distance (radius) from a molecule's center of mass.⁷ After the RDF was calculated, the atomistic model was mapped to a single site coarse-grain model, and the data taken from the atomistic simulation was used to determine the necessary coarse-grain interaction potentials. These potentials were then employed to simulate the system while treating each molecule as a sphere rather than an eighteen-atom system with two atom types. The potentials which were determined for all of the cyclohexane simulations are shown below (Figure 3.4).

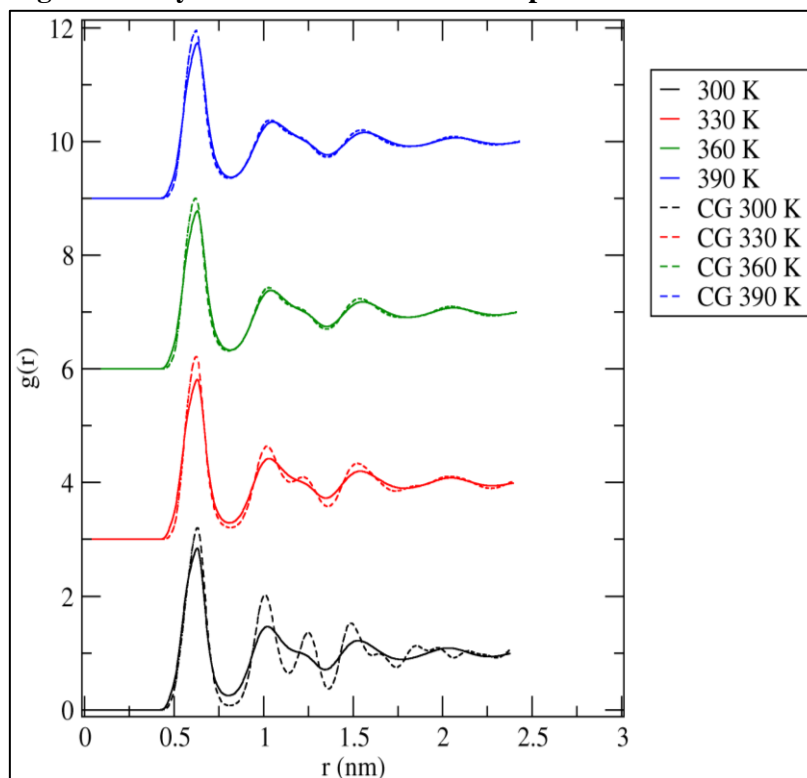
Figure 3.4: Potentials Determined for Cyclohexane Simulations



As shown by the graph the coarse-grain interaction potentials starts at an infinite value and generally decreases until it experiences an asymptote at 0 kJ/mol at some value of r . The value of $U(r = 0)$ approaching infinity indicates the enforcement of volume exclusivity dictating that two molecules cannot share the same location within the system. The horizontal asymptote at $U = 0$ indicates the decay of molecular interactions with distance. The section between these two areas of the graph, approximately 0.5 – 1.5 nm for this system, is the most interesting and informative when addressing the structure of the system. Here the expected trends are shown with all graphs demonstrating a smooth and generally decreasing trend in this area of interest along with distinct minima which is similarly positioned on the x-axis for all temperatures. Additionally, features such as the shoulder at ~ 0.75 nm grow less distinct with increasing temperature and decreasing density which is an expected trend.¹⁶

Finally, coarse-grain simulations were carried out in an effort to test whether or not they were able to replicate the predictions of the atomistic model. This is most easily done by comparing the RDF's generated by both methods of simulation (See Figure 3.5). All RDFs were combined on the same graph and were separated by temperature using an offset of 3 on the y-axis.

Figure 3.5: Cyclohexane RDFs at all Temperatures



Among the most distinct features of these RDFs is the shoulder which occurs at approximately 1.25 nm. All of the simulations display this feature clearly though it becomes less pronounced at higher temperatures. Additionally, another trend is exhibited whereby the peaks of each RDF move slightly further out with each increase in temperature. The coarse-grain RDFs are also very similar to the atomistic RDFs at all temperatures except 300 K, though they consistently overpredict the magnitude of the RDF's oscillation. The fit of the coarse-grain RDFs

also becomes significantly better with the 300 K being the most different and the 390 K simulations resulting in two almost indistinguishable RDFs.

Each of these models were also compared to experimental data to determine their veracity to analogous real systems. First, each model was checked to ensure that the specified temperature was achieved during each simulation (See Table A.3). Then for the purposes of this paper the most convenient property to address is density. A search of the literature returned the values shown below alongside those provided by the simulations (Table 3.1).⁵

Table 3.1: Comparison of Cyclohexane Density Data

System	Literature Value (kg/m ³)	Atomistic Simulation (kg/m ³)	Coarse-Grain Simulation (kg/m ³)
Cyclohexane 300K	773.9 (298.15 K)	852.734	845.905
Cyclohexane 330 K	740.4 (333.15 K)	835.454	834.446
Cyclohexane 360 K	720.1 (353.15 K)	817.924	815.635
Cyclohexane 390 K	677.8 (393.15 K)	799.852	799.838

The literature values provided above are approximate due to a lack of measurements having been taken at the exact simulation conditions though all reported values were recorded within ± 10 K of the stated temperature.⁴ Additionally, the values provided from the atomistic simulations are averages over the entirety of the simulation's run time of 20 ns. The density of the systems from the coarse grain simulations had to be calculated from the known number of molecules, the system volume, and cyclohexane's molecular weight of 84.16 g/mol as they were not reported by GROMACS (See Table A.4 for Coarse-grain system data). The high degree of agreement between both simulations is expected as the volume of the coarse-grain system and the number of molecules it contains were both based on data from the atomistic simulations. The slight differences are a result of the variance in volume which occur when running an isobaric-

isothermal atomistic simulations. There is a significant difference between the experimental values and those returned from the simulation with the simulation overpredicting the density by $\sim 100 \text{ kg/m}^3$ on average. However the accuracy of the simulation density could likely be corrected by finding a new source for the initial atomistic interaction parameters.

It is also notable that the coarse-grained models demonstrated increased computing efficiency as expected, though the magnitude of the increase in computing speed remains astounding. This is shown below in Table 3.2.

Table 3.2: Comparison of Cyclohexane Simulation Time

System	Atomistic Simulation Time (ns/hr)	Coarse-Grain Simulation Time (ns/hr)
Cyclohexane 300 K	0.227	20.833
Cyclohexane 330 K	0.229	20.833
Cyclohexane 360 K	0.228	21.277
Cyclohexane 390 K	0.235	21.144

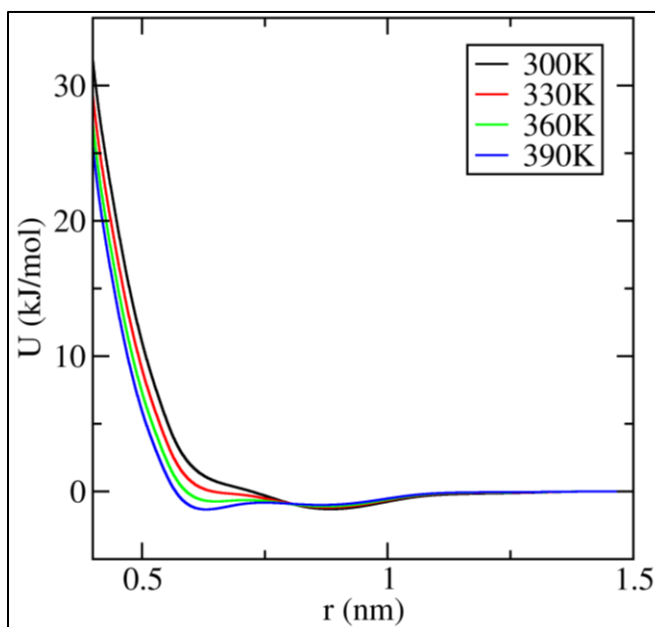
The atomistic simulations were run in parallel on 2 nodes with a total of 40 processor cores while the coarse grain simulations were only run on 1 node with 24 processor cores. This indicates that the difference in computing speed is even larger than it seems. Even ignoring this consideration the computer was able to simulate the coarse-grain model at a rate of approximately 90x the atomistic model's simulation rate. This represents an incredible increase in computing capability and indicates that the application of this model would save a large amount of time compared to the atomistic alternative.

3.2 1,4-Dioxane Simulation

The initial NVT and NPT equilibration simulations for dioxane completed in a manner comparable to the cyclohexane simulations. All temperatures converged to within .051 K and all pressure to within 2.5 bar (See Tables A.1 and A.2 in Appendix A).

The atomistic simulations were then performed in the same manner as the cyclohexane systems. The data recovered from these simulations was then employed to create a coarse-grain model for the system and the system was then simulated again. To this end the potentials shown in Figure 3.6 were determined.

Figure 3.6: Potentials Determined for Dioxane Simulations

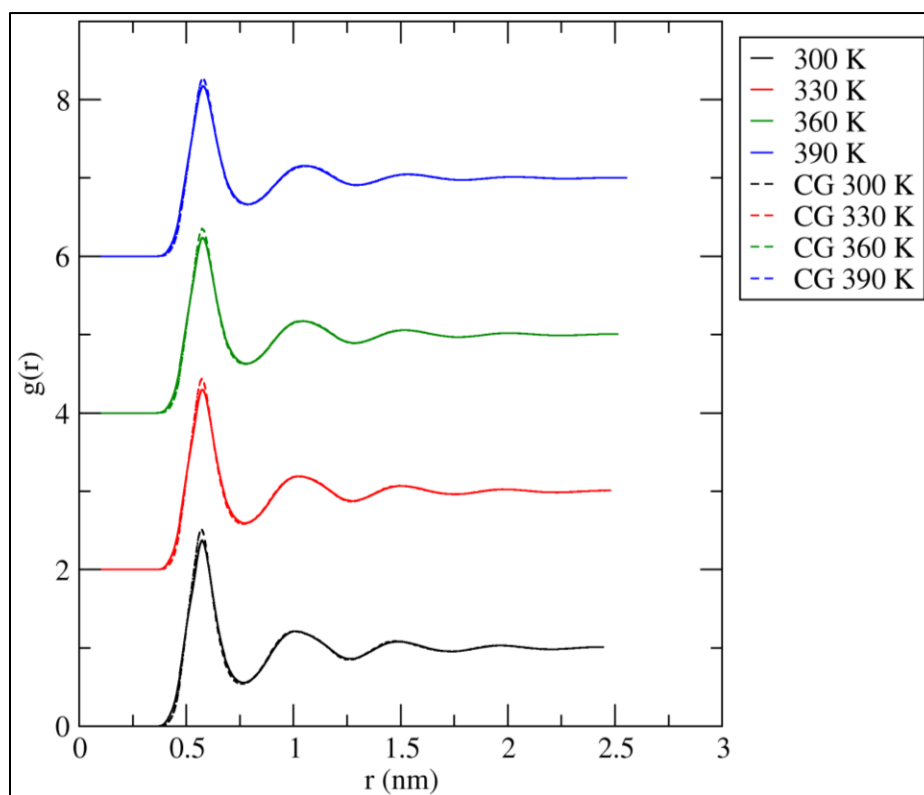


The potential for dioxane exhibited the same features as cyclohexane at both extrema associated with volume exclusivity and the decay of intermolecular forces with distance. Additionally, the same pattern was shown whereby the potential decreases more rapidly with increasing temperature and decreasing density. Yet the minimum is also noticeably closer than on the cyclohexane graphs and the overall shape of the curves is distinct from those for cyclohexane

demonstrating the differences in the intermolecular interactions associated with these two species.

A graph of the overlapping RDF's from the dioxane simulations is shown below (Figure 3.5). For this graph an offset of 2 on the y-axis was sufficient to allow for the data sets to be distinguished.

Figure 3.7: Dioxane RDFs at all Temperatures



Similarly to the potential graph the dioxane RDFs display the same expected trends present in the cyclohexane RDFs. In particular the movement of the graph features outward and the quality of the coarse-grain fit with increasing temperature and decreasing density. Also, the coarse-grain models again tend to overpredict the oscillation of the graphs. However unlike the cyclohexane results, the coarse-grain RDFs fit exceptionally well to the atomistic RDFs at all temperatures.

Each dioxane simulation was checked to ensure it reached the desired temperature (See Table A.3), and then the densities of the simulations were compared with values found through a search of the literature (See Table 3.3).⁶

Table 3.3: Comparison of Dioxane Density Data

System	Literature Value (kg/m ³)	Atomistic Simulation (kg/m ³)	Coarse-Grain Simulation (kg/m ³)
Dioxane 300K	1027.950 (298.15 K)	1010.020	1010.446
Dioxane 330 K	1005.420 (218.15 K)	972.777	971.308
Dioxane 360 K	X	933.931	930.675
Dioxane 390 K	X	892.790	889.191

The literature values which were located were once again found within an approximate range of ± 12 K of the specified temperature.⁶ The Coarse-Grain values were calculated in a similar manner to cyclohexane using 1,4-dioxane's molecular weight as 88.11 g/mol. No literature values were located at any points reasonably close to either 360 or 390 K. Here the densities are almost identical with the simulation values being within ± 20 kg/m³ of the experimental data. This indicates that the parameters used in this simulation were of high quality and can effectively predict system structure at the state points tested.

Once again the coarse-grain simulation proved to result in significant increases in computing efficiency at all conditions (See Table 3.4)

Table 3.4: Comparison of Dioxane Simulation Time

System	Atomistic Simulation Time (ns/hr)	Coarse-Grain Simulation Time (ns/hr)
Dioxane 300 K	0.267	16.949
Dioxane 330 K	0.244	17.543
Dioxane 360 K	0.254	17.543
Dioxane 390 K	0.256	17.857

The coarse-grain models performed marvelously by outperforming the atomistic models by an average of 70x the computing rate while being delegated fewer computing resources. This is significantly more impressive given the accuracy of the results.

Overall both coarse-grain models performed well with all conditions except cyclohexane at 300 K returning RDFs which are very close to the original atomistic graphs for all values of r . Additionally, the dioxane model demonstrated the ability to accurately predict system density within 20 kg/m^3 . The cyclohexane model overpredicted the density significantly, though this can likely be fixed by rerunning the simulations after adjusting the atomistic interaction parameters.

Chapter 4. Conclusion

Coarse-grain modelling is an exciting and innovative approach toward addressing the issue of the vast amount of computing power required for molecular simulation. As demonstrated in this work these models can effectively replicate the outputs of atomistic simulations for cyclohexane and dioxane. For all of the cases studied except for cyclohexane at 300 K, the coarse-grain models fit exceptionally well to the atomistic results. Additionally, all expected trends such as the decay of molecular forces with distance and temperature were clearly demonstrated in the potentials suggesting that the application of the coarse-graining methodology in these cases will produce sensible results. Given the 70 to 90x increased computation speed these models were able to operate at, this methodology most certainly looks promising for application to these and other solvents during the simulation of process mixtures.

For this particular investigation the next steps would focus on addressing the deviations in the density of the cyclohexane simulations, analyzing the energetics of the simulations, and considering the use of multisite coarse-grain models. Another search of the literature should be done to find alternative values for the cyclohexane parameters in an attempt to correct the density of the simulated systems. System properties such as pressure need to be addressed and various methods applied in order to converge the values to be more representative of an analogous real world system. Additionally, each of the two molecules which were mapped to a single site could also be divided into two or 3 sites. This would require the determination of the proper bond angles and interactions between these different sites, but this data is also able to be extracted from atomistic simulations run during this investigation. It is possible that the application

of multisite models could improve the accuracy of the coarse-grain simulations without sacrificing all of the gains in computing efficiency from using a single site model.

Appendix A: Data

Table A.1: Temperature Data for Atomistic NVT Equilibration

Simulation	Average Temperature (K)	Standard Deviation (K)
Cyclohexane 300 K	299.997	1.51889
Cyclohexane 330 K	329.987	1.61365
Cyclohexane 360 K	360.004	1.76585
Cyclohexane 390 K	389.99	1.91522
Dioxane 300 K	299.972	1.53017
Dioxane 330 K	330.051	1.66577
Dioxane 360 K	360.027	1.87245
Dioxane 390 K	390.033	1.97263

Table A.2: Pressure Data for Atomistic NPT Equilibration

Simulation	Average Pressure (bar)	Standard Deviation (bar)
Cyclohexane 300 K	0.280484	406.2
Cyclohexane 330 K	-10.617	431.882
Cyclohexane 360 K	3.53457	470.413
Cyclohexane 390 K	-5.99704	465.781
Dioxane 300 K	-0.634286	421.225
Dioxane 330 K	-2.36519	416.39
Dioxane 360 K	-1.54533	415.068
Dioxane 390 K	.723558	444.678

Table A.3: Average Temperature and Pressure of Atomistic Production Simulations

Simulation	Average Temperature (K)	Average Pressure (bar)
Cyclohexane 300 K	299.998	8.393
Cyclohexane 330 K	330.023	.737
Cyclohexane 360 K	360.019	2.528
Cyclohexane 390 K	389.997	2.961
Dioxane 300 K	299.982	1.342
Dioxane 330 K	329.985	-2.489
Dioxane 360 K	360.046	1.540
Dioxane 390 K	389.991	-3.283

Table A.4: Coarse-Grain System Volume and Molecular Data

Simulation	Total Molecules	System Volume (nm ³)
Cyclohexane 300 K	677	111.849
Cyclohexane 330 K	677	113.385
Cyclohexane 360 K	677	116.000
Cyclohexane 390 K	677	118.291
Dioxane 300 K	834	120.764
Dioxane 330 K	834	125.630
Dioxane 360 K	834	131.115
Dioxane 390 K	834	137.232

BIBLIOGRAPHY

1. Hollingsworth, Scott A, and Ron O Dror. Molecular Dynamics Simulation for All. *Neuron*, Vol. 99, 6 (2018): 1129-1143. DOI:10.1016/j.neuron.2018.08.011.
2. Hug, Susanna, Classical Molecular Dynamics in a Nutshell. In *Biomolecular Simulations*; Monticelli, Luca, Salonen, Emppu; Methods in Molecular Biology Series 924; Springer Science: New York, NY, 2013, pp 127-152.
3. Noid, W. G., Systematic Methods for Structurally Consistent Coarse-Grain Models. In *Biomolecular Simulations*; Monticelli, Luca, Salonen, Emppu; Methods in Molecular Biology Series 924; Springer Science: New York, NY, 2013, pp 487-532.
4. Monticelli, Luca, and Tieleman, D. Peter, Force Fields for Classical Molecular Dynamics. In *Biomolecular Simulations*; Monticelli, Luca, Salonen, Emppu; Methods in Molecular Biology Series 924; Springer Science: New York, NY, 2013, pp 197-214
5. Beg, Shafkat A. et al., Saturated Liquid Densities of Benzene, Cyclohexane, and Hexane from 298.15 K to 473.15 K. *J. Chem. Eng. Data.*, Vol 38, 3 (1993): 461-464. DOI: 10.1021/je00011a035
6. Nain, Anil K. et al., Densities, Refractive Indices, and Excess Properties of Binary Mixtures of 1,4-Dioxane with Benzene, Toluene, o-Xylene, m-Xylene, p-Xylene, and Mesitylene at Temperatures from (288.15 to 318.15) K. *J. Chem. Eng. Data.*, Vol 53, 11 (2008). 2654-2665. DOI: 10.1021/je800579j
7. Martinez-Seara, Hector, and Róg, Tomasz. Molecular Dynamics Simulations of Lipid Bilayers: Simple Recipe of How to Do It. In *Biomolecular Simulations*; Monticelli, Luca, Salonen, Emppu; Methods in Molecular Biology Series 924; Springer Science: New York, NY, 2013, pp 197-214

8. Crowl, Daniel A., and Louvar, Joseph F., *Chemical Process Safety: Fundamentals with Applications*; Pearson: New York, NY, 2019, pp 219-243
9. Singh, Nidhi, and Wenjin Li. "Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications." *International Journal of Molecular Sciences*; Vol. 20,15 3774. 1 Aug. 2019, doi:10.3390/ijms20153774
10. M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, and E. Lindahl, (2015). "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, 1-2, 19-25.
11. A. W. Schuettelkopf and D. M. F. van Aalten (2004). PRODRG – a tool for high throughput crystallography of protein-ligand complexes. *Acta Crystallographica*. D60,1355-1363
12. Dunn, NJH; Lebold, KM; DeLyser, MR; Rudzinski, JF; Noid, WG. (2017). BOCS: Bottom-Up Open-Source Coarse-Graining Software. *J. Phys. Chem. B*. 122, 13, 3363-3377.
13. Moulton, Orthonas A. et al. (2016), Atomistic Molecular Dynamics Simulations of Carbon Dioxide Diffusivity in n-Hexane, n-Decane, n-Hexadecane, Cyclohexane, and Squalane. *J. Phys. Chem. B*. 120, 12890-12900
14. A. O. Yazaydin and R. W. Thompson (2006) Simulating the Vapour–Liquid Equilibria of 1,4-Dioxane, *Molecular Simulation*; 32:8, 657-662, DOI: 10.1080/08927020600883277
15. Noid, W G et al. "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models." *The Journal of Chemical Physics* Vol. 128, 24 (2008): 244114. DOI:10.1063/1.2938860

16. Lebold, Kathryn M. and Noid, W. G. "Systematic Study of Temperature and Density Variations in Effective Potentials for Coarse-grained Models of Molecular Liquids." *J Chem. Phys.* Vol. 150, (2009): 014104. DOI: 10.1063/1.5050509

Education

Pennsylvania State University August 2017 – May 2021
Bachelor of Science in Chemical Engineering

Relevant Courses

- Molecular Thermodynamics (Chem 466)
- Introduction to Chemical Engineering Thermodynamics (CHE 220)
- Organic Chemistry I & II (Chem 210 & 212)
- Experimental Methods (STAT 401)

Skills

Familiar with Linux OS, R, Mathematica

Research

Molecular Dynamics and Coarse-Grain Modeling

- Learned about Modeling Methods and Theory
- Performed Atomistic and Coarse-Grain Simulations

Notable Honors/ Awards

Eagle Scout June 2017
• National Eagle Scout Association (NESA) Member
Schreyer Honors College May 2021
