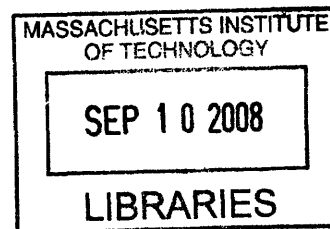


**A Novel Algorithm for Creating Density Dependent, Coarse-Grained  
Models for the Simulation of Surfactant Systems**

by

Erik Christian Allen

B.S., Iowa State University (2002)



Submitted to the Department of Chemical Engineering  
in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 2008

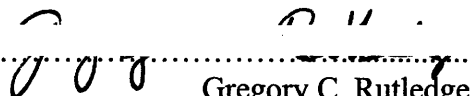
©2008 Erik Christian Allen. All rights reserved.

The author hereby grants to MIT permission to reproduce  
and to distribute publicly paper and electronic  
copies of this thesis document in whole or in part  
in any medium now known or hereafter created.

Signature of Author.....

Erik Christian Allen  
Department of Chemical Engineering  
August 28, 2008

Certified By.....

  
Gregory C. Rutledge  
Professor of Chemical Engineering  
Thesis Supervisor

Accepted By.....

William M. Deen  
Professor of Chemical Engineering  
Chairman, Committee for Graduate Students

# A Novel Algorithm for Creating Density Dependent, Coarse-Grained Models for the Simulation of Surfactant Systems

by

Erik Christian Allen

Submitted to the Department of Chemical Engineering  
on August 4, 2008, in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

## Abstract

Large-scale simulations of solvated molecules that treat the solvent explicitly are very computationally expensive, and as a result work has been done on modifying the potentials to treat solvent implicitly. Implicit solvation is well-known in Brownian Dynamics of dilute solutions, but offers promise to speed up many other types of molecular simulations as well, including studies of proteins and colloids where the local density can vary considerably. This work examines implicit solvent potentials within a more general coarse-graining framework. While a pairwise potential between solute sites is relatively simple and ubiquitous, an additional parameterization based on the local solute concentration has the possibility to increase the accuracy of the simulations with only a marginal increase in computational cost. In this thesis we describe a method in which the radial distribution function (RDF) and excess chemical potential of solute insertion ( $\mu^{ex}$ ) for a system of Lennard-Jones particles are first measured in a fully explicit, all-particle simulation, and then reproduced across a range of solute particle densities in an implicit solvent simulation. The resulting potentials are density-dependent, implicit solvent (DDIS) potentials.

We then test the transferability of DDIS potentials to mixtures and systems of chains without additional optimization. We find that RDF transferability to mixtures is very good and RDF errors in systems of chains increase linearly with chain length. Excess chemical potential transferability is good for mixtures at low solute concentration, chains, and chains of mixed composition; at higher solute concentrations in mixtures, chemical potential transferability fails due to the unique property of DDIS potentials that inserting a single particle changes the densities of all neighboring particles. Based these results, we demonstrate that DDIS potentials derived for pure solutes can be used effectively in

the study of many important systems including those involving mixtures, chains and chains of mixed composition.

Finally, the DDIS potentials are used to examine the self-assembly of a model surfactant system. We demonstrate that the coarse-grained DDIS potentials generated by this method accurately predict the trends in critical micelle concentration (CMC) for two surfactant types, but that the absolute values of the predicted CMC are an order of magnitude higher than previously established estimates for the same surfactants using atomistic simulations. Additionally, the micelles formed are less densely packed than the corresponding all-atom micelles, leading to a larger average aggregation number. By examining a series of simulations of increasing molecular complexity, we identify the source of this error with the transferability of the DDIS potentials. The results suggest that deriving the DDIS potentials directly from simulations of chain molecules in solvent could improve the ability of such potentials to reproduce surfactant properties accurately.

The method for deriving DDIS potentials is extremely general and can be applied to study a variety of solvated systems of any chemical complexity. For example, the results of this work can be extended to study problems in protein folding, drug uptake in micellar systems, and biological membranes.

Thesis Supervisor: Gregory C. Rutledge

Title: Lamot du Pont Professor Professor of Chemical Engineering

Thesis Supervisor: Kenneth J. Beers

Title: St. Laurent Career Development Visiting Assistant Professor of Chemical Engineering

*For my parents, Tom and Julie*

## Acknowledgements

*The only real valuable thing is intuition.*

-Albert Einstein

*A man's errors are his portals of discovery.*

-James Joyce

Although the work of completing this thesis was an individual endeavor, it would not have happened without the truly amazing assistance and support of those closest to me. This brief thank you doesn't fully reflect their impact, but it's a start.

I'd like to first thank my advisors, Ken Beers and Greg Rutledge. My time at MIT can be split into two phases, and I feel extremely fortunate to have received exactly the right type of guidance at exactly the right time. From Professor Beers, I learned to view the whole world as an interesting problem, to be attacked with vigor. Who knew that Quantum Chemistry, sailboat tacking, and Sudoku could all be solved with Matlab? From Professor Rutledge Rutledge, I learned how to narrow the search from the whole world to that problem on which I could make the biggest impact. From him I learned how to answer a question thoroughly and completely.

I'd also like to thank members of the Rutledge and Beers research labs, especially Kirill Titievsky, Brian Stephenson, Vikram Kuppa, Ahmed Ismail, Miguel Amat, Minglin Ma, and Fred Bernardin. These individuals offered technical knowledge and moral support.

Thanks also to my thesis committee, Professors Robert Cohen and Alan Hatton, who provided invaluable insights. They stayed with me through project changes and uncertainty, and guided me expertly toward the final product you see here.

I owe a debt of gratitude to the Department of Energy's Computational Science Graduate Fellowship (CSGF). In addition to providing the funding that allowed me to complete my thesis research, the CSGF program put me in contact with a large community of like-minded peers in both academia and at the DOE. I credit the program with expanding my understanding and appreciation of all computational tools. It has been a tremendous opportunity.

To my parents: they aren't scientists by trade, but you wouldn't know it from the analytical way they approach the world. My parents taught me to question everything, assume nothing, and always follow your intuition. They were always there for me, ready to listen with open mind. This work couldn't have happened without them.

To my brother and sister-in-law: Jeff and Alison, you guys are supposed to be the younger siblings, but I have learned so much from you about the importance of

friendship, rocking, and taking the path that will make you happiest, regardless of convention.

To Beez, RDB, Tmon, Joel, Pieffer, Jeff, and Luke: long live the hedge!

To all my friends, from all phases of life: One of the biggest lessons I learned while in graduate school is that it's all about friendship. Thank you for sustaining me in good times and bad, and I hope I did the same for you. To Ben, RDB, Tyler, Joel, Kris, Jane, Jake, Sharon, Kevin, Joe, Wally, Luke, Bryan, Jim, Simon, Carrie, and Marcie – Thanks!

Last, but not least, I want to thank Jenn. It's hard to remember now, but there was a time not too long ago when I was walking alone. I'm so happy that I will never say that again. Your smiling face brings endless joy, your intelligence challenges me daily, and your humor...well, your intelligence is still good. J/K, J/K for real.

# Table of Contents

ABSTRACT .....	2
ACKNOWLEDGEMENTS.....	5
TABLE OF CONTENTS .....	7
LIST OF FIGURES .....	9
LIST OF TABLES .....	12
<b>1 INTRODUCTION.....</b>	<b>13</b>
1.1 MOTIVATION AND OBJECTIVES.....	13
1.2 BACKGROUND.....	15
1.2.1 <i>Coarse-Graining</i> .....	15
1.2.2 <i>Transferability of Coarse-Grained Potentials</i> .....	18
1.2.3 <i>Implicit Solvent Models</i> .....	20
1.2.4 <i>Surfactant Solutions: Theory and Simulations</i> .....	22
1.3 OUTLINE OF A NEW COARSE-GRAINING APPROACH .....	24
1.4 THESIS OVERVIEW.....	26
1.5 REFERENCES .....	27
<b>2 A NOVEL ALGORITHM FOR CREATING COARSE-GRAINED, DENSITY DEPENDENT IMPLICIT SOLVENT MODELS .....</b>	<b>33</b>
2.1 INTRODUCTION .....	33
2.2 CONSERVED PROPERTIES .....	36
2.3 COARSE-GRAINED POTENTIAL FORM .....	37
2.4 POTENTIAL PARAMETERIZATION .....	41
2.4.1 <i>Metrics</i> .....	41
2.4.2 <i>Global Density Dependent Potentials</i> .....	42
2.4.3 <i>Local Density Dependent Potentials</i> .....	45
2.5 SIMULATION PROTOCOL.....	51
2.5.1 <i>All Atom Simulations</i> .....	51
2.5.2 <i>Global Density Dependent Simulations</i> .....	52
2.5.3 <i>Local Density Dependent Simulations</i> .....	52
2.5.4 <i>External Field Simulations</i> .....	53
2.6 RESULTS AND DISCUSSION.....	53
2.6.1 <i>All-atom Simulation</i> .....	53
2.6.2 <i>Global Density Dependent Potentials</i> .....	54
2.6.3 <i>Pairwise potential transferability to other state points</i> .....	56
2.6.4 <i>Local Density Dependent Potentials</i> .....	58
2.6.5 <i>External Field Simulations</i> .....	60
2.7 CONCLUSION.....	62
2.8 REFERENCES .....	64
<b>3 EVALUATING THE TRANSFERABILITY OF COARSE-GRAINED, DENSITY DEPENDENT IMPLICIT SOLVENT MODELS TO MIXTURES AND CHAINS .....</b>	<b>81</b>
3.1 INTRODUCTION .....	81
3.2 THEORY .....	84
3.2.1 <i>Density Notation Conventions</i> .....	84
3.2.2 <i>DDIS Potential Review</i> .....	84
3.2.3 <i>Mixing Rule</i> .....	86
3.2.4 <i>Limits of Transferability</i> .....	87
3.2.5 <i>Special Cases</i> .....	89

3.2.6	<i>Solute Enhancement Ratio</i> .....	91
3.3	SIMULATION PROTOCOL.....	92
3.4	RESULTS AND DISCUSSION.....	93
3.4.1	<i>Particle Types Used in this chapter</i> .....	93
3.4.2	<i>DDIS Potentials</i> .....	95
3.4.3	<i>Mixtures</i> .....	97
3.4.4	<i>Chains</i> .....	98
3.4.5	<i>Chains of Mixed Composition</i> .....	101
3.5	CONCLUSION.....	102
3.6	APPENDIX.....	103
3.7	REFERENCES.....	105
<b>4</b>	<b>SIMULATION OF SURFACTANT SYSTEMS USING COARSE-GRAINED, DENSITY DEPENDENT IMPLICIT SOLVENT MODELS</b> .....	<b>124</b>
4.1	INTRODUCTION.....	124
4.2	THEORY.....	127
4.2.1	<i>Density Notation Conventions</i> .....	127
4.2.2	<i>DDIS Potential Review</i> .....	128
4.2.3	<i>Simulation Metrics</i> .....	129
4.2.4	<i>Surfactant Theory</i> .....	131
4.3	SIMULATION PROTOCOL.....	133
4.3.1	<i>All-Atom Model</i> .....	133
4.3.2	<i>Simulation Details</i> .....	135
4.4	RESULTS AND DISCUSSION.....	136
4.4.1	<i>DDIS Potentials</i> .....	136
4.4.2	<i>Surfactant Results</i> .....	138
4.4.3	<i>Analysis of Transferability</i> .....	140
4.5	CONCLUSIONS.....	143
4.6	REFERENCES.....	146
<b>5</b>	<b>CONCLUSIONS AND FUTURE DIRECTIONS</b> .....	<b>161</b>
5.1	CONCLUSIONS.....	161
5.2	FUTURE RESEARCH.....	166
5.2.1	<i>Methodological Improvements</i> .....	166
5.2.2	<i>Applications</i> .....	169
<b>APPENDIX A AB INITIO STUDY OF THE BINDING STRENGTH OF POSS-CATION COMPLEXES</b>		<b>171</b>
A.1	INTRODUCTION.....	171
A.2	METHODOLOGY.....	173
A.3	RESULTS AND DISCUSSION.....	174
A.3.1	<i>Addition of cyclohexyl side chains</i> .....	176
A.3.2	<i>Presence of a Cl<sup>-</sup> counterion</i> .....	176
A.3.3	<i>Replace Li<sup>+</sup> with CH<sub>3</sub>NH<sub>3</sub><sup>+</sup></i> .....	177
A.3.4	<i>Feasibility of Lithium transfer between two POSS molecules</i> .....	177
A.4	SUMMARY.....	178
A.5	REFERENCES.....	179
<b>APPENDIX B ALGORITHMS</b> .....		<b>187</b>
B.1	INTRODUCTION.....	187
B.2	CALCULATE LOCAL DENSITY.....	187
B.3	CALCULATE SYSTEM ENERGY.....	188



## List of Figures

- Figure 1-1: Flow Diagram for our modified coarse-graining technique. Multiple feedback loops allow for rigorous iterative improvement at all steps of the process. .... 32
- Figure 2-1: Comparison of free energy changes upon particle removal in the (a) all-particle and (b) implicit solvent, density-dependent cases. The density-dependent potential introduces a secondary free energy change due to the change in energy models associated with a change in global average solute density of the system.... 66
- Figure 2-2: Comparison of the RDF generated using the converged potential for  $X_S=0.1$  (open circles) and the target RDF (solid line)..... 67
- Figure 2-3: Pairwise potentials that reproduce the target RDF across a range of global solute densities,  $\rho_G$ . Bare Lennard-Jones potential shown for comparison as dark solid line (a) solute densities below the total particle density (—,  $X_S=0.10$ ;  $\cdots$ ,  $X_S=0.40$ ; ---,  $X_S=0.70$ ; -·-·,  $X_S=0.90$ ) (b) solute densities above the total particle density (—,  $X_S=1.10$ ;  $\cdots$ ,  $X_S=1.20$ ) ..... 68
- Figure 2-4: Self-interaction energy,  $\mu(\rho_G)$  for  $\rho^*=0.55, T^*=1.35$  ..... 69
- Figure 2-5: Comparison of measured (open circles) and target (solid line) radial distribution functions at  $X_S=0.10$  for reduced temperatures (a)  $T^*=0.83$  (b)  $T^*=1.08$  (c)  $T^*=1.58$  (d)  $T^*=1.83$ . (a) and (b) are below the fitting temperature, (c) and (d) above..... 70
- Figure 2-6: Comparison of measured (open circles) and target (solid line) radial distribution functions at  $X_S=0.10$  for system volumes (a)  $V/V_0=0.90$  (b)  $V/V_0=0.95$  (c)  $V/V_0=1.05$  (d)  $V/V_0=1.10$ . (a) and (b) are more dense than the fitting density, (c) and (d) less dense. .... 71
- Figure 2-7: Excess chemical potential as a function of system temperature. (Circles, all-particle simulation results; squares,  $X_S=1.00$ ; crosses,  $X_S=0.10$ ) ..... 72
- Figure 2-8: Comparison of simulated  $\mu^{ex}$  values with target value as a function of  $\rho_G$  (Circles, original simulation using global density values; squares, after optimization; solid line, all-particle simulation; dashed lines, desired simulation accuracy)..... 73
- Figure 2-9: Comparison of simulated RDF with worst fit ( $X_S=1.00$ ) (open circles) and the all-particle RDF (solid line). .... 74
- Figure 2-10: Comparison of simulated  $\mu^{ex}$  values with the exact value for the all-particle simulation as a function of  $\rho_G$  for  $r_D=2\sigma$  (Circles, simulation data; solid line, all-particle simulation; dashed lines indicate an error of  $0.06\epsilon$ )..... 75
- Figure 2-11: Solute fraction as a function of distance from center of the box (—, all-atom; ---, local density approach;  $\cdots$ , global density approach)..... 76
- Figure 3-1: Comparison of free energy changes upon particle insertion in the (a) all-particle and (b) density-dependent, implicit solvent cases. The density-dependent potential introduces a secondary free energy change due to the change in energy models associated with a change in global average solute density of the system... 108
- Figure 3-2: Comparison of excess chemical potential in all-atom (solid line) and coarse-grained (circles) simulations. Left: B-type particles in solvent Z. Right: C-type

particles in solvent Z. The dashed lines demarcate errors of $0.06\epsilon_Z$ , and are provided as a guide to compare the relative errors between the two coarse-grained potentials.....	109
Figure 3-3: Comparison of worst-fit solute RDF in all-atom (solid line) and coarse-grained (circles) simulations. Left: B-type particles in solvent Z. Right: C-type particles in solvent Z.....	110
Figure 3-4: Coarse-grained two-body term for local solute density $\rho_s^L/\rho=0.5$ (—A,—B,···C).....	111
Figure 3-5: Coarse-grained one-body term as a function of local solute density (Circles: A-type particles, Squares: B-type particles, Crosses: C-type particles).....	112
Figure 3-6: Worst case RDF reproduction for mixtures. Left Side: A/B/Z mixture, Right Side: A/C/Z mixture (Solid Line – all-atom results; Circles –DDIS results). .....	113
Figure 3-7: Comparison of coarse-grained and all-atom excess chemical potential for mixtures, broken down by total solute density, $\rho_s$ . Left Side: $\mu_B^{ex}(\rho_A,\rho_B)$ in A/B/Z mixtures. Right Side: $\mu_C^{ex}(\rho_A,\rho_C)$ in A/C/Z mixtures. (Solid Line – All-atom values; Circles – $\rho_s/\rho=0.2$ ; Squares – $\rho_s/\rho=0.5$ ; Crosses – $\rho_s/\rho=0.9$ ). .....	114
Figure 3-8: Worst case RDF reproduction for tetramers of solute particles (a) A-type, (b) B-type, (c) C-type. (Solid Line – All-atom; Circles – DDIS results).....	115
Figure 3-9: Comparison of coarse-grained and all-atom excess chemical potential for monomers, dimers, and tetramers as a function of solute density, $\rho_s$ . (a) A-type, (b) B-type, (c) C-type. (Solid Line – All-atom; Circles – Monomers; Squares – Dimers; Crosses – Tetramers) .....	116
Figure 3-10: $E_L$ standard error (units of $\epsilon_Z$ ) as a function of chain length for local DDIS and global density-dependent potentials. (Circles – Local potential; Squares – Global potential).....	117
Figure 3-11: Comparison of C-C particle RDFs for all-atom, local potential, and global potential for C-type tetramers in solvent Z at $\rho_C/\rho=0.2$ (Solid Line – all-atom; Circles – global potential; Squares – local potential).....	118
Figure 3-12: All-atom (solid line) and coarse-grained (circles) $\mu^{ex}(\rho_s)$ for $A_1B_1$ molecule. ....	119
Figure 3-13: Comparison of measured $\mu_{ex}^B(\rho_A,\rho_B)$ transferability with and without mixture correction function for $\rho_s/\rho=0.9$ . Results indicate that parameterizing the self-interaction energy on the density of both particle types improves transferability, and that allowing the pure-component values to vary improves transferability even more (Straight line – All-atom values; Squares – original mixing rule; Circles – modified mixing rule). .....	120
Figure 4-1: Comparison of excess chemical potential in all-atom (solid line) and coarse-grained (circles) simulations. Left: H-type particles in solvent W. Right: T-type particles in solvent W. ....	151
Figure 4-2: Comparison of worst-case fits of solute RDF in all-atom (solid line) and coarse-grained (circles) simulations. Left: H-type particles in solvent W at $\rho_H=0.2$ . Right: T-type particles in solvent W at $\rho_T=0.1$ . ....	152
Figure 4-3: Comparison of RDFs in all-atom (solid line) and coarse-grained (circles) simulations. (a): $\rho_T/\rho=0.1$ . (b): $\rho_T/\rho=0.5$ (c): $\rho_T/\rho=1.0$ . ....	153

Figure 4-4: Coarse-grained two-body term for local solute density $\rho_s^L/\rho=0.2$ (—H,--T). The dark line shows the T-type Lennard-Jones interaction for comparison.....	154
Figure 4-5: Coarse-grained one-body term as a function of local solute density (Circles: H, Squares: T) .....	155
Figure 4-6: Free surfactant density (circles) as a function of total surfactant density. Solid line is the 45° line corresponding to a condition of no micellar aggregates. Left: $H_1T_4$ ; Right: $H_1T_5$ .....	156
Figure 4-7: Simulation snapshot of 216 $H_1T_4$ surfactant molecules in implicit solvent showing the formation of micellar aggregates. H-type particles are black, T-type particles are grey. ....	157
Figure 4-8: Micelle aggregation number distribution for $H_1T_4$ surfactants. ....	158
Figure 4-9: Comparison of all-atom (solid lines) and implicit solvent (circles) RDFs in a 1:4 H:T mixture at $\rho_s/\rho=0.3$ . (a): H-H RDF; (b) H-T RDF; (c) T-T RDF. ....	159
Figure 4-10: Comparison of all-atom (solid lines) and implicit solvent (circles) RDFs for a single micelle composed of 20 $H_1T_4$ surfactants in 3418 solvent molecules W. (a): H-H RDF; (b) H-T RDF; (c) T-T RDF.....	160
Figure A-1: Structural similarity of the POSS molecule to crown ether. (a) POSS (b) Crown ether.....	182
Figure A-2: GAUSSIAN98 output – POSS- $Li^+$ Complex .....	183
Figure A-3: POSS/ $Li^+$ Energy as a function of lithium position. The energy minimum is achieved approximately 1 angstrom away from the POSS cage face .....	184
Figure A-4: Predicted IR Spectrum of POSS (solid) and POSS/ $Li^+$ (dotted). DFT calculations predict the presence of an additional IR peak at $\sim 1000cm^{-1}$ .....	185
Figure A-5: Structure of two POSS molecules joined by attractive force of a single lithium ion. Results predicted by energy minimization in GAUSSIAN98.....	186

## List of Tables

Table 2-1: Test system properties. ....	77
Table 2-2: Transferability of the original effective potential, as measured by the error in the per-particle energy with changing temperature. ....	78
Table 2-3: Transferability of the original effective potential, as measured by the error in the per-particle energy with changing density. ....	79
Table 2-4: Errors produced by use of a local density dependent potential. For the case $r_D = r_c = 5\sigma$ , errors are below the specified tolerance of 1%. However, a smaller cutoff radius $r_D = 2\sigma$ introduces significant errors in the per-particle energy. ....	80
Table 3-1: Parameters for solute types used in this chapter. B-type solute is “solvent-philic”, as indicated by the positive free energy of transfer. C-type is “solvent-phobic”, as indicated by a negative free energy of tr. ....	121
Table 3-2: Values of $E_L$ for systems of monomers ( $N=1$ ), dimers ( $N=2$ ), and tetramers ( $N=4$ ) for chains of A, B, and C type solute particles. ....	122
Table 3-3: Values of $\mu^{ex}$ for systems of monomers ( $N=1$ ), dimers ( $N=2$ ), and tetramers ( $N=4$ ) for chains of A, B, and C type solute particles. ....	123
Table 4-1: Key parameters for all-atom particle types. H-type particles are “solvent-philic”, as indicated by the positive free energy of transfer. T-type are “solvent-phobic”, as indicated by a negative free energy of transfer and high solute enhancement ratio. ....	148
Table 4-2: Behavior of $H_1T_4$ and $H_1T_5$ solutions using the all-atom and DDIS potentials. ....	149
Table 4-3: Solute enhancement ratios (SERs) for interactions in single micellar aggregate simulations. “AA” denotes all-atom simulation, “DDIS” density dependent implicit solvent. ....	150
Table A-1: Energy Convergence with Respect to Basis Set. ....	180
Table A-2: Summary of Binding Energies for all syst. studied. ....	181

# 1 Introduction

## 1.1 MOTIVATION AND OBJECTIVES

A common class of molecular simulations involves a small number of solute molecules in a large bath of solvent (typically water). Examples systems include proteins, surfactants, bilayers, and other biological structures. Despite the continuing increase in computational power, the simulation of this class of systems remains computationally intractable. The fundamental limitation is the presence of the solvent, which represents more than 99% of the particles, and a comparable fraction of the computational load.

Molecular simulations of these systems (including Monte Carlo and Molecular Dynamics) represent an inefficient investment of computational resources because we are generally not interested in the details of the solvent behavior, but its presence is necessary to correctly predict the behavior of the more interesting solute. Therefore, we can realize large benefits by developing models in which the solvent is not simulated explicitly, but instead treat its effect by modifying the solute particle interactions. Such models are referred to as implicit solvent models.

The aim of this dissertation is to implement an implicit solvent model through the process of coarse-graining (CG). Here, coarse-graining refers to any general algorithm in which the computational load of a molecular simulation is reduced by a corresponding reduction in the degrees of freedom (i.e. particles) of the simulation. The novelty of our CG technique is that it represents an inversion of the traditional CG process. Whereas

previous approaches have emphasized the optimization algorithm and subsequently applied the technique to various systems, we first choose a motivating example and design a CG methodology specifically tuned to the needs of that example. In addition to the inversion, our approach also naturally broadens the scope of the traditional CG procedure to include considerations outside the specific optimization algorithm.

The specific application we select is the aggregation of surfactant molecules into micelles. Surfactants are an ideal model molecule with which to study the effectiveness of coarse-graining techniques because they produce complex solution behavior from simple molecular structures. In addition, surfactants have a long history of experimental, theoretical, and simulation study to draw upon.

Surfactant solutions contain solute heterogeneities that persist over a long time scale. For that reason, we seek an implicit solvent potential that is responsive to the local environment by parameterizing the potential by the instantaneous local density of solute particles. Given its dependence on the solute density, we call such potentials density dependent, implicit solvent (DDIS) potentials.

**The specific objectives of this thesis are to:**

- **Develop a coarse-graining methodology to generate DDIS potentials from all-atom simulations of monomeric solute in solvent.**
- **Examine the transferability of the generated DDIS potentials to other state points without the use of further optimization.**

- **Evaluate the performance of the DDIS potentials to predict thermodynamic properties of surfactant solutions, including critical micelle concentration.**

## **1.2 BACKGROUND**

### **1.2.1 Coarse-Graining**

Coarse-graining refers to any general algorithm in which the computational load of a molecular simulation is reduced by a corresponding reduction in the degrees of freedom (i.e. particles) of the simulation. Such a general definition allows for a wide variety of techniques to be classified as coarse-graining. For the purposes of this work, we can classify coarse-graining techniques into two broad classes: non-algorithmic coarse-graining and algorithmic (or systematic) coarse-graining.

Non-algorithmic coarse-graining refers to any technique in which a reduced representation of the system is made in an intuitive fashion, or without an underlying all-atom simulation. Common examples of non-algorithmic CG models are Kremer-Grest bead-spring polymer simulations<sup>84</sup> and Dissipative Particle Dynamics (DPD)<sup>85</sup>. In both cases, a simple nonbonded interaction potential is chosen to qualitatively reproduce a desired system behavior. However, the results of these simulations are not directly comparable to an all-atom simulation.

In systematic coarse-graining, by contrast, one seeks to create a reduced model of a molecular system that matches some properties of an underlying all-atom simulation. A

systematic coarse-graining scheme usually involves a short simulation in which every particle is explicitly included, which is used to generate descriptive “data” about the exact system. This simulation is followed by a matching procedure in which the coarse-grained potential is created and applied to a reduced number of particles to reproduce the data of the underlying all-particle simulation. In this thesis, we focus exclusively on issues related to systematic coarse-graining; that is, CG systems explicitly related to an underlying all-atom simulation.

An excellent body of work already exists in the field of systematic coarse-grained molecular simulation. This work spans a range of techniques and applications. Various types of data have been used for CG potential fitting, including forces<sup>1-7</sup>, reversible work<sup>8</sup>, radial distribution functions<sup>9-12</sup> (RDFs), and experimental results<sup>13-15</sup>.

Additionally, a wide variety of fitting procedures have been proposed, including simplex optimization<sup>16-21</sup>, radial distribution function inversion<sup>22-31</sup>, wavelet transform<sup>32,33</sup>, and semi-grand canonical Monte Carlo<sup>13</sup>. We review here some of the main contributions to the field.

Gregory Voth’s group has proposed a fitting method that minimizes the difference between the aggregate force acting on the atoms comprising a coarse-grained interaction site, and the measured force in the coarse-grained simulation<sup>1-6</sup>. The process begins with a short all-atom simulation in which the force vector acting on each atom is registered at every time step. The all-atom to coarse-grained particle mapping is then defined, at which point the modeler can identify the total force acting on each coarse-grained



particle. From this aggregated force data, a linear regression is performed that determines the interparticle potential that minimizes the error between measured and fit force.

Alexander Lyubartsev and coworkers have proposed an algorithm by which a radial distribution function (RDF) generated from an all-atom simulation is reproduced in a corresponding coarse-grained simulation<sup>27-30</sup>. The algorithm is a modification of the original process described by Schommers<sup>22,23</sup>. The premise of the RDF-matching method is predicated on Henderson's Theorem<sup>33</sup>, which guarantees the existence of a unique pairwise potential for any given RDF, but Lyubartsev's method does not necessarily solve for this exact potential. Instead, he relies on an iterative procedure in which an initial guess of the coarse-grained potential is refined by successive simulations. Despite this limitation, it has been argued that Lyubartsev's method is likely to converge to the optimum result in many common cases<sup>34</sup>.

The Muller-Plathe<sup>15-20</sup> fitting method also derives its CG potentials from RDFs. However, while the potentials generated from the Lyubartsev method are of arbitrary shape, the Muller-Plathe potentials are fit to a predefined functional form. The parameters of this functional form are generated through an iterative simplex optimization. For a small number of unknown parameters, this method is quite effective, but the computational load increases rapidly with increasing model complexity.

Kremer<sup>76-79</sup> and Faller<sup>80-83</sup> have developed CG polymer models in which the nonbonded interactions are fit to RDFs. Kremer's approach is to use a Lennard-Jones type potential and to manually adjust the exponents of the potential to best reproduce the RDF. These approaches are not explicitly systematic CG methods, as they rely on manual adjustment of the parameters and other considerations (e.g. group contribution theory) to set the parameters.

An open issue is that the methods described above do not generate the same CG potential given the same system. For example, RDF matching and force matching can produce fundamentally different results for identical coarse-grained representations of a system. Therefore, in the absence of any other information, neither method can be considered a demonstrably correct coarse-graining approach. Indeed the choice of fitting data set might be situation specific, and might incorporate variables outside those mentioned above.

### **1.2.2 Transferability of Coarse-Grained Potentials**

The value of a coarse-grained potential is determined in large part by its utility, or “transferability”, to situations outside of the one to which it was fit, because transferability is the only way that the upfront cost of performing the fit can be recouped. As suggested by Johnson *et al.*<sup>44</sup>, transferability can be classified as one of two types: “observable transferability”, and “state point transferability”. Observable transferability (called “representability” by Johnson *et al.*) refers to the ability of a CG potential that is fit to one set of simulation observables to reproduce accurately another set of simulation

observables. State point transferability, by contrast, refers to ability of a CG potential that is fit at one thermodynamic state point (temperature, density, system size, and composition) to predict the same observables at other thermodynamic state points. Both types of transferability have been previously addressed in the literature.

Lyubartsev<sup>45</sup> examined the observable transferability of a CG potential for lipid molecules in water. The potential utilized RDF matching, and was subsequently shown to reproduce reasonably well the density profile of a lipid bilayer. Noid *et al.*<sup>5</sup> showed that for isotropic, homogeneous materials a force-matched potential will also reproduce the system RDF. Frequently, however, CG algorithms are used to simulate systems on time scales that are inaccessible via all-atom simulation. In these cases, demonstrating observable transferability is impossible, since the all-atom simulation cannot be performed for purposes of comparison. Further, theoretical study by Louis<sup>46</sup> demonstrated that observable transferability breaks down even in simple cases; for example, he showed that potentials fit to system RDFs do not reproduce system energy, and *vice-versa*. As a result, he suggested that only more complex potential forms, incorporating density dependence or many-body interactions, might overcome these representability problems.

State point transferability has been studied many times as well. Henderson's seminal theoretical work<sup>33</sup> indicates that the pairwise potential that generates a given RDF is unique to within an additive constant for systems with only pairwise interactions. However, the RDF of a system depends on the state point; it does not follow that the

potential obtained by RDF inversion at one state point will generate the correct RDF at other state points. Recent simulations of polymers<sup>47</sup> by this approach have exhibited limited state-point transferability of such potentials.

The study of transferability is key to generating good CG strategies, and the recent literature indicates that CG schemes frequently break down when applied to properties or state points far from the point of fit. Increasing the robustness of CG potentials can only be accomplished by intelligent trial and error.

### **1.2.3 Implicit Solvent Models**

Implicit solvent models are those in which the explicit treatment of solvent-solvent and solvent-solute interactions is replaced by a modified solute-solute interaction potential designed to capture some of the solvent-mediated behavior. Typically, the solvent is assumed to be much smaller, and to equilibrate much faster, than the solute. These models have received a great deal of interest in recent years, from arenas including proteins<sup>48</sup>, surfactants<sup>74</sup>, membranes<sup>49</sup>, tissues<sup>10</sup> and macromolecules<sup>11,50,51</sup>. Because many of the systems of interest are solvent rich, the implicit solvent formulation offers the possibility to improve greatly the efficiency of computer simulations of such systems, because it permits a vast reduction in the number of degrees of freedom.

Simulation in the absence of a solvent (generally water) has previously been achieved by the addition of a “solvation free energy” term to the energy model, which is derived from an underlying physical theory to reduce or eliminate the need for optimization. Two

previous approaches to the implicit solvent problem bear special mention. The solvent-accessible surface area (SASA) model<sup>52</sup> adds its solvation free energy term based on the surface area of the solute exposed to solvent. The SASA model requires the computation of a surface area for the solute molecule(s) at each time step during a molecular dynamics simulation, greatly increasing the computational cost. Although theoretically motivated, SASA does not guarantee the matching of properties between all-atom and implicit solvent simulations.

A second approach is the Effective Energy Function (EEF)<sup>53</sup> of Lazaridis *et al.*, in which the implicit water effects are treated as a perturbation to the existing all-atom pairwise potential between solute sites. An attractive perturbation is added for hydrophobic effects and a repulsive perturbation for hydrophilic effects, with the parameters fit such that the free energy change produced in transferring a solute particle from a fully solvent-exposed location to a location deeply buried within a solute-rich environment matches the experimentally derived free energy of transfer. Though computationally simple, the perturbation introduced by the EEF model necessarily changes the internal structure of solute-rich regions of the simulation (as measured by the radial distribution function), such that the density of hydrophobic regions of the simulation is significantly higher than intended, with the first peak of the RDF shifted to shorter distance than that observed in an equivalent all-atom simulation.

In this work, we examine the process of creating an implicit solvent potential within the coarse-graining framework. In many coarse-graining applications, a small number of

particles or degrees of freedom are combined and represented by a new, “coarse-grain” particle with fewer degrees of freedom. A coarse-grained, implicit solvent model achieves a similar reduction in degrees of freedom by eliminating some number of particles (the “solvent”) while retaining others (the “solute”) in unaltered form, except for the requisite correction to the solute interaction potential to reflect solvent-mediated interaction. Thus, implicit solvent models can be considered a special case of the CG framework described above.

#### **1.2.4 Surfactant Solutions: Theory and Simulations**

The design choices made in the modified CG approach described above are system dependent. The four key questions lose meaning without a specific application with which to refer. We selected simulations of surfactant systems to further study our CG framework.

Surfactant molecules are composed of a hydrophilic head group that is soluble in water and a hydrophobic tail group that is not. At sufficient concentration in solution, surfactant molecules will spontaneously aggregate into structures known as micelles, in which hydrophilic head groups surround and shield the hydrophobic tail groups from the solvent.

Theoretical approaches<sup>54,55</sup> treat micellization as arising from a set of competing free energy effects. The transfer of hydrophobic tails from melt to solution promotes micelle formation, but is offset by a surface energy penalty and the loss of surfactant translational

entropy. These competing effects create a well-defined free energy minimum as a function of aggregation number, so that the resulting micelles are of nearly uniform size. The concentration at which surfactant molecules spontaneously form micelles is referred to as the critical micelle concentration, or CMC, and in theoretical treatments is directly related to the aforementioned free energy considerations. Both the CMC and the average aggregation number are important properties that characterize the self-assembling nature of surfactants and for which we desire efficient, predictive capabilities.

The literature on particle-based surfactant simulations contains a diverse set of particle representation approaches<sup>56-75</sup>, which can be grouped according to two key design decisions: the level of detail used to describe the surfactant molecule itself, and explicit or implicit treatment of the solvent. Explicit solvent simulations<sup>56-72</sup> are computationally burdensome because of the low value of experimentally measured CMC's ( $10^{-6}$ - $10^{-2}$  M). At these concentrations, a significant majority of the simulation cell is composed of the relatively uninteresting solvent. As a result, the direct measurement of the CMC by explicit solvent simulations is impractical, even for very coarse-grained surfactant and solvent representations. Instead, the CMC and aggregation number distribution are generally determined indirectly, through the use of free energy simulations<sup>63,67,68,72</sup>.

By reducing the number of particles represented, and thus also the computational burden, implicit solvent simulations of surfactant systems allow investigation of much longer time and length scales. By this approach, Lazaridis *et al.*<sup>74</sup> studied the formation of

dodecylphosphocholine (DPC) micelles using the Effective Energy Function (EEF1) implicit solvent model. The simulated CMC was close to the reported experimental value. Von Gottberg *et al.*<sup>75</sup> studied  $A_2B_2$  surfactants in implicit solvent using stochastic dynamics (SD), in which each surfactant bead was intended to represent a Kuhn segment (approximately ten methylene groups) of a polymer chain. They successfully simulated a fully equilibrated micelle size distribution at multiple surfactant concentrations, from which they were able to determine the system CMC.

The literature on the molecular simulation of surfactants suggests that implicit solvent simulations allow the measurement of thermodynamic properties including CMC and average aggregation number. Therefore, creating an implicit solvent energy model via coarse-graining could be quite advantageous. Properties of interest could be measured with little computational burden and the results would have a clear correspondence to an underlying all-atom energy model that is presumed to be a more accurate representation of the system.

### **1.3 OUTLINE OF A NEW COARSE-GRAINING APPROACH**

Our coarse-graining methodology is a four-step process. Each step answers a key question:

- 1) What property or properties does one seek to conserve during the coarse-graining process?
- 2) What is the potential form that can sufficiently capture the property or properties of interest?



- 3) Once a potential form has been selected, how does one optimize the potential to reproduce the properties of interest?
- 4) How transferable is the resulting CG potential to other properties and state points?

Question 1 begins as a hypothesis, which can be validated or invalidated by the results of the coarse-graining. In other words, it is a testable hypothesis. An iterative process in which increasingly complex functional forms are tested until an adequate reproduction of the observables is achieved answers questions 2 and 3. Question 4 is an empirical test of the potential's capability and the assumptions that went into the CG process. Following the four-step process generates information with which the modeler can modify their original hypothesis and thus systematically improve the quality of their CG potential. This framework is described by the flow sheet given in Figure 1-1. Note that there are multiple feedback loops, which allow one to repeatedly revisit all relevant design decisions.

In essence, we approach the coarse-graining problem as a more general statistical inversion problem, in which the input parameters of a model (in this case, a molecular simulation model) are tuned so as to reproduce a desired set of observable properties. In the special case of CG simulation, the observables are taken directly from an underlying all-atom simulation, but the actual fitting algorithm could just as easily be chosen to reproduce experimental measurements, simulations previously reported in the literature, or even the intuitions of the modeler. The success or failure of such coarse-grained potentials is measured empirically by its ability to accurately reproduce the properties

that the user expresses interest in. In fact, the set of observables used to train a coarse-grained potential might be very successful in predicting some properties of a molecular simulation but an abject failure in others.

The inversion framework is common in the development of all-atom potentials. For example, J. Ilja Siepmann's<sup>35-43</sup> group at the University of Minnesota has created the TraPPE (Transferable Potentials for Phase Equilibria) potential for the prediction of liquid phase properties. As a result of iterative study and improvement, their research has indicated that reproducing the vapor-liquid equilibrium curve for a given molecule produces a robust potential that accurately reproduces many other system properties including solubilities, structure factors, and self-diffusion coefficients.

## **1.4 THESIS OVERVIEW**

The objective of this thesis is to propose a coarse-graining framework that is robust, flexible, and extensible. We invert the traditional conception of coarse-graining in molecular simulations by first selecting a motivating application, and designing a CG methodology around that application. As a motivating example, we investigate the behavior of the new framework within the problem of surfactant simulation.

Chapter 2 addresses the fundamental design choices we made in creating a coarse-grained, implicit solvent simulation. These design choices included the matching parameters and form of potential, and were made with the terminal application of surfactant simulation in mind. Chapter 3 addresses the state-point transferability of the

corresponding CG potentials to more complex systems, including solute mixtures, chains, and surfactant-like systems. Chapter 4 builds on all our previous work to investigate the behavior of surfactant solutions using our CG potentials. Finally, Chapter 5 summarizes our findings, and suggests future avenues of research that could add accuracy and computational efficiency to the model.

## 1.5 REFERENCES

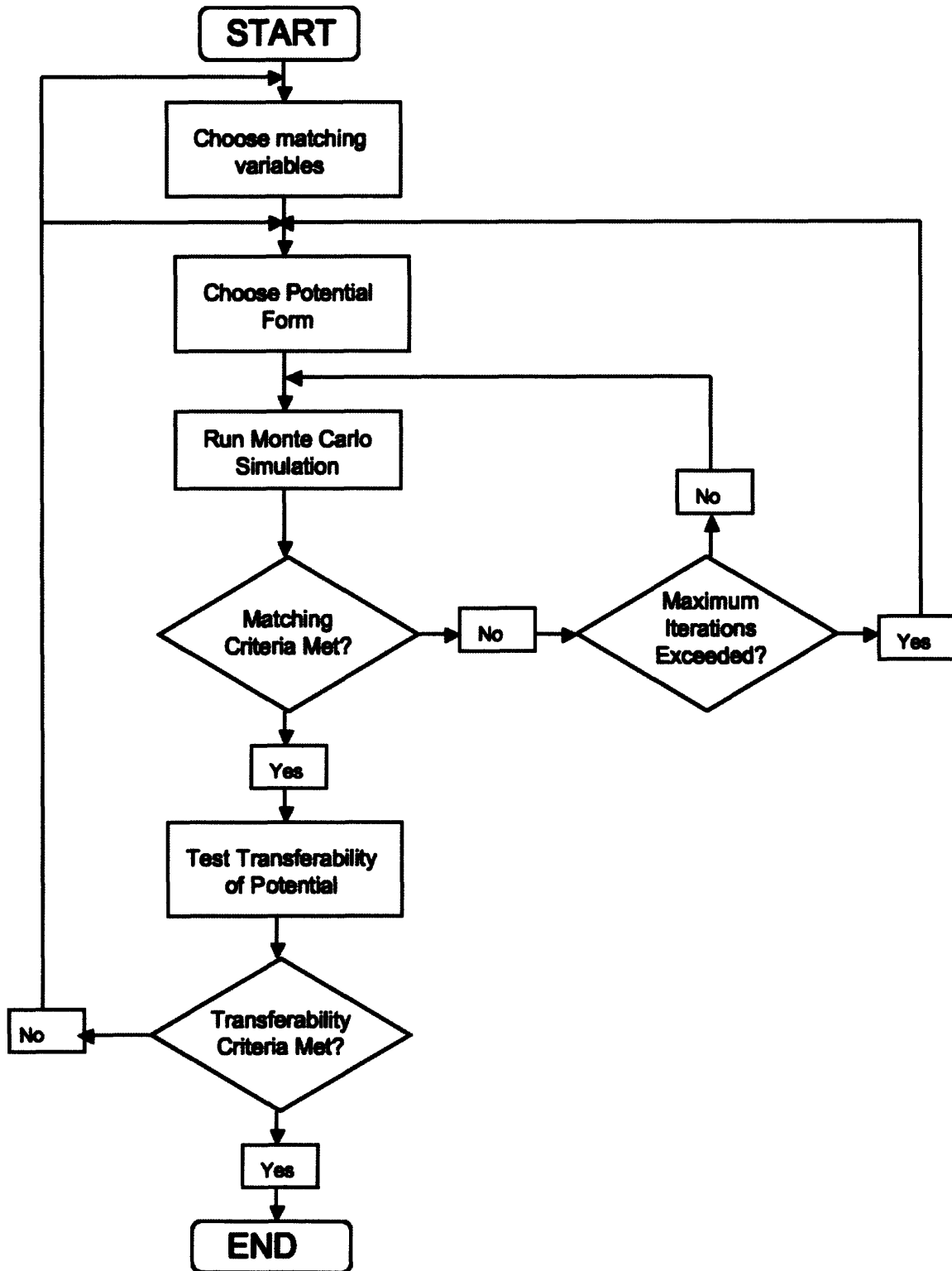
1. Q. Shi, S. Izvekov, and G. A. Voth, *J. Phys. Chem. B* **110**, 15045 (2006).
2. S. Izvekov and G. A. Voth *J. Chem. Phys.* **125**, 151101 (2006)
3. G. S. Ayton, W. G. Noid, and G. A. Voth, *Mrs Bull.* **32**, 929 (2007).
4. P. Liu, S. Izvekov, and G. A. Voth, *J. Phys. Chem. B* **111**, 11566 (2007).
5. W. G. Noid, J. W. Chu, G. S. Ayton, and G. A. Voth, *J. Phys. Chem. B* **111**, 4116 (2007).
6. J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, *Biophys. J.* **92**, 4289 (2007).
7. J.D. McCoy and J.G. Curro, *Macromolecules* **31**, 9362 (1998).
8. K. Kremer and F. Muller-Plathe, *Mol. Simulat.* **28**, 729 (2002).
9. N. B. Wilding, *J. Chem. Phys.* **119**, 12163 (2003).
10. M. Bathe and G. C. Rutledge, *J. Comput. Chem.* **24**, 876 (2003).
11. M. Bathe, G. C. Rutledge, A. J. Grodzinsky, and B. Tidor, *Biophys. J.* **88**, 3870 (2005).
12. G. C. Rutledge, *Phys. Rev. E* **63** 021111 (2000).
13. S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, *J. of Chem. Phys.* **119**, 7043 (2003).

14. W. Shinoda, R. Devane, and M. L. Klein, *Mol. Simulat.* **33**, 27 (2007).
15. R. Faller, H. Schmitz, O. Biermann, and F. Muller-Plathe, *J. Comput. Chem.* **20**, 1009 (1999).
16. H. Meyer, O. Biermann, R. Faller, D. Reith, and F. Muller-Plathe, *J. Chem. Phys.* **113**, 6264 (2000).
17. D. Reith, H. Meyer, and F. Muller-Plathe, *Macromolecules* **34**, 2335 (2001).
18. D. Reith, M. Putz, and F. Muller-Plathe, *J. Comput. Chem.* **24**, 1624 (2003).
19. G. Milano, S. Goudeau, and F. Muller-Plathe, *J. Polym. Sci. Pol. Phys.* **43** (8), 871 (2005).
20. G. Milano and F. Muller-Plathe, *J. Phys. Chem. B* **109**, 18609 (2005).
21. W. Schommers, *Phys. Lett. A* **43**, 157 (1973).
22. W. Schommers, *Phys. Rev. A* **28**, 3599 (1983).
23. D. Levesque, J.J. Weis, and L. Reatto, *Phys. Rev. Lett.* **54**, 454 (1985).
24. G. Toth and A. Baranyai, *Mol. Phys.* **97** (3), 339 (1999).
25. G. Toth and A. Baranyai, *J. Mol. Liq.* **85**, 3 (2000).
26. G. Toth and A. Baranyai, *J. Phys.: Condens. Matter* **17**, S159 (2005).
27. A. P. Lyubartsev and A. Laaksonen, *J. Phys. Chem.* **100**, 16410 (1996).
28. A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E* **55**, 5689 (1997).
29. A. P. Lyubartsev and A. Laaksonen, *Comput. Phys. Comm.* **122**, 57 (1999).
30. A. P. Lyubartsev and A. Laaksonen, *Chem. Phys. Lett.* **325**, 15 (2000).
31. A.E. Ismail, G.C. Rutledge, G. Stephanopoulos, *J. Chem. Phys.* **118**, 4414 (2003).
32. A.E. Ismail, G. Stephanopoulos, G.C. Rutledge, *J. Chem. Phys.* **118**, 4424 (2003).
33. R.L. Henderson, *Phys. Lett. A* **49**, 197 (1974).

34. A.K. Soper, Chem Phys. **202**, 295 (1996).
35. M.G. Martin and J.I. Siepmann, J. of Phys. Chem. B **102**, 2569 (1998).
36. M.G. Martin and J.I. Siepmann, J. of Phys. Chem. B **103**, 4508 (1999).
37. B. Chen, B and J.I. Siepmann, J. of Phys. Chem. B **103**, 5370 (1999).
38. C.D. Wick, M.G. Martin, and J.I. Siepmann, J. of Phys. Chem. B **104**, 8008 (2000).
39. B. Chen, J.J. Potoff, and J.I. Siepmann, J. of Phys. Chem. B, **105**, 3093 (2001).
40. J.M. Stubbs, J.J. Potoff, and J.I. Siepmann, J. of Phys. Chem. B, **108**, 17596 (2004).
41. C.D. Wick, J.M. Stubbs, N. Rai, and J.I. Siepmann, J. of Phys. Chem. B, **109**, 18974 (2005).
42. N. Lubna, G. Kamath, J.J. Potoff, N. Rai, and J.I. Siepmann, J. of Phys. Chem. B, **109**, 24100 (2005).
43. N. Rai, and J.I. Siepmann, J. of Phys. Chem. B, **111**, 10790 (2007).
44. M.E. Johnson, T. Head-Gordon, and A.A. Louis, J. Chem. Phys. **126**, 144509 (2007).
45. A. P. Lyubartsev, Eur. Biophys. J. Biophys. **35**, 53 (2005).
46. A.A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
47. P.G. Bolhuis, A.A. Louis, J.P. Hansen, and E.J. Meijer, J. Chem. Phys. **114**, 4296 (2001).
48. B. Zagrovic, E.J. Sorin, and V. Pande, J. Mol. Biol. **313**, 151 (2001).
49. K.H. Lin, N.A. Baker, and J.A. McCammon, Biophys. J **83**, 1374 (2002).
50. P.G. Bolhuis, A.A. Louis, J.P. Hansen, and E.J. Meijer, J. Chem. Phys. **114**, 4296 (2001).
51. P.G. Bolhuis, A.A. Louis, and J.P. Hansen, Phys. Rev. E **64**, 021801 (2001).
52. D. Eisenberg and A.D. McLachlan, Nature **319**, 199 (1986).

53. T. Lazaridis and M. Karplus, *Prot. Struct. Funct. Genet.* **35**, 133 (1999).
54. S. Pubbada and D. Blankschtein, *J. Chem. Phys.* **92**, 3710 (1990).
55. L. Maibaum, A.R. Dinner, and D. Chandler, *J. Phys. Chem. B* **108**, 6778 (2004).
56. B. Jonsson, O. Edholm, and O. Teleman, *J. Chem. Phys.* **85**, 2259 (1986).
57. K. Watanabe, M. Ferrario, and M.L. Klein, *J. Chem. Phys.* **92**, 819 (1988).
58. A.D. MacKerell, *J. Phys. Chem.* **99**, 1846 (1995).
59. D.P. Tielman, S.J. Marrink, and H.J.C. Berendsen, *Biochim. Biophys. Acta* **1331**, 235 (1997).
60. M. Tarek, S. Bandyopadhyay, and M.L. Klein, *J. Mol. Liq.* **78**, 1 (1998).
61. S.J. Marrink, D.P. Tieleman, and A.E. Mark, *J. Phys. Chem. B* **104**, 12165 (2000).
62. D.P. Tielman, D. van der Spoel, and H.J.C. Berendsen, *J. Phys. Chem. B* **104**, 6389 (2000).
63. N. Yoshii, K. Iwahashi, and S. Okazaki, *J. Chem. Phys.* **124**, 184901 (2006).
64. B. Smit, *Phys. Rev. A* **37**, 3431 (1988).
65. B. Smit, P.A.J. Hilbers, K. Esselink, L.A.M. Rupert, N.M. van Os, and A.G. Schlijper, *J. Phys. Chem.* **95**, 6361 (1991).
66. B. Smit, K. Esselink, P.A.J. Hilbers, N.M. van Os, L.A.M. Rupert, and I. Szleifer, *Langmuir* **9**, 9 (1993).
67. P.G. Bolhuis and D. Frankel, *Physica A* **244**, 45 (1997).
68. M. Kinoshita and Y. Sugai, *Chem. Phys. Lett.* **313**, 685 (1999).
69. S.O. Nielsen, C.F. Lopez, G. Srinivas, and M.L. Klein **119**, 7043 (2003).
70. S.J. Marrink, A.H. de Vries, and A.E. Mark **108**, 750 (2004).
71. R. Pool and P.G. Bolhuis, *J. Chem. Phys. B* **109**, 6650 (2005).

72. M. Hatakeyama and R. Faller, *Phys. Chem. Chem. Phys.* **9**, 4662 (2007).
73. H. Sinto, S. Marisada, M. Miyahara, and K. Higashitani, *Langmuir* **20**, 2017 (2004).
74. T. Lazaridis, B. Mallik, and Y. Chen, *J. Phys. Chem. B* **109**, 15098 (2005).
75. F.K. von Gottberg, K.A. Smith, and T.A. Hatton, *J. Chem. Phys.* **106**, 9850 (1997).
76. V.A. Harmandaris, D. Reith, N.F.A. van der Vegt, and K. Kremer, *Macromol. Chem. Phys.*, **208**, 2109 (2007).
77. W. Tschop, K. Kremer, J. Batoulis, T. Buerger, and O. Hahn, *ActaPolym.* **49**, 61 (1998).
78. C. Abrams and K. Kremer, *Macromolecules* **36**, 260 (2003).
79. V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, and K. Kremer, *Macromolecules* **39**, 6708 (2006).
80. R. Faller, H. Schmitz, O. Biermann, and F. Muller-Plathe, *J. Comput. Chem.* **20**, 1009 (1999).
81. J. Ghosh, and R. Faller, *Mol. Simulat.* **33**, 759 (2007).
82. Q. Sun, and R. Faller, *J. of Chem. Phys.* **126**, 144908 (2007).
83. Q. Sun, F.R. Pon, and R. Faller, *Fluid Phase Equilibria* **261**, 35 (2007).
84. K. Kremer and G.R. Grest, *J. Chem. Phys.* **92**, 5057 (1990).
85. R.D. Groot and P.B. Warren, *J. Chem. Phys.* **107**, 4423 (1997).



**Figure 1-1:** Flow Diagram for our modified coarse-graining technique. Multiple feedback loops allow for rigorous iterative improvement at all steps of the process.



## 2 A Novel Algorithm for Creating Coarse-Grained, Density Dependent Implicit Solvent Models

### 2.1 INTRODUCTION

Implicit solvent models are those in which the explicit treatment of solvent-solvent and solvent-solute interactions is replaced by a modified solute-solute interaction potential designed to capture some of the solvent-mediated behavior. Typically, the solvent is assumed to be much smaller, and to equilibrate much faster, than the solute. These models have received a great deal of interest in recent years, from arenas including proteins<sup>1</sup>, surfactants<sup>2</sup>, membranes<sup>3</sup>, tissues<sup>4</sup> and macromolecules<sup>5-7</sup>. Because many of the systems of interest are solvent rich, the implicit solvent formulation offers the possibility to improve greatly the efficiency of computer simulations of such systems, because it permits a vast reduction in the number of degrees of freedom.

Here, we examine the process of creating an implicit solvent potential within the coarse-graining framework<sup>8</sup>. In coarse-graining, one seeks to create a reduced model of a molecular system that matches some properties of an underlying all-atom simulation. A coarse-graining scheme usually involves a short simulation in which every particle is explicitly included, which is used to generate descriptive “data” about the exact system. This simulation is followed by a matching procedure in which the coarse-grained potential is created and applied to a reduced number of particles to reproduce the data of the underlying all-particle simulation. In many coarse-graining applications, a small

number of particles or degrees of freedom are combined and represented by a new, “coarse-grain” particle with fewer degrees of freedom. A coarse-grained, implicit solvent model achieves a similar reduction in degrees of freedom by eliminating some number of particles (the “solvent”) while retaining others (the “solute”) in unaltered form, except for the requisite correction to the solute interaction potential to reflect solvent-mediated interaction. The resulting coarse-grain interaction potential is a potential of mean force between solute particles that serves to preserve the solution structure. Reproducing the full thermodynamic behavior of the all-particle simulation, however, requires some additional accounting for the lost solvent-solvent interactions. In contrast to the traditional assumptions of Brownian dynamics that allow one to treat the solvent implicitly as a heat bath, the coarse-graining method is inherently a fitting process that seeks to optimize a trade-off between accuracy and computational efficiency through a reduction of degrees of freedom between interacting particles.

To explain our approach here, we begin with three basic questions that have guided this work:

- 1) What property or properties does one seek to conserve during the coarse-graining process to maximize the utility of the resulting potential?
- 2) What is the simplest possible potential form that can sufficiently capture the property or properties of interest?
- 3) Once a potential form has been selected, how does one optimize the potential to reproduce the properties of interest?

Coarse-graining studies have traditionally been concerned primarily with the last question, with much of the excellent work thus far focused on reproducing the forces of an underlying all-particle simulation<sup>9,10</sup>, or on replicating the radial distribution function<sup>11-17</sup>. Although both approaches are reasonable, neither is a demonstrably complete description of the system, and the appropriateness of one or the other may be situation specific.

In contrast to coarse-graining, the implicit solvent literature has typically focused on the first two questions. Simulation in the absence of a solvent (generally water) is achieved by the addition of a “solvation free energy” term to the energy model, which is derived from an underlying physical theory to reduce or eliminate the need for optimization. Two previous approaches to the implicit solvent problem bear special mention. The solvent-accessible surface area (SASA) model<sup>18</sup> adds its solvation free energy term based on the surface area of the solute exposed to solvent. The SASA model requires the computation of a surface area for the solute molecule(s) at each time step during a molecular dynamics simulation, greatly increasing the computational cost. Although theoretically motivated, SASA does not guarantee the matching of properties between all-atom and implicit solvent simulations.

A second approach is the Effective Energy Function (EEF)<sup>19</sup> of Lazaridis *et al.*, in which the implicit water effects are treated as a perturbation to the existing all-atom pairwise potential between solute sites. An attractive perturbation is added for hydrophobic effects and a repulsive perturbation for hydrophilic effects, with the parameters fit such

that the free energy change produced in transferring a solute particle from a fully solvent-exposed location to a location deeply buried within a solute-rich environment matches the experimentally derived free energy of transfer. Though computationally simple, the perturbation introduced by the EEF model necessarily changes the internal structure of solute-rich regions of the simulation (as measured by the radial distribution function), such that the density of hydrophobic regions of the simulation is significantly higher than intended, with the first peak of the RDF shifted to shorter distance than that observed in an equivalent all-atom simulation.

In this chapter, we develop a coarse-grained, implicit solvent potential with explicit consideration of each of the three basic questions stated above. Sections II, III, and IV of this chapter deal with each of the three questions in turn. Section V describes the computational methodology. In Section VI, we apply the coarse-graining method to a simple system in which the solute and solvent are identical Lennard-Jones (LJ) particles.

## **2.2 CONSERVED PROPERTIES**

To address the question of which properties to conserve in moving from an all-atom to implicit solvent simulation, we first consider the types of systems in which we are interested. One of these is the protein folding problem, wherein a protein's minimum energy conformation depends in large part on the pattern of hydrophobic and hydrophilic amino acids along the chain. This relative hydrophobicity of amino acids is defined in an implicit solvent model through the free energy of transfer of a solute particle from a fully solvent-exposed environment to a fully screened environment. Additionally, the packing

density of solute particles is important. If the free energy of transfer of a hydrophobic group is correctly reproduced in the implicit solvent model, but the density of a cluster of these hydrophobic groups is too high, the resulting packing may result in unrealistic intramolecular strain, or in the shielding or exposure of apolar groups. The radial distribution function serves as a proxy for these structural characteristics.

Another class of system of interest is that in which surfactant solutions undergo micelle formation. Blankschtein's model<sup>20</sup> of micelle formation shows that the critical micelle concentration and average micelle size depend not only on the free energy of transfer of the hydrophobic tail groups, but also on the surface tension between tail and water (which depends both on free energy considerations *and* solute density considerations) and the packing density of both heads and tails.

Motivated by these examples, we suggest that a good implicit solvent model for such problems should replicate both the free energy of transfer of solute groups and the radial distribution function of solute particles (especially at high concentrations). The quantities are general to all molecular simulations and measurable in a computationally reasonable amount of time. Approaches that address only one of these factors in describing protein folding or micellization are likely to be incomplete.

### **2.3 COARSE-GRAINED POTENTIAL FORM**

Having identified the radial distribution function and the free energy of transfer as the most important properties to capture, we next turn to the question of the form of

interaction model required to replicate these properties. We note that the answer to this question is inherently subjective, since the quality of reproduction desired by the model builder determines the complexity of the model required. In the case of implicit solvation, the form of this potential can be derived from the  $N$ -particle partition function,  $Z_N$ , in a straightforward way, as follows:

$$Z_N = \int_V dr^N \exp[-\beta V(r^N)] \quad (1)$$

where  $V(r^N)$  is the exact interaction energy, which depends explicitly on the locations of all  $N$  particles. Dividing the particles into  $n$  solute and  $N-n$  solvent and integrating over the solvent degrees of freedom, this partition function becomes:

$$\begin{aligned} Z_N &= \int_V dr_{solv}^{N-n} \int_V dr_{solute}^n \exp[-\beta V(r_{solv}^{N-n}, r_{solute}^n)] \\ &= C_{solv} \int_V dr_{solute}^n \exp[-\beta V^{EFF}(r_{solute}^n)] \end{aligned} \quad (2)$$

where  $V^{EFF}$  is an effective potential of mean force between solute particles that includes solute-solvent interactions, and  $C_{solv}$  is a constant that arises from solvent-solvent interactions. Rewriting  $C_{solv} = \exp[-\beta \mu n]$ , the partition function becomes:

$$Z_N = \int_V dr^n \exp \left[ -\beta \sum_{i=1}^n \left( \mu + \sum_{j=1}^n V^{EFF}(r_{ij}) \right) \right]. \quad (3)$$

Here, the “solute” subscript has been dropped for ease of notation. Pairwise potentials have been assumed, and  $i$  and  $j$  index the  $n$  solute particles. For sufficiently simple systems with pair-wise interactions only, Henderson’s theorem guarantees the existence of a unique pairwise potential to within an additive constant for a given radial distribution function (RDF). Thus, a constant can be added to the potential of mean force between explicit solute particles in order to account for the missing solvent-solvent interactions in a mean field sense, without altering the RDF produced. Although  $\mu$  is often set to zero in many effective potentials, it is apparent that a nonzero  $\mu$  is necessary to reproduce the free energy of transfer or excess chemical potential. This is because  $V^{EFF}$  alone fails to capture the total energy of the excluded solute-solvent and solvent-solvent interactions. As another straightforward illustration of this fact, consider the extreme situation of insertion of a solute particle into a bath of solvent particles at infinite dilution. In the all-atom simulation, this insertion produces a non-zero excess chemical potential. In the implicit solvent simulation, however, there are no particles with which the solute atom can interact, and the measured excess chemical potential of a pairwise potential in the absence of the self-interaction constant would necessarily be zero. Setting  $\mu = \mu_{ex}$  in this case can provide the necessary match of excess chemical potential. According to Henderson’s theorem, introduction of the additive constant  $\mu$  corrects the energy but does not alter the structure of the fluid associated with  $V^{EFF}$ .

In general, the effective potential form proposed here is a function of the thermodynamic state point, due to the integration over solvent particle positions taken in Eq (2). This is supported on both theoretical<sup>16</sup> and empirical<sup>5</sup> grounds. As such, both the one-body ( $\mu$ )

and two-body ( $V^{EFF}$ ) terms may depend on the temperature  $T$ , total particle number  $N$ , total particle density  $\rho$ , and solute particle density  $\rho_S$ . For the present study, we focus on parameterization as a function of the solute particle density  $\rho_S$ , which we believe to be most relevant to the problems of protein folding and micellization. In the results section, we also investigate the limits of transferability of such potentials to other state points defined by small changes in  $N$ ,  $T$ , and  $\rho$ .

Thus, we arrive at the final form of the potential:

$$E_i = \mu(\rho_{S,i}) + \frac{1}{2} \sum_{j \neq i} V^{EFF}(r_{ij}, \rho_{S,i}) \quad (4)$$

where  $E_i$  is the effective energy of particle  $i$ ,  $V^{EFF}$  is a pairwise potential between solute particles that is explicitly a function of solute particle density in the vicinity of particle  $i$ ,  $\rho_{S,i}$ , and  $\mu$  is a “self-interaction” term (also a function of solute particle density). The inclusion of a density-dependent self-interaction energy makes this potential similar in form to the Embedded Atom Method<sup>23</sup>. In this chapter, we consider two estimates of the solute particle density. The first is the global average density of solute particles, which we denote  $\rho_G$ , and the second is the instantaneous local density of solute particles, which we denote  $\rho_L$ ; both  $\rho_G$  and  $\rho_L$  are referred to collectively as  $\rho_S$  where the subscript “S” stands for solute. Finally,  $\rho$  without subscript refers to the state point density, considering all particles (solute and solvent).



## 2.4 POTENTIAL PARAMETERIZATION

### 2.4.1 Metrics

For all of the results reported below, we define three metrics by which to assess the accuracy of the implicit solvent potentials to replicate the all-atom systems. The first of these is the local average number of solute particles,  $N_L$ , within the interaction radius,  $r_C$ :

$$N_L = \rho_G \int_0^{r_C} g(r, \rho_G) 4\pi r^2 dr. \quad (5)$$

The second metric is the local energy per particle, defined as:

$$E_L = \rho_G \int_0^{r_C} V(r) g(r, \rho_G) 4\pi r^2 dr \quad (6)$$

where  $\rho_G$  is the global average density of solute particles in the simulation,  $V(r)$  is the exact interaction potential, and  $g(r, \rho_G)$  is the coarse-grained solute-solute RDF, which may be density-dependent. We expect an accuracy of better than 1% error in both  $N_L$  and  $E_L$ .  $E_L$  provides a measure of the error in  $g(r, \rho_G)$ , compared to the exact  $g(r)$  for the explicit system, in units of energy for comparison to  $\mu^{ex}(\rho_G)$ , described next. As a third metric, we also computed  $\mu^{ex}(\rho_G)$ , the excess chemical potential, using standard procedures as described below. For this metric, we expect an error of no more than  $0.05 k_B T$ , or  $\sim 2\%$ , which is comparable to the typical experimental error<sup>21</sup>. This accuracy should hold across a wide range of solute particle densities.

## 2.4.2 Global Density Dependent Potentials

In the global density-dependent formulation, both terms in Eq. (4) are parameterized on the bulk density  $\rho_G$  of solute particles in the simulation cell. Each particle in the simulation interacts through the same potential. Because the self-interaction potential does not impact the radial distribution function of the resulting simulation, the creation of the global density dependent potential can be broken down into two separate steps, one of which is used to match the radial distribution function and the other to match the excess chemical potential.

In the first step, the pairwise portion of the potential,  $V^{EFF}(r, \rho_G)$ , is generated using the iterative method proposed first by Schommers<sup>11,12</sup>. Successive approximations of the potential are created using the update rule:

$$V_{i+1}(r, \rho_G) = V_i(r, \rho_G) - k_b T \ln \left( \frac{g_{TARGET}(r, \rho_G)}{g_i(r, \rho_G)} \right) \quad (7)$$

where  $V_i$  represents the estimate of  $V^{EFF}(r, \rho_G)$  at iteration  $i$ , and  $g_i(r)$  the RDF generated by  $V_i$ . Schommers' method is expected to converge for initial  $V_1$  that is close to  $V^{EFF}(r, \rho_G)$ , but convergence is slow for poor initial guesses. This is only one of a number of update methods that have been proposed to solve the RDF inversion problem<sup>7,13,17</sup>.

For the lowest density simulation, we begin with the Boltzmann inversion of the RDF as a trial function  $V_1$ :

$$V_1(r, \rho_G) = -k_b T \ln(g_{TARGET}(r, \rho_G)) \quad (8)$$

For subsequent higher density systems, the converged solution of the next lower density is used as the initial guess. Schommers' iterative procedure is followed until the error function

$$error_i = \frac{\int_0^{r_c} |g_i(r, \rho_G) - g_{TARGET}(r, \rho_G)| dr}{\int_0^{r_c} g_{TARGET}(r, \rho_G) dr} \quad (9)$$

decreases to within a specified tolerance of 0.3%. This fit is much closer than is needed to achieve a 1% error in the local particle number or per-particle energy.

The second step in the fitting method is to calculate the self-interaction energy,  $\mu(\rho_G)$ , that correctly reproduces the excess chemical potential,  $\mu^{ex}(\rho_G)$ . The complexity that arises in free energy calculations in a density-dependent potential is shown in Figure 2-1. For an all-particle potential (Figure 2-1(a)),  $\mu^{ex}$  is measured using a free energy method such as thermodynamic integration<sup>24,25</sup> or the acceptance ratio method<sup>25,26</sup>, which measures the energy difference,  $\Delta F_1$ , between a reference particle that interacts with the surrounding medium in one case, and one that acts as a phantom (non-interacting) particle in the second case. In an implicit solvent, density-dependent potential (Figure 2-1(b)), however, the insertion of a reference particle changes the density of the system and thereby introduces a secondary free energy effect,  $\Delta F_2(\rho_S)$ , such that:

$$\mu^{ex} = -(\Delta F_1 + \Delta F_2) \quad (10)$$

$\Delta F_2(\rho_S)$  is the energy change associated with changing between energy models at the initial density ( $\rho_S=n/V$ ) and the final density ( $\rho_S=(n-1)/V$ ):

$$\Delta F_2\left(\rho_S = \frac{n}{V}\right) = \Delta F_{2,1}(\rho_S) + \Delta F_{2,2}(\rho_S) \quad (11)$$

where  $\Delta F_{2,1}(\rho_S)$  and  $\Delta F_{2,2}(\rho_S)$  are the one-body and two-body contributions, respectively, and are given by:

$$\Delta F_{2,1}\left(\rho_S = \frac{n}{V}\right) = (n-1)\mu\left(\frac{n-1}{V}\right) - n\mu\left(\frac{n}{V}\right) \quad (12)$$

$$\Delta F_{2,2}\left(\rho_S = \frac{n}{V}\right) = \frac{(n-1)^2}{2V} \int_0^{r_c} g(r, \rho_S) \left[ V^{EFF}\left(r, \frac{n-1}{V}\right) - V^{EFF}\left(r, \frac{n}{V}\right) \right] 4\pi r^2 dr \quad (13)$$

In the above discussion, we have used the density notation  $\rho_S$  to indicate that the analysis is relevant to both the global and local solute density cases. Combining Eqs. (10), (11), (12), and (13) yields an equation for  $\mu(\rho_G=n/V)$ :

$$\mu\left(\rho_G = \frac{n}{V}\right) = \frac{1}{n} \left[ \mu_{ex}\left(\frac{n}{V}\right) + \Delta F_1\left(\frac{n}{V}\right) + \Delta F_{2,2}\left(\frac{n-1}{V}\right) \right] + \frac{(n-1)}{n} \mu\left(\frac{n-1}{V}\right) \quad (14)$$

For an  $N$  particle system, there are  $N$  such equations (for  $n=1..N$ ). We set  $\mu(\rho_G=0) = \mu^{ex}(\rho_G=0)$ . For values of  $n$  where  $\Delta F_1(\rho_G=n/V)$  and  $V^{EFF}(r, \rho_G=n/V)$  are not measured, they are approximated by linear interpolation from the nearest known values.

If a suitable target RDF and excess chemical potential exist, one can also perform the fitting algorithm described for solute particle densities exceeding the total particle density: that is, the implied density of solvent particles is negative. Although physically unrealistic for systems parameterized on global density, these results are important for the local density parameterization described next, since the local density of solute particles will frequently exceed the average total particle density through temporary fluctuations when  $\rho_G$  is sufficiently close to  $\rho$ .

### 2.4.3 Local Density Dependent Potentials

Most solute/solvent systems of interest exhibit a heterogeneous distribution of particles, with local areas of solute concentration or depletion. Such heterogeneities may be instantaneous, in the case of thermal fluctuations, or long-lived, as in the cases of micellization or phase separation. In these cases, the global density does not represent the local environment that a solute particle experiences at a given point in time. It has been suggested that improved transferability might result from using local density-dependent potentials<sup>27</sup>. Once again, if we consider solute density as a proxy for the number of missing solvent-solvent and solute-solvent interactions, the instantaneous local density is intuitively a more appropriate proxy than the ensemble-averaged global density.

In the local density dependent formulation, the potential that any particular particle experiences at any given time is dependent on the density of solute particles within a

cutoff radius,  $r_D$ . Thus, the potential experienced by a particle is a dynamic property. The local density dependent potential then has two required components: (i) the two body interaction potential  $v^{EFF}(r, \rho_L)$ , which we have implemented as a two-dimensional array specified at regularly spaced intervals of  $r$  and  $\rho_L$ , the instantaneous local density, and (ii) the self-interaction energy,  $\mu(\rho_L)$ , which we have implemented as a one-dimensional array at regularly spaced intervals of  $\rho_L$ . Values of  $v^{EFF}$  and  $\mu$  that are intermediate to the specified points are determined by linear interpolation. Interactions between particles that experience different local densities are evaluated as the average of the interaction computed at each density.

In this chapter, the cutoff radius for counting particles within the local density approximation,  $r_D$ , is set equal to the potential cutoff,  $r_C$ . Since the local density is a proxy for missing solvent interactions, it is logical to consider the relevant local density as the density within the range of potential interaction. This choice should be most accurate for systems where the variations in local (solute) density occur over distances that are large compared to the range of the interaction potential itself. We experimented with three methods for determining local density: equal weighting of every particle within the cutoff radius, a linear decay in particle weighting with inter-particle spacing, and a quadratic decay in particle weighting with inter-particle spacing. The behavior of the system under all three assumptions was essentially the same, and so all results are reported here using the linear weighting method.

As with the global density dependent case, we break the problem of potential fitting into two steps: first fitting the pairwise potential, and then fitting the self-interaction term. Two factors must be considered with regard to fitting local density dependent potentials: first, Henderson's theory assures an optimal solution to the *global* density dependent fitting problem, with a unique pairwise potential reproducing the desired RDF and a unique self-interaction energy reproducing the excess chemical potential (both of these properties are ensemble averaged). The *local* density dependent case, by contrast, offers no such guarantees, since the potential is required to reproduce RDFs and excess chemical potentials across a range of simulations at varying global densities. As such, the quality of fit deteriorates somewhat in return for a better representation of density fluctuations. Second, a rich literature exists on the topic of fitting global density dependent potentials; it therefore makes sense to treat the local density fitting as a transformation of the existing global density solution. With the preceding in mind, the approach followed here is to use the set of previously determined global density potentials, obtained over a range of solute densities, to obtain a first approximation to the local density potential. To fit the pairwise potential, we first note that the radial distribution function (RDF) of a system at global solute density  $\rho_G$  can be expanded as follows:

$$\rho_G g(r, \rho_G) = \sum_i P(\rho_{L,i}) \rho_{L,i} g(r, \rho_{L,i}) \quad (15)$$

where  $P(\rho_{L,i})$  is the probability of observing local density  $\rho_{L,i}$ , and  $g(r, \rho_{L,i})$  is the radial distribution function about particles that experience an instantaneous local density  $\rho_{L,i}$ .  $g(r, \rho_{L,i})$  can then be expanded in powers of the density<sup>11</sup> as:

$$g(r, \rho_{L,i}) = \exp\left[\frac{-v^{EFF}(r, \rho_{L,i})}{k_B T}\right] \gamma(r, \rho_{L,i}) \quad (16)$$

with

$$\gamma(r, \rho_{L,i}) = 1 + \rho_{L,i} a(r) + \frac{\rho_{L,i}^2}{2!} b(r) + \dots \quad (17)$$

The coefficients of Eq. (17) are cluster integrals of the effective potential<sup>11</sup>. Inserting Eq. (16) into Eq. (15) results in:

$$\rho_G g(r, \rho_G) = \sum_i P(\rho_{L,i}) \rho_{L,i} \exp\left[\frac{-v^{EFF}(r, \rho_{L,i})}{k_B T}\right] \gamma(r, \rho_{L,i}) \quad (18)$$

Schommers<sup>11</sup> states that  $\gamma(r, \rho_S)$  is only mildly dependent on the potential, since  $v^{EFF}(r, \rho_S)$  plays the role of an integral quantity in  $\gamma(r, \rho_S)$ . Therefore, we assume that  $\gamma(r, \rho_G) = \gamma(r, \rho_{L,i})$  when  $\rho_{L,i} = \rho_G$ . For a set of  $m$  simulations at  $m$  distinct  $\rho_G$  values, this generates a set of  $m$  equations for each inter-particle separation  $r$ , with the local density dependent potentials as the only unknown values. These linear equations can be solved once the global density-dependent potentials are known. Eq. (18) gives a simple framework to



derive appropriate local density dependent potentials once the equivalent global potentials are known. Although the  $\gamma$ -equivalence assumption may introduce a small source of error in estimating the optimum local potential, in our experience the ability to avoid a computationally expensive optimization of the local pairwise potential makes this trade-off worthwhile. Additionally, we demonstrate below that the magnitude of the introduced error is small enough to be considered inconsequential.

The second part of the fitting process is choosing appropriate values for the self-interaction constants  $\mu(\rho_L)$ . For a large value of the density cut-off radius, the distribution of local densities  $P(\rho_L)$  is tightly peaked around  $\rho_G$ , and one can again assume that a direct transfer of global values to local values, *i.e.*  $\mu(\rho_L)=\mu(\rho_G)$  when  $\rho_L=\rho_G$ , offers a reasonable first estimate.

In our experience, this initial guess offers an acceptable approximation of the RDF, with differences in per-particle energy (Eq. (6)) between the global and local density models of  $\sim 0.3\%$  and differences in local solute number  $N_L$  (Eq. (5)) of  $\sim 0.2\%$ ; the approximation for  $\mu^{ex}$  is not as good (errors up to  $0.11 k_B T$ , or  $5\%$ ). Additionally, the RDF is less sensitive to changes in  $\mu(\rho_L)$  than  $\mu^{ex}$  is. Therefore, we use an iterative procedure to determine the set of  $\mu(\rho_L)$ 's that reproduce the simulated  $\mu^{ex}(\rho_G)$ 's over a range of  $\rho_G$ .

The contributions to  $\mu^{ex}$  can be separated into those that explicitly depend on  $\mu(\rho_L)$  and those that do not. From the global density theory, the only term in Eq. (10) with explicit

reference to  $\mu$  is the  $\Delta F_{2,1}(\rho_S)$  term. We assume that the other terms in the equation remain unchanged under changes in  $\mu$ , and produce an update equation of the form:

$$\Delta F_{2,1}(\rho_L)^{NEW} = \Delta F_{2,1}(\rho_L)^{OLD} + \Delta\Delta F_{2,1}(\rho_L) \quad (19)$$

$$\Delta\Delta F_{2,1}(\rho_L) = \mu_{MEASURED}^{ex}(\rho_G = \rho_L) - \mu_{TARGET}^{ex}(\rho_G = \rho_L) \quad (20)$$

Plugging Eq. (12) into Eq. (19) leads to the following update for  $\mu$ :

$$\Delta\mu\left(\rho_L = \frac{n}{V}\right) = \left(\frac{n-1}{n}\right)\Delta\mu\left(\rho_L = \frac{n-1}{V}\right) - \frac{1}{n}\left[\mu_{MEASURED}^{ex}\left(\rho_G = \frac{n}{V}\right) - \mu_{TARGET}^{ex}\left(\rho_G = \frac{n}{V}\right)\right] \quad (21)$$

where  $\Delta\mu(\rho_L)$  is the step change to take in  $\mu(\rho_L)$ . We solve the equation series by setting  $\Delta\mu(\rho_L=0)$  to 0 and solving by substitution for all larger values of  $\rho_L$ . For values of  $n$  where  $\mu_{MEASURED}^{ex}(\rho_G)$  is not known, it is approximated by linear interpolation from the nearest known values. In our experience, the update  $\Delta\mu$  suggested by Eq. (21) is too large, and so this value is scaled by a factor  $\alpha$  to improve convergence:

$$\mu(\rho_L)^{NEW} = \mu(\rho_L)^{OLD} + \alpha\Delta\mu(\rho_L) \quad (22)$$

For the system studied here,  $\alpha=0.5$  worked effectively.

## 2.5 SIMULATION PROTOCOL

### 2.5.1 All Atom Simulations

All atom simulations using the bare LJ potential were performed using GROMACS (Version 3.3)<sup>28</sup> in the NVT ensemble with periodic boundary conditions. We employed a total of  $4 \times 10^6$  time steps with step size of 2fs, for a total of 8 ns of simulation.

Trajectories were printed every 2 ps, and used to generate simulated RDFs. Temperature was controlled with a Berendsen thermostat. A typical 1000-particle simulation required approximately 2 hours on a 2.80GHz Intel Xeon CPU.

Excess chemical potentials were calculated by inserting a single Lennard Jones particle into the simulation cell using thermodynamic integration (TI)<sup>24,25</sup>; the standard GROMACS  $\lambda$  switching function was used to integrate between an initial state in which the test particle interactions with the remaining system were switched off and a final state in which the test particle interactions were fully enabled. Switching was applied to nonbonded interactions only, with soft core interactions to avoid singularities and using a soft core interaction parameter  $\alpha=0.51$ . A total of 31  $\lambda$  values were used ( $\lambda = [0.00 \ 0.03 \ 0.07 \ 0.10 \ \dots \ 0.93 \ 0.97 \ 1.00]$ ). Each simulation employed  $4 \times 10^5$  time steps with step size of 2 fs, for a total of 800 ps of simulation. This simulation length was chosen because the estimated uncertainty in the chemical potentials was estimated to be  $0.04 k_B T$  or less for all simulations. Free energy simulations were performed in parallel, each requiring on the order of one hour on a 2.80GHz Intel Xeon CPU.

## 2.5.2 Global Density Dependent Simulations

Simulation protocol for the global density-dependent, implicit solvent simulations was identical to that of the all atom simulations, with the exception of the potential used.

Custom coarse-grained potentials were implemented using the GROMACS table option.

## 2.5.3 Local Density Dependent Simulations

The local density dependent case requires a form of the energy model that is not readily available in common molecular dynamics codes. As a result, we used instead a Monte Carlo (MC) code of our own design. Data was generated via Monte Carlo in the NVT ensemble using periodic boundary conditions. We employed a total of  $4 \times 10^8$  translation moves of single particles, with a maximum step size of  $0.4\sigma$ . Trajectories were printed every  $4 \times 10^4$  moves and used to generate simulation RDFs. Because the movement of a single particle affects the local density (and, by extension, the effective potential) of a large number of particles, a naïve MC energy calculation would be prohibitively expensive. To avoid recalculating the local density of every particle at every time step, the density of non-moving particles was updated every  $N$  moves using the multiple time step Monte Carlo algorithm<sup>29</sup>. A typical 1000-particle simulation required approximately 40 hours on a 2.80GHz Intel Xeon CPU. The longer simulation time is a function of both the local density dependence and the switch from MD to MC. To compare the global and local density dependent simulation times, we also performed a set of MC simulations using the global density dependent potential. The local density dependent simulation period is only about 50% longer than an equivalent, global density dependent MC simulation.

Excess chemical potentials were calculated by inserting a single LJ particle into the simulation cell using the Bennett Acceptance Ratio (BAR) method<sup>26</sup>. Two  $\lambda$  values were used ( $\lambda = [0.00, 1.00]$ ) with an initial state comprising a non-interacting test particle and a final state having a fully interacting test particle. Prior to performing the local density simulations, the BAR method was validated by comparing the chemical potential measurements obtained by TI and BAR. Results were within  $0.05\epsilon$  for all densities.

#### **2.5.4 External Field Simulations**

To demonstrate the effect of density variations under the global and local density approximations, and to mimic micellization using the simple LJ system used here, some simulations were performed in the presence of an external field that acts only on solute particles. The simulation methodology is identical to the local density case, with the exception of the additional external potential. This potential is implemented as a point particle in the center of the simulation cell, which attracts solute particle and repels solvent particles. The interaction of the solute particles with the point particle is identical in the all-atom and implicit solvent simulations.

### **2.6 RESULTS AND DISCUSSION**

#### **2.6.1 All-atom Simulation**

The system considered in this work consists of a bath of  $N$  identical particles, interacting via the Lennard-Jones (LJ) potential, at constant volume and temperature. The system is

coarse-grained by arbitrarily selecting a fraction  $X_S=n/N$  of the particles to be “solute”, and the remainder “solvent”. Such a system offers a number of simplifications for testing purposes. The main one is that, for a given  $N, V, T$ , the solute-solute radial distribution function and excess chemical potential are independent of the degree of coarse-graining, or fraction of particles designated as solute,  $X_S$ , and equal to that of the conventional  $N$ -particle LJ system. As such, we require only one simulation to generate statistics for the explicit, all-particle system, and the results of this simulation are generally applicable across the entire range of  $X_S=(0,1]$ . The specification of a system in which the system RDF is independent of density makes this equivalent to an ISO- $g^{(2)}$  process<sup>30</sup>.

The test system used in this work is described in Table 2-1, and is known to be in the liquid region of the Lennard-Jones phase diagram<sup>31</sup>.  $A^*$  denotes a quantity defined in reduced units, where  $\rho^* = \rho\sigma^3$ ,  $T^* = k_B T/\epsilon$ , and  $r_c^* = r_c/\sigma$ .

A base simulation at the given  $N, \rho^*, T^*$  was used to generate the radial distribution function and excess chemical potential. The coarse-graining schemes described in this chapter are used to fit the potential across a range of solute fractions from 0.1 to 1.2 in increments of 0.1. The excess chemical potential of the simulation is -2.34 (in units of  $\epsilon$ ) and the RDF is given in Figure 2-2.

## 2.6.2 Global Density Dependent Potentials

The maximum number of iterations required to solve for any potential was 14, with the average number being 9.6. The maximum error, as defined by Eq. (9) was 0.3%.

For perspective, Figure 2-2 shows the RDF generated for the solved potential  $V^{EFF}$  at  $X_S=0.1$  compared to the target RDF. The results are so close as to be indistinguishable. The agreement for other values of  $X_S$  is equally good. The average per-particle energy error and local density error are on the order of 0.1%.

The resulting pairwise potentials are given in Figure 2-3. The location of the minimum in energy remains relatively constant across all global solute densities,  $\rho_G=X_S\rho$ . At low densities there is a local maximum at  $\sim 1.7\sigma$  that gradually disappears as the solute particle density is increased. The form of the pairwise potential as a function of  $X_S$  illustrates clearly how the effective potential transitions smoothly from a potential of mean force at low concentrations to a Lennard-Jones-like potential at high concentrations. At  $X_S=1.00$ , we recover the LJ potential, as expected.

The results for  $\mu(\rho_G)$  are shown in Figure 2-4. At low densities ( $X_S<0.7$ ), the value of the self-interaction term is very close to  $\mu^{ex}$ , reflecting the fact that the change in chemical potential due to pairwise interactions at these densities is very close to zero.  $\mu(\rho_G)$  captures this missing free energy change. At a density equal to the total particle density, the self-interaction term is not exactly zero, as it is in the Lennard-Jones potential. This is a peculiar feature of the density dependent framework; upon deletion of a solute particle from a simulation of  $n$  explicit solute particles and  $N-n$  implicit solvent particles, the potential behaves as though the resulting system consists of  $n-1$  solute particles and  $N-n+1$  implicit solvent particles. In effect, the deletion of one solute particle leads to the insertion of one implicit solvent particle.

### 2.6.3 Pairwise potential transferability to other state points

As discussed previously, the coarse-grained potentials described here are dependent on thermodynamic state and are not generally transferable to other system densities, temperatures, or sizes. Frequently, this limitation is unimportant in implicit solvent simulations because one is mainly interested in changes of solute composition at fixed total density. Nevertheless, in this section we address the limits of transferability of the effective potential obtained at one state point  $(N_1, \rho_1, T_1)$  and applied to another state point  $(N_2, \rho_2, T_2)$ . To do so, we compared the RDFs of all-particle simulations with those of coarse-grained simulations for different values of  $T^*$ ,  $\rho^*$ , and  $N$ , using the effective potential derived for the state point given by Table 2-1 with  $X_S=0.1$ . This concentration was chosen because the effective potential at that point is most dissimilar from the LJ potential.

Figure 2-5 shows the RDF generated at  $X_S=0.10$  across a range of temperatures using the effective potential generated at  $T^*=1.53$ . These results are compared to the all-particle simulation RDFs. At temperatures lower than the fitting temperature, shown in Figures 5(a) and 5(b), the RDFs generated by the effective potential are more sharply peaked than the target RDFs, while at temperatures higher than the fitting temperature, Figures 5(c) and 5(d), the RDFs are less sharply peaked. However, the location of the first peak of the RDF remains at the correct inter-particle distance across the entire temperature range. Table 2-2 lists the error in the average per-particle energy as a function of temperature change. For the lowest temperature simulation ( $T^*=0.83$ ), the average per-particle energy



error is approximately 5%. This state point is in the two-phase region of the LJ phase diagram<sup>31</sup>. At all other temperatures, the magnitude of the error is less than 3%, with errors of 1% for temperature changes of 10%.

Figure 2-6 compares all-atom and coarse-grained simulation results at different densities. As with the temperature variation experiments, the location of the first peak of the RDF remains correct. Table 2-3 demonstrates that changes in the density of up to 10% have less than a 1% impact on the per-particle energy.

Within the current system, then, it appears that a temperature or density change of up to 10% introduces at most a 1% error in the average per-particle energy.

Finally, we find that an increase in the number of particles (at constant total density,  $\rho$ ) by a factor of ten only introduces a 0.8% error in the per-particle energy. For this system, it appears that the correct effective potential is insensitive to system size. For inhomogeneous systems, especially those containing a phase boundary, one cannot assume system size independence.

The error in  $\mu^{ex}$  introduced by transferring the effective potential to another state point is more significant. Figure 2-7 compares the excess chemical potential for  $X_S=1.00$  and  $X_S=0.10$  with the true all-particle value, as a function of system temperature. The results for  $X_S=1.00$  match well with the target values for all of the single phase systems, but the results for  $X_S=0.10$  do not match well for  $T \neq T_{FIT}$ . This is because  $\mu^{ex}$  values at low solute

density (high level of coarse graining) are determined almost entirely by the self-interaction constant, which was assumed to depend only on the density of explicit solute particles. The results are similar for transferring the effective potential from one system density to another.

#### 2.6.4 Local Density Dependent Potentials

For the special case with identical solute and solvent particles, the solution of Eq. (18) under the  $\gamma$ -equivalence assumption requires that the local pairwise potential be identical to the global pairwise potential at an equivalent density, that is  $v^{EFF}(r, \rho_L) = V^{EFF}(r, \rho_G)$  when  $\rho_L = \rho_G$ .

For  $\mu(\rho_L)$ , we use as an initial guess  $\mu(\rho_L) = \mu(\rho_G)$  when  $\rho_L = \rho_G$ . After three iterations, the value of  $\mu^{ex}$  at each simulated density is identical to the value determined from the all-particle simulation to within the specified tolerance of  $\pm 0.07\epsilon$ . The measured profile of  $\mu^{ex}$  with solute particle density using the initial guess and after convergence is shown in Figure 2-8. Clearly, the initial approximation yields a very good result, with all values within  $\pm 0.12\epsilon$  of the target excess chemical potential.

Table 2-4 lists the errors in per-particle energy and local solute number for the coarse-grain simulations, and demonstrates that all these values are below the specified tolerance of 1%. Figure 2-9 compares the RDF's obtained using coarse graining with the local density parameterization and the all-particle simulations for the worst case scenario

( $X_S=1.00$ ). Even for this example, the local density RDF and all-particle RDF are very nearly indistinguishable.

The results in Figures 8 and 9 indicate that the local density representation is comparable in accuracy to the global density representation. This is due in large part to the selection of a large cutoff radius,  $r_D$ . Setting  $r_D = r_c = 5\sigma$  reduces the fluctuations in local density so that the distribution of local densities  $P(\rho_L)$  is tightly peaked around  $\rho_G$  for the homogeneous systems studies here. For every solute density utilized here, we measured the standard deviation in  $\rho_L$  as a function of  $\rho_G$ . In every case, the standard deviation is less than 5% of the total particle density. In the limit that  $r_D$  approaches the size of the simulation cell, the local density approximation reduces exactly to the global density approach.

Reducing  $r_D$  renders the  $\gamma$  equivalence assumption less accurate, such that  $v^{EFF}(r, \rho_L)$  is not longer well approximated by  $V^{EFF}(r, \rho_G)$ , and may require that  $v^{EFF}$  be re-optimized for the particular choice of  $r_D$ . To illustrate this, we repeated the simulations with  $v^{EFF}(r, \rho_L) = V^{EFF}(r, \rho_G)$  but  $r_D$  reduced from  $5\sigma$  to  $2\sigma$ . Table 2-4 and Figure 2-10 display the results. The accuracy in all three metrics is considerably worse.

Further evidence of the equivalence of the global and local formulations for  $r_D = 5\sigma$  comes from simulations of model transfer to other temperatures. Just as we did with the global results, we performed simulations to measure both RDF and excess chemical potential at temperatures other than the fitting temperature. These simulations used the same

temperature values as were used to study the transferability of the global potentials, with  $X_S=0.10$ . Though results are not included here for the sake of space, the local results were identical to the global results within our specified limits of accuracy at all temperatures.

### 2.6.5 External Field Simulations

While the local and global density approximations are found to be similar in the aforementioned cases, where solute particles are homogeneously distributed throughout the system, the global density approximation is expected to break down for situations where the solute and solvent are distinct components and are inhomogeneously distributed through the system. Such inhomogeneous distributions are common, for example, in phase separation and in surfactant systems, where aggregation into micelles creates stable areas of high solute concentration. In such cases, the global density is not representative of the environment of the majority of solute particles.

To demonstrate the advantage of the local density approximation in this case, we replicate the aggregation of solute into a micelle for the system studied here by the addition of a point particle attractor in the center of the simulation cell. The point particle attracts solute particles with potential  $\phi_A$ :

$$\phi_A(r) = \begin{cases} -6\epsilon & \text{for } r \leq 5\sigma \\ 6\epsilon \left( \frac{1}{(5\sigma)^2} (r - 5\sigma)^2 - 1 \right) & \text{for } 5\sigma < r \leq 10\sigma \\ 0 & \text{for } r > 10\sigma \end{cases} \quad (23)$$

and repels solvent particles with potential  $\phi_B$ :

$$\phi_B(r) = \begin{cases} \infty & \text{for } r \leq 5\sigma \\ 0 & \text{for } r > 5\sigma \end{cases} \quad (24)$$

The strength of these potentials was chosen such that the density of solute near the center of the cell approached the total particle density, i.e.  $\rho_L \approx \rho$  in the vicinity of the point attractor, and  $\rho_L \approx 0$  for locations far from the attractor.

An all-atom simulation of 1000 solute particles and 9000 solvent particles ( $X_S=0.1$ ) was performed, followed by implicit solvent simulations using either the local density potential or the global density potential determined earlier for  $X_S=0.1$ . No additional fitting was performed in this step, as we were attempting to assess the relative transferability of the two density dependent approaches.

The superiority of the local density approach is hinted at in Figure 2-11, where the solute particle fraction is plotted as a function of distance from the point attractor. The local density dependent solute profile is much closer to the target profile than the global density dependent profile, with average errors (defined by Eq. (9)) of 3% and 20% respectively. Additionally, we measured  $\mu^{ex}$  for solute particle insertion, obtaining an all-atom value of  $-4.93 \pm 0.08$  (units of  $\epsilon$ ). The global potential returned  $-4.57 \pm 0.02$ , while the local density model yielded  $-4.92 \pm 0.02$ , which is identical to the all-atom value within simulation uncertainty. Thus, it appears that the local density approach is more amenable

to the study of heterogeneous systems. Further application of the local density approach to study micellar systems will be the subject of a future report.

## 2.7 CONCLUSION

In this chapter, we have defined a framework to create coarse-grained potentials for use in implicit solvent simulations in three steps: defining the goals of coarse-graining in terms of the desired conserved quantities, suggesting a potential that can achieve those goals, and only then defining a potential optimization scheme. In doing so, we have demonstrated the need for implicit solvent potentials to include a density dependence in order to capture fully the behavior at both low and high solute concentrations, especially in heterogeneous systems. The pairwise portion of the density dependent potential behaves as a potential of mean force at low concentrations and converges toward the bare Lennard-Jones potential at high concentrations, with a smooth transition in the intermediate ranges. Additionally, a “self-interaction” energy is required if one wishes to fit both the radial distribution function (RDF) and excess chemical potential ( $\mu^{ex}$ ) simultaneously.

The resulting potentials are independent of system size, but are in general functions of both temperature and density. However, for the system studied here, it appears that the pairwise portion of the potential may be used for temperature and density changes of up to 10% without violating acceptable standards of accuracy. The excess chemical potential is more sensitive to changes in state, and thus the self-interaction constant needs to be updated more frequently than the pairwise potential.

We have demonstrated that one may parameterize the potential based either on a global density dependence (i.e. the average solute particle density of the simulation), or on a local density dependence that varies in space and time. For sufficiently large cut-offs in the definition of local density, the results of a global density dependent fitting method (such as that described by Lyubartsev<sup>13</sup> or Schommers<sup>11</sup>) can be used as an initial guess for the appropriate local density potential and, depending on the desired accuracy of the potential, can be used without alteration. If greater accuracy is desired in the local density potential, the proposed iterative procedure improves results in only a few iterations.

We note briefly that the use of *solute* particle density as the parameterization variable is only one possibility. One could equally well parameterize the potential based on solvent density. The difference in these two approaches would manifest itself in calculation of the excess chemical potential. Under parameterization based on solute particle density, the insertion (deletion) of a solute particle leads to the deletion (insertion) of an implicit solvent particle. Essentially, the potential treats solute particle insertion as equivalent to a particle swap, rather than a particle insertion. By contrast, a parameterization based on implicit solvent particle density would treat a solute particle insertion as a true insertion, since the density of implicit solvent particles would remain unchanged by the insertion of a solute particle. The difference in assumption would lead to changes in the self-interaction energy, but both parameterizations are internally self-consistent. The relative advantages or disadvantages of these alternate approaches is unclear at this point, and we simply note that either form can be fit using the optimization schemes described above.

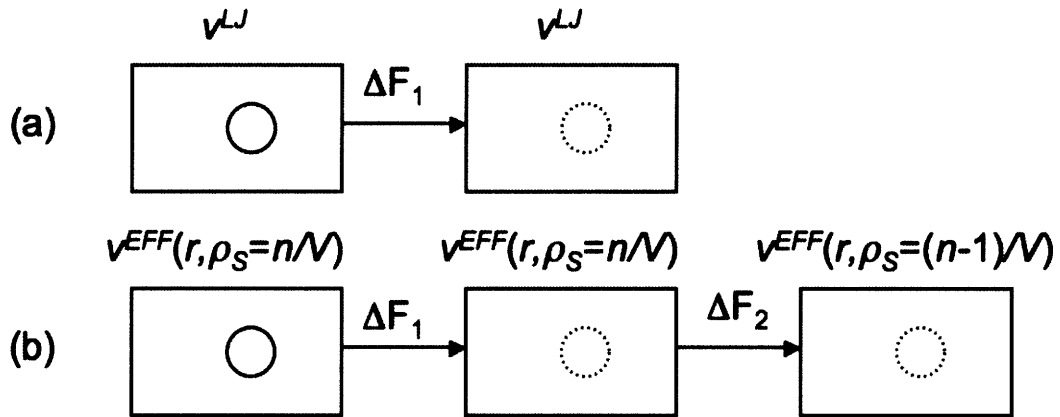
The utility of a local density dependence is limited in systems with a homogeneous distribution of solute particles because the distribution of instantaneous local densities is tightly clustered around the average global density. However, for systems of heterogeneous solute distribution we expect the local density approach to offer superior performance. This is indeed observed in the system with externally applied potential and indicates the potential utility of the local density approach for a number of physically important systems.

## 2.8 REFERENCES

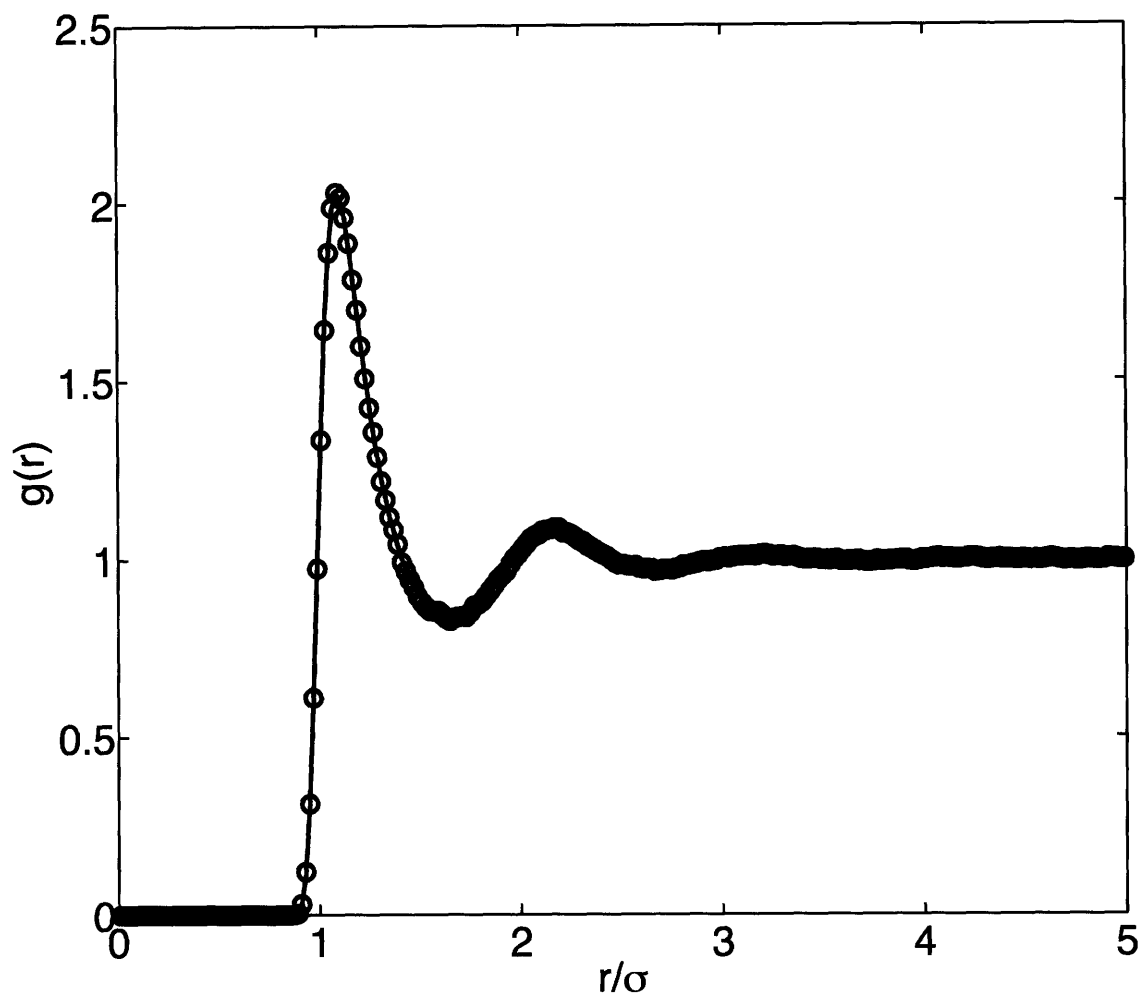
1. B. Zagrovic, E.J. Sorin, and V. Pande, *J. Mol. Biol.* **313**, 151 (2001).
2. T. Lazaridis, B. Mallik, and Y. Chen, *J. Phys. Chem. B* **109**, 15098 (2005).
3. K.H. Lin, N.A. Baker, and J.A. McCammon, *Biophys. J* **83**, 1374 (2002).
4. M. Bathe, G.C. Rutledge, A.J. Grodzinsky, and B. Tidor, *Biophys. J.* **88**, 3870 (2005).
5. P.G. Bolhuis, A.A. Louis, J.P. Hansen, and E.J. Meijer, *J. Chem. Phys.* **114**, 4296 (2001).
6. P.G. Bolhuis, A.A. Louis, and J.P. Hansen, *Phys. Rev. E* **64**, 021801 (2001).
7. M. Bathe and G.C. Rutledge, *J. Comput. Chem.* **24**, 876 (2003).
8. D. Reith, M. Putz, and F. Muller-Plathe, *J. Comput. Chem.* **24**, 1624 (2003).
9. S. Izvekov and G.A. Voth, *J. Phys. Chem. B* **109**, 2469 (2005).
10. S. Izvekov and G.A. Voth, *J. Chem. Phys.* **123**, 134105 (2005).
11. W. Schommers, *Phys. Lett. A* **43**, 157 (1973).
12. W. Schommers, *Phys. Rev. A* **28**, 3599 (1983).



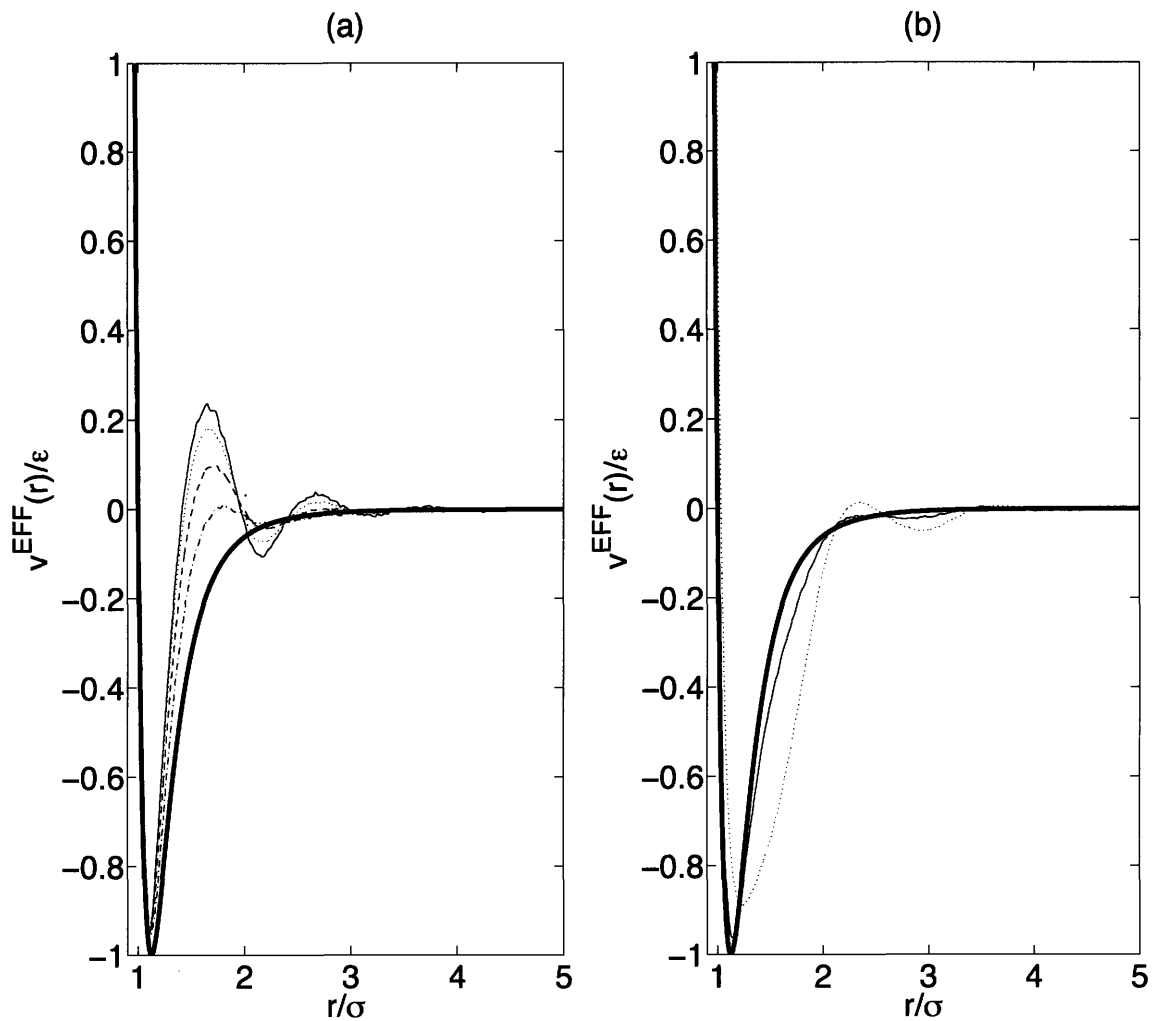
13. A.P. Lyubartsev and A. Laaksonen, *Phys. Rev. E* **52**, 3730 (1995).
14. G.C. Rutledge, *Phys. Rev. E* **63**, 021111 (2000).
15. H. Meyer, O. Biermann, R. Faller, D. Reith, and F. Muller-Plathe, *J. Chem. Phys.* **113**, 6264 (2000).
16. F.H. Stillinger, H. Sakai, and S. Torquato, *J. Chem. Phys.* **117**, 286 (2002).
17. N.B. Wilding, *J. Chem. Phys.* **119**, 12163 (2003).
18. D. Eisenberg and A.D. McLachlan, *Nature* **319**, 199 (1986).
19. T. Lazaridis and M. Karplus, *Prot. Struct. Funct. Genet.* **35**, 133 (1999).
20. C. Sarmoria, S. Puvvada, and D. Blankschtein, *Langmuir* **8**, 2690 (1992).
21. M.R. Shirts, J.W. Pitera, W.C. Swope, and V.S. Pande, *J. Chem. Phys.* **119**, 5740 (2005).
22. R.L. Henderson, *Phys. Lett. A* **49**, 197 (1974).
23. M.S. Daw and M.I. Baskes, *Phys. Rev. B* **29**, 6443 (1984).
24. T.P. Straatsma and J.A. McCammon, *J. Chem. Phys.* **95**, 1175 (1991).
25. M.R. Shirts, E. Bair, G. Hooker, and V.S. Pande, *Phys. Rev. Lett.* **91**, 140601 (2003).
26. C.H. Bennett, *J. Comput. Phys.* **22**, 245 (1976).
27. A.A. Louis, *J. Phys.: Condens. Matter* **14**, 9187 (2002).
28. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J.C. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).
29. B. Hetenyi, K. Bernacki, and B.J. Berne, *J. Chem. Phys.* **117**, 8203 (2002).
30. F.H. Stillinger, S. Torquato, J.M. Eroles, and T.M. Truskett, *J. Phys. Chem.* **105**, 6592 (2001).
31. B. Smit, *J. Chem. Phys.* **96**, 8639 (1992).



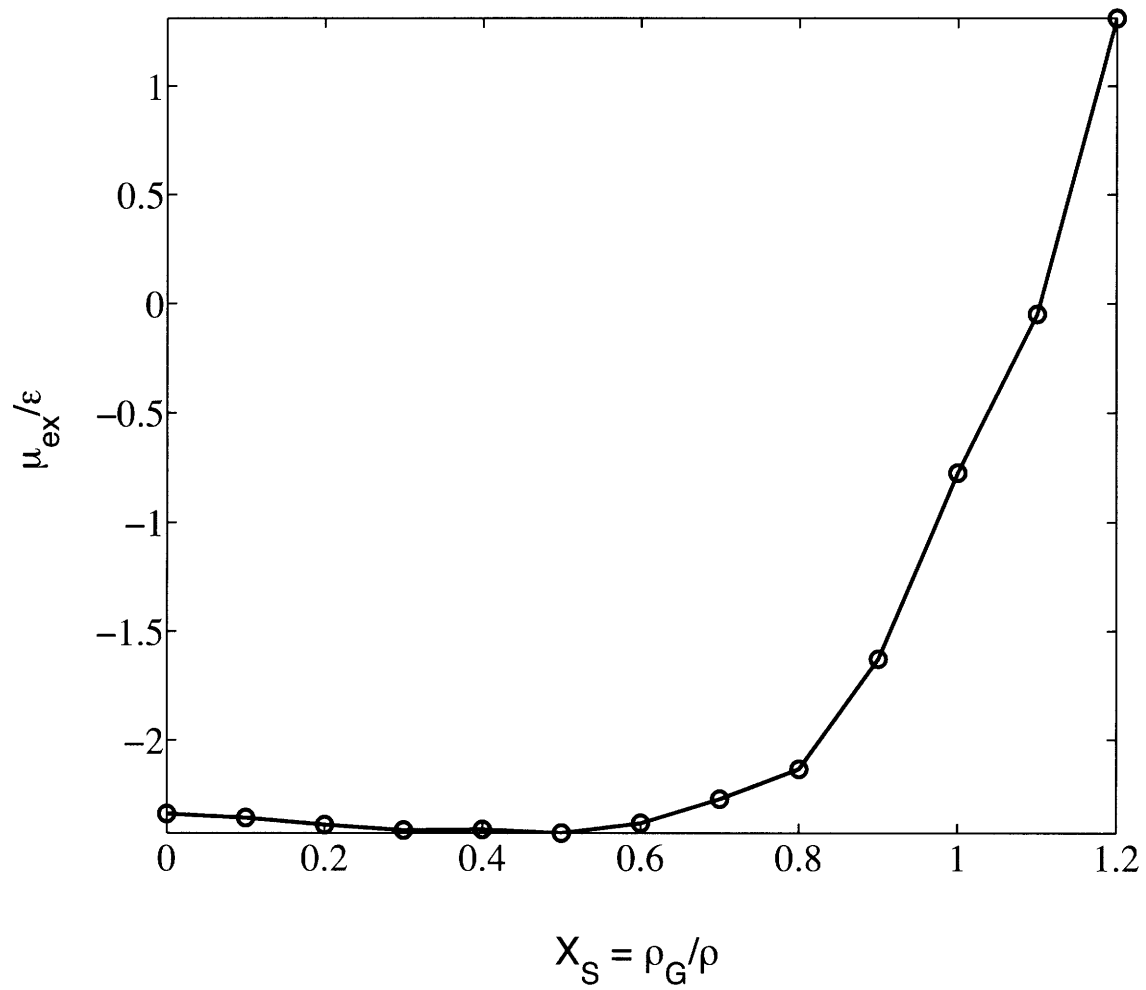
**Figure 2-1:** Comparison of free energy changes upon particle removal in the (a) all-particle and (b) implicit solvent, density-dependent cases. The density-dependent potential introduces a secondary free energy change due to the change in energy models associated with a change in global average solute density of the system.



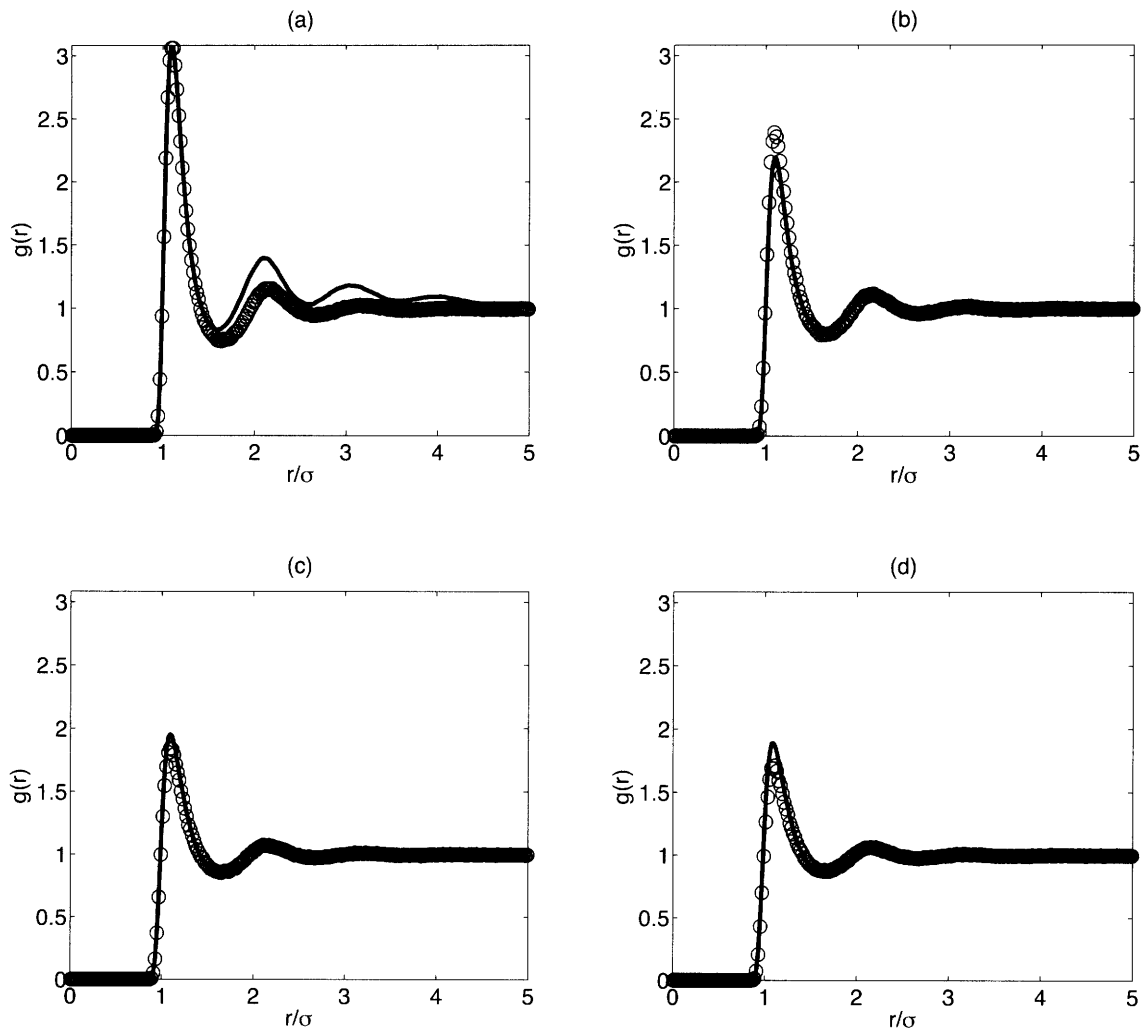
**Figure 2-2:** Comparison of the RDF generated using the converged potential for  $X_S=0.1$  (open circles) and the target RDF (solid line).



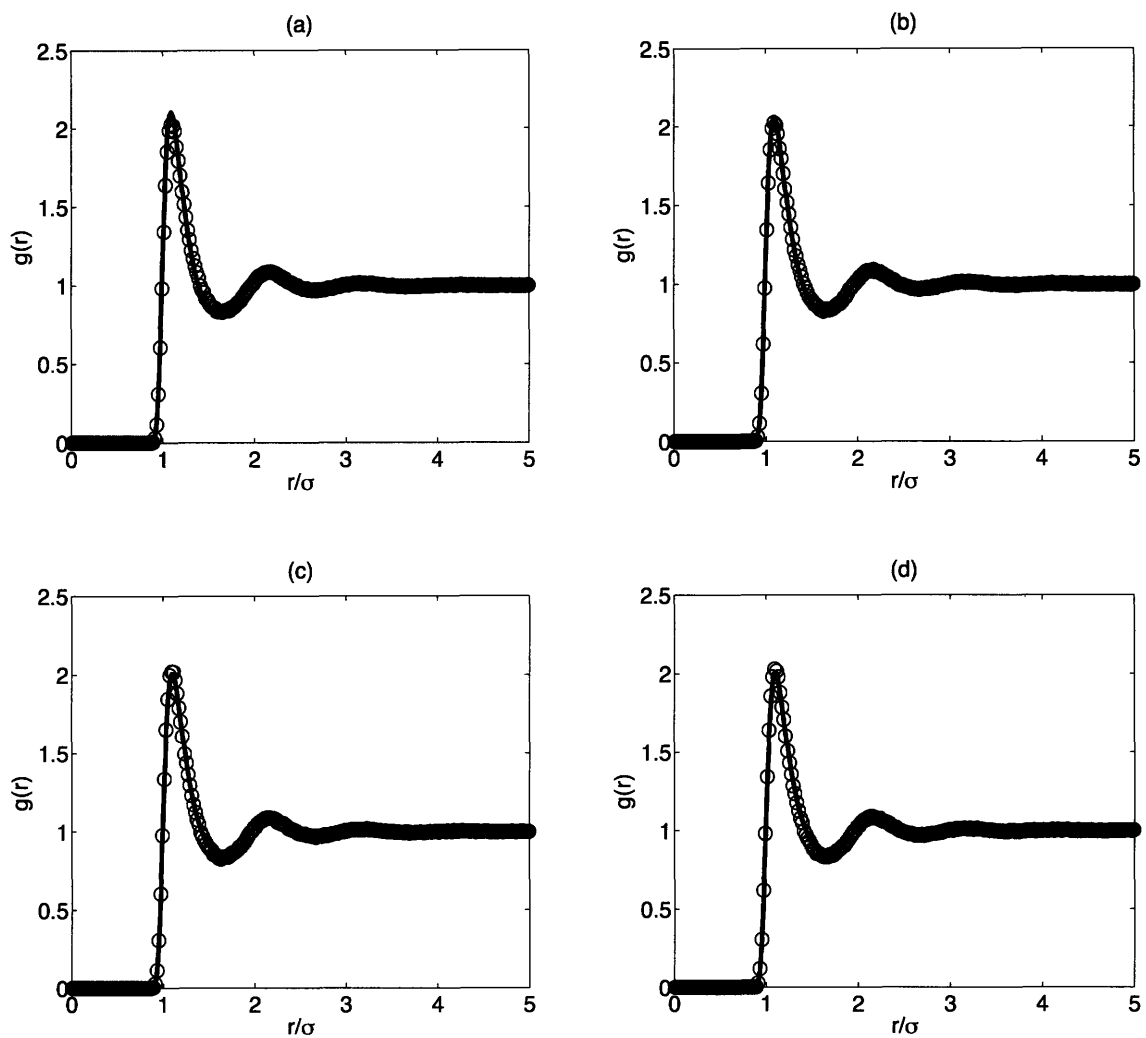
**Figure 2-3:** Pairwise potentials that reproduce the target RDF across a range of global solute densities,  $\rho_G$ . Bare Lennard-Jones potential shown for comparison as dark solid line (a) solute densities below the total particle density (—,  $X_S=0.10$ ;  $\cdots$ ,  $X_S=0.40$ ; ---,  $X_S=0.70$ ; -·-,  $X_S=0.90$ ) (b) solute densities above the total particle density (—,  $X_S=1.10$ ;  $\cdots$ ,  $X_S=1.20$ )



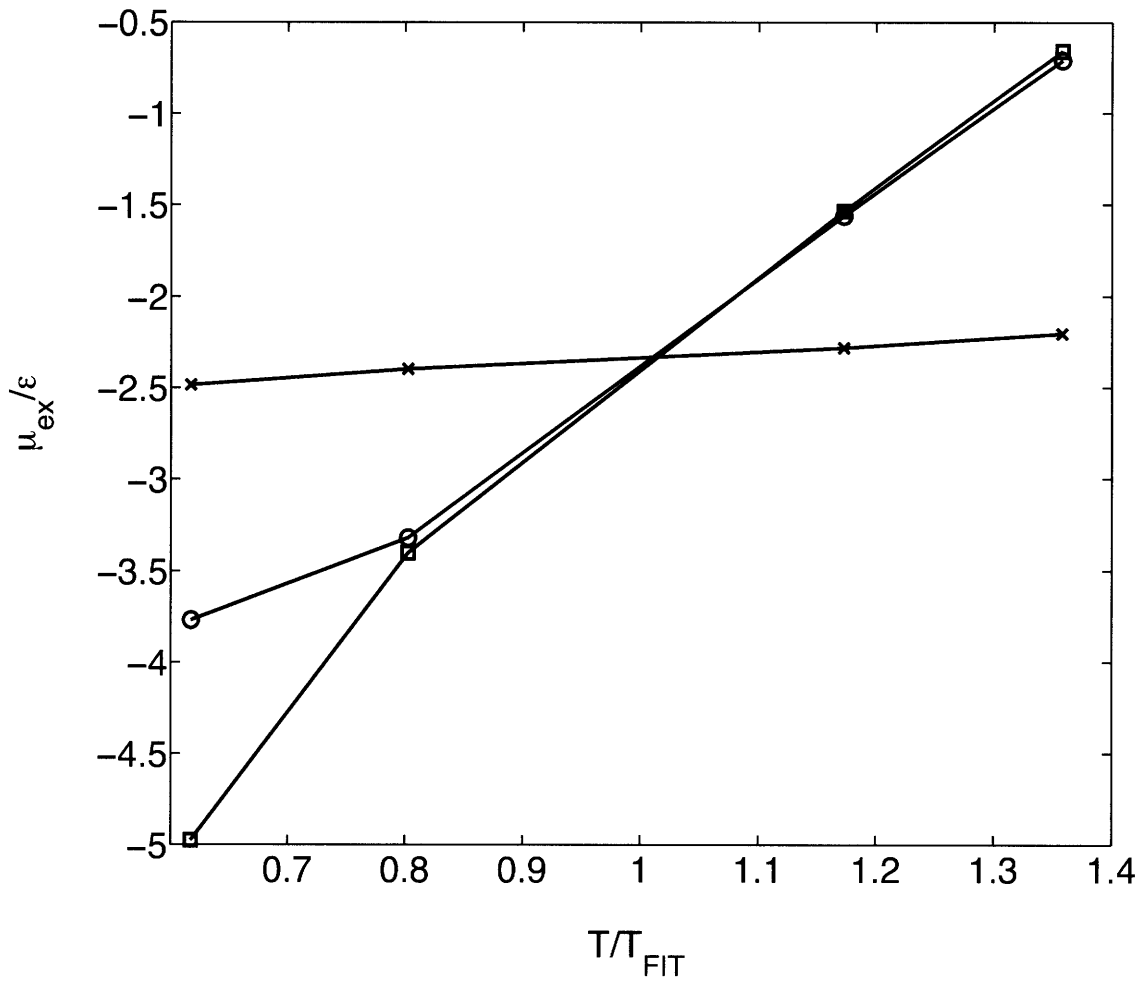
**Figure 2-4:** Self-interaction energy,  $\mu(\rho_G)$  for  $\rho^*=0.55, T^*=1.35$



**Figure 2-5:** Comparison of measured (open circles) and target (solid line) radial distribution functions at  $X_S=0.10$  for reduced temperatures (a)  $T^*=0.83$  (b)  $T^*=1.08$  (c)  $T^*=1.58$  (d)  $T^*=1.83$ . (a) and (b) are below the fitting temperature, (c) and (d) above.

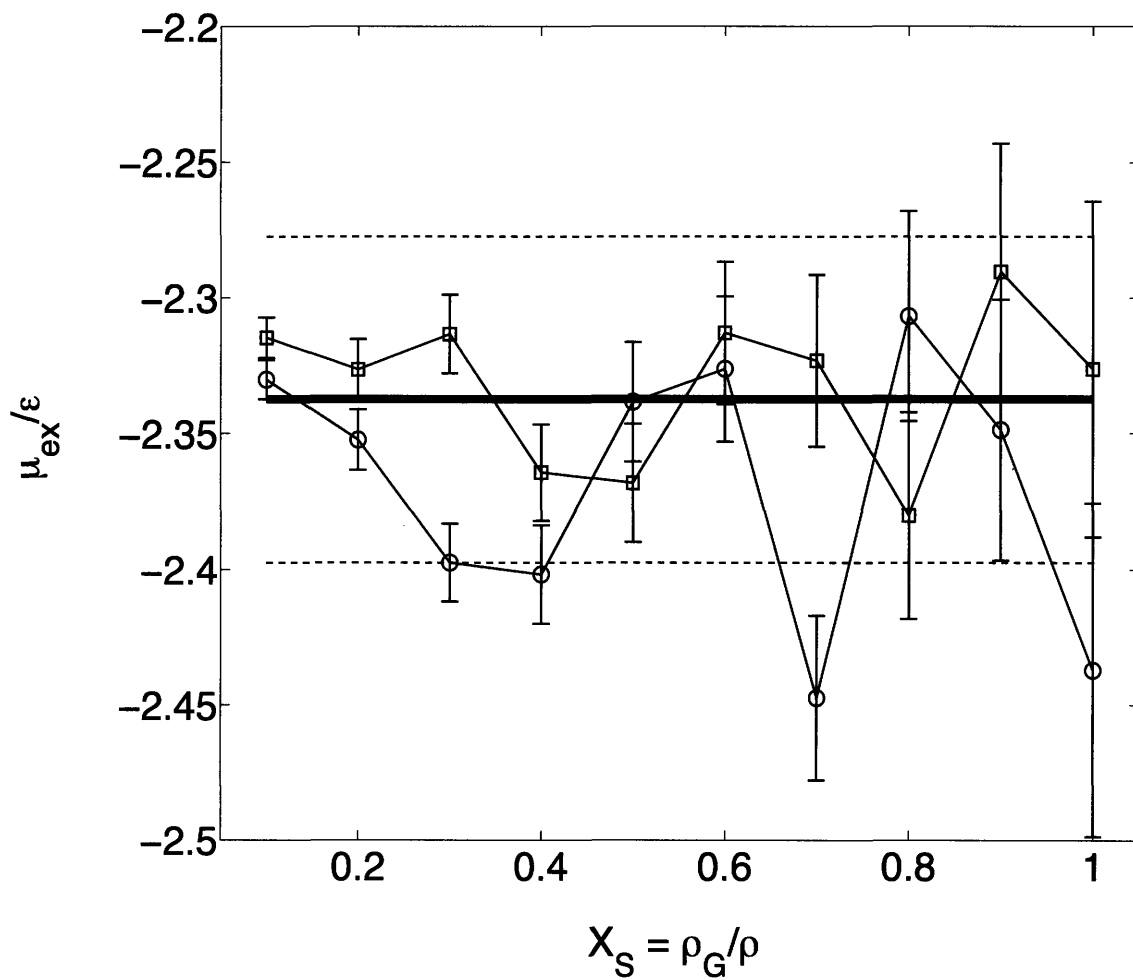


**Figure 2-6:** Comparison of measured (open circles) and target (solid line) radial distribution functions at  $X_S=0.10$  for system volumes (a)  $V/V_0=0.90$  (b)  $V/V_0=0.95$  (c)  $V/V_0=1.05$  (d)  $V/V_0=1.10$ . (a) and (b) are more dense than the fitting density, (c) and (d) less dense.

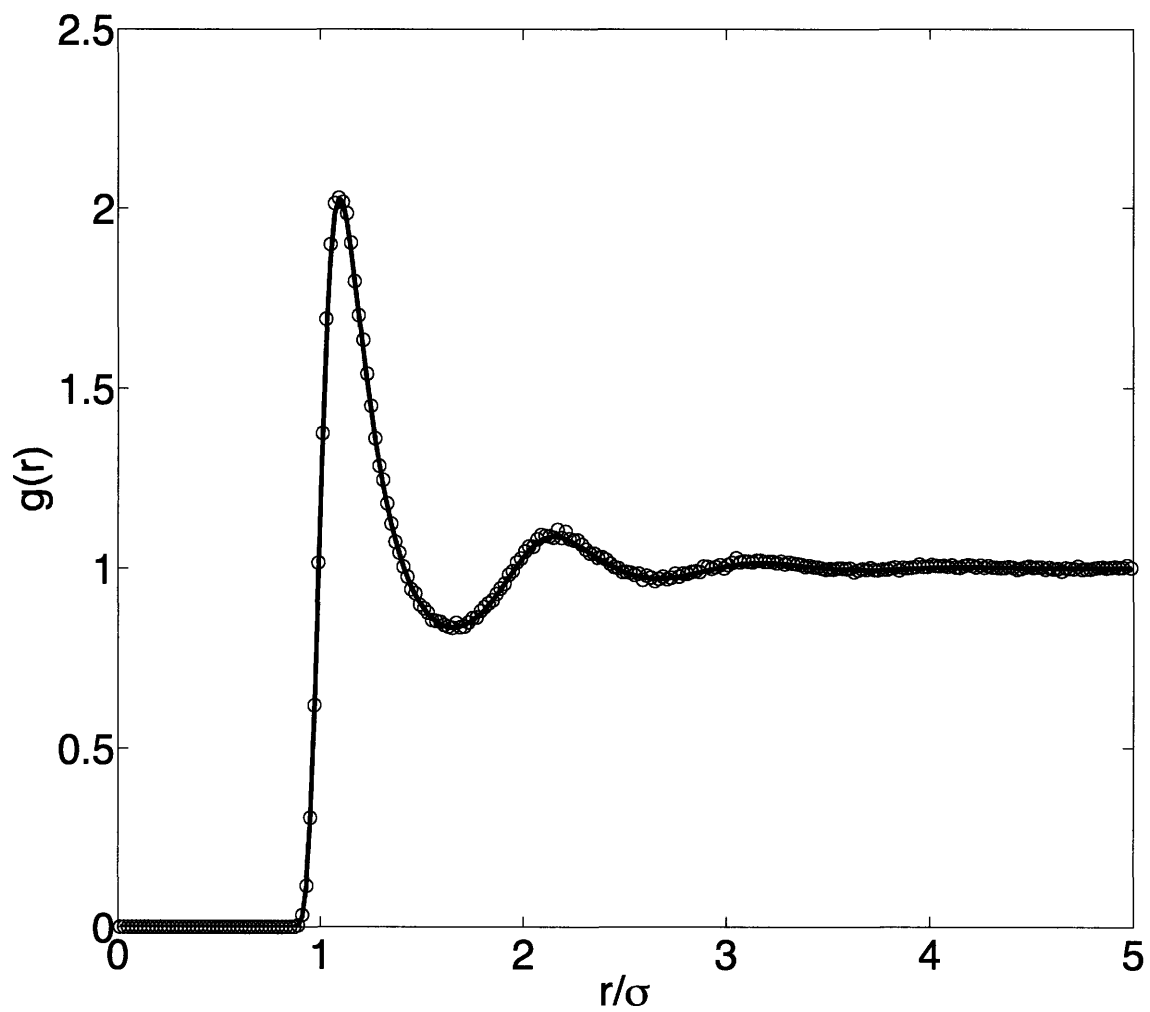


**Figure 2-7:** Excess chemical potential as a function of system temperature. (Circles, all-particle simulation results; squares,  $X_S=1.00$ ; crosses,  $X_S=0.10$ )

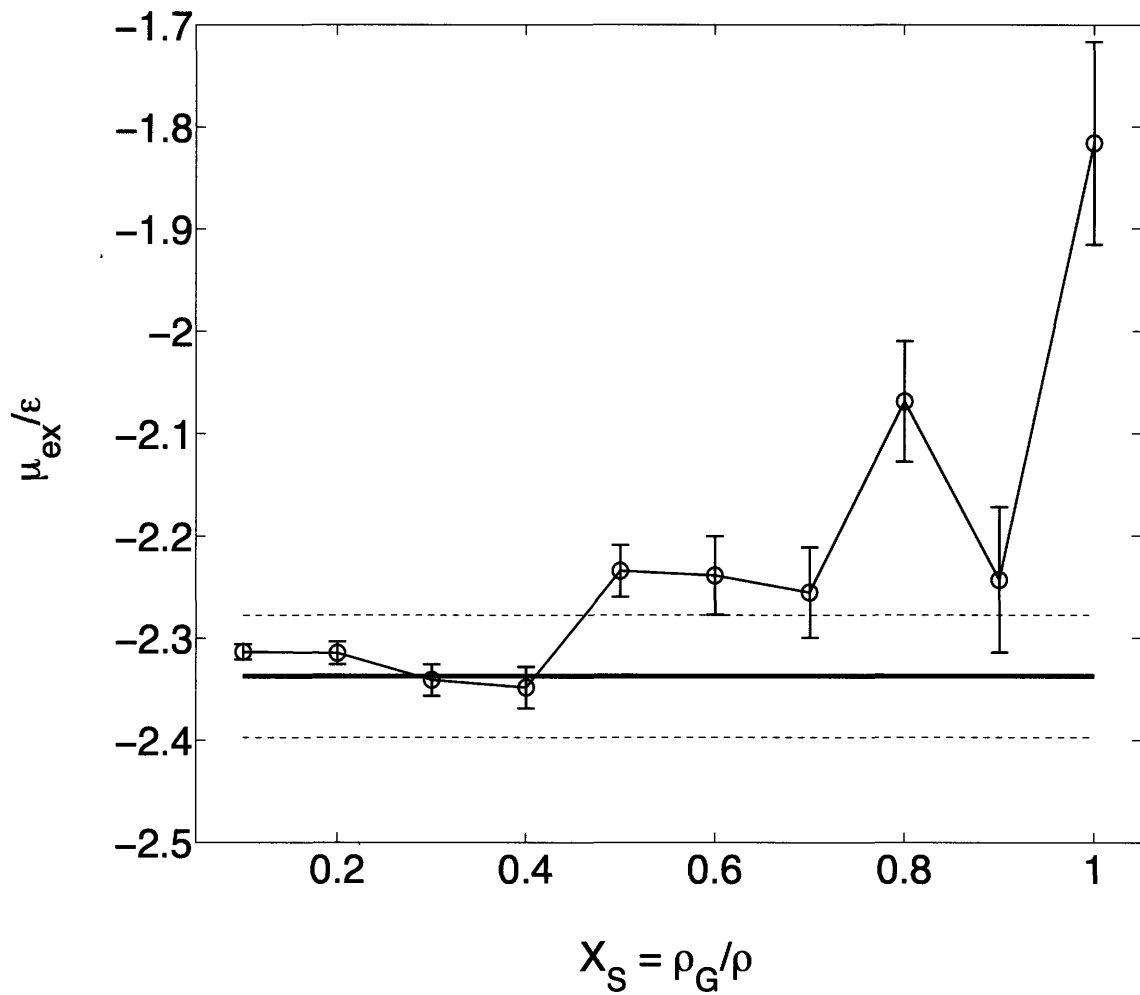




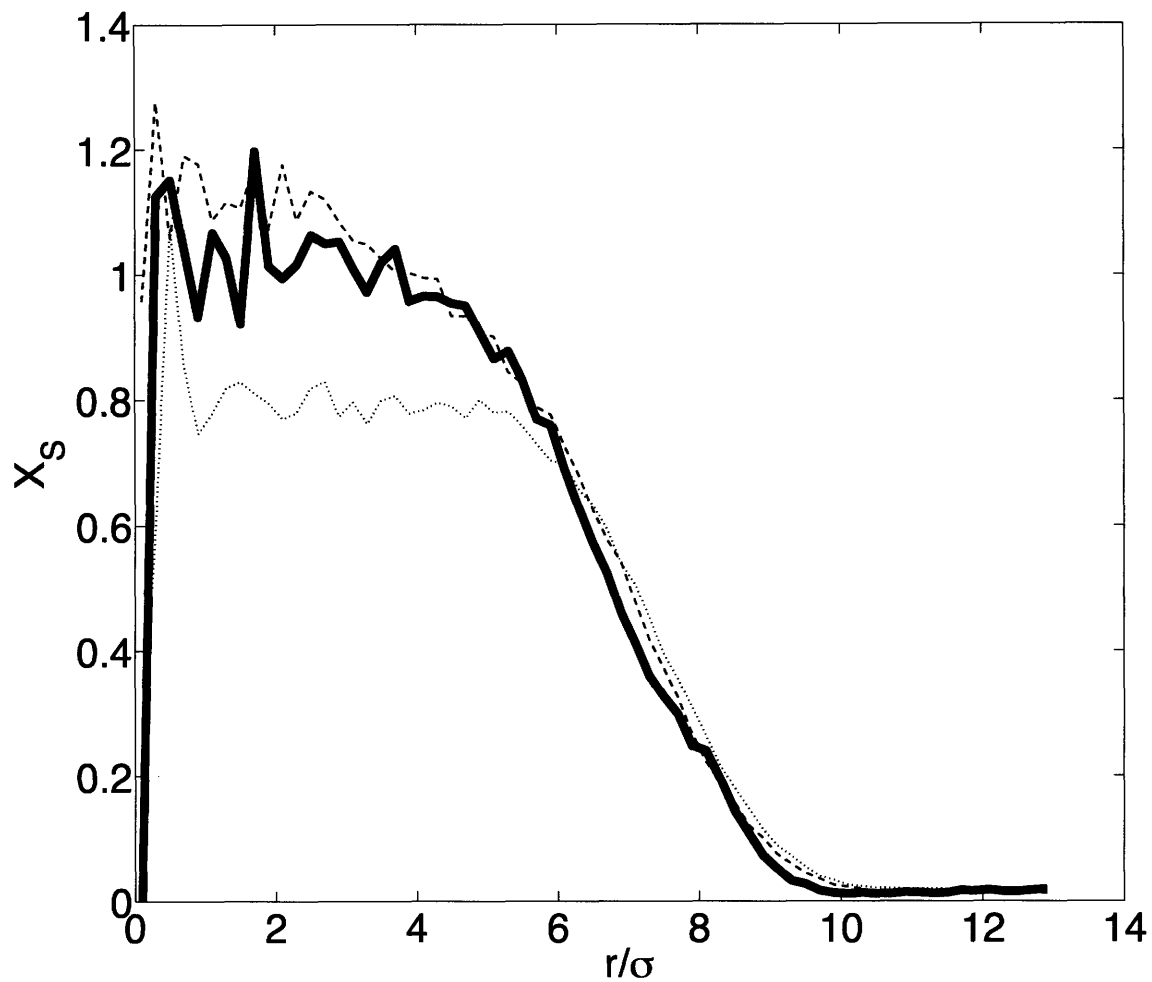
**Figure 2-8:** Comparison of simulated  $\mu^{ex}$  values with target value as a function of  $\rho_G$   
 (Circles, original simulation using global density values; squares, after optimization; solid line, all-particle simulation; dashed lines, desired simulation accuracy)



**Figure 2-9:** Comparison of simulated RDF with worst fit ( $X_S=1.00$ ) (open circles) and the all-particle RDF (solid line).



**Figure 2-10:** Comparison of simulated  $\mu^{ex}$  values with the exact value for the all-particle simulation as a function of  $\rho_G$  for  $r_D=2\sigma$  (Circles, simulation data; solid line, all-particle simulation; dashed lines indicate an error of  $0.06\epsilon$ )



**Figure 2-11:** Solute fraction as a function of distance from center of the box (—, all-atom; ---, local density approach; ···, global density approach)

Property	Value
$\epsilon = \epsilon_{Solute-Solute} = \epsilon_{Solute-Solvent} = \epsilon_{Solvent-Solvent}$	1
<i>Solvent</i>	
$\sigma = \sigma_{Solute-Solute} = \sigma_{Solute-Solute} = \sigma_{Solvent-Solvent}$	1
<i>Solute-Solute</i>	
$T^*$ (reduced temperature)	1.35
$\rho^*$ (reduced total particle density)	0.55
$N$ (number of particles)	1000
$r_c^*$ (cut-off radius)	5
$X_S$	[0.1,0.2...,1]

**Table 2-1:** Test system properties.

$T^*$	$(T^*-T^*_{fit})/$ $T^*_{fit}$	Local solute number error (%)	Per-particle energy error (%)
0.83	-39%	+8.0%	+5.0%
1.08	-20%	+0.3%	-2.8%
1.22	-10%	+0.1%	-1.3%
1.49	+10%	+0.1%	+1.3%
1.58	+17%	+0.2%	+2.0%
1.83	+36%	+0.2%	+3.0%

**Table 2-2:** Transferability of the original effective potential, as measured by the error in the per-particle energy with changing temperature.

$V/V_0$	Local solute number error (%)	Per-particle energy error (%)
0.90	+0.1%	-0.2%
0.95	+0.1%	+0.1%
1.05	+0.2%	+0.5%
1.10	+0.3%	+0.9%

**Table 2-3:** Transferability of the original effective potential, as measured by the error in the per-particle energy with changing density.

$\rho/\rho$	Case $r_D=5.0\sigma$		Case $r_D=2.0\sigma$	
	Local solute number error (%)	Per-particle energy error (%)	Local solute number error (%)	Per-particle energy error (%)
0.10	+0.1%	-0.0%	+0.4%	+1.9%
0.20	+0.1%	+0.4%	+0.3%	+2.9%
0.30	+0.1%	-0.6%	+0.3%	+2.6%
0.40	+0.2%	-0.3%	+0.4%	+2.3%
0.50	+0.2%	+0.5%	+0.2%	+0.6%
0.60	+0.1%	+0.3%	+0.2%	-0.2%
0.70	+0.2%	+0.2%	+0.2%	-1.2%
0.80	+0.1%	+0.2%	+0.2%	-1.2%
0.90	+0.2%	+0.5%	+0.2%	-1.0%
1.00	+0.3%	+0.6%	+0.2%	-0.4%

**Table 2-4:** Errors produced by use of a local density dependent potential. For the case  $r_D=r_c=5\sigma$ , errors are below the specified tolerance of 1%. However, a smaller cutoff radius  $r_D=2\sigma$  introduces significant errors in the per-particle energy.



### **3 Evaluating the Transferability of Coarse-Grained, Density Dependent Implicit Solvent Models to Mixtures and Chains**

#### **3.1 INTRODUCTION**

Coarse-graining techniques have received growing interest as methods to extend the time and length scales of molecular simulations. In its most essential form, a coarse-graining algorithm is a statistical fitting process that systematically reduces the number of degrees of freedom from that of an all-atom simulation. This reduction is achieved either by aggregating multiple atomic coordinates into a single coarse-grained (CG) particle or by the deletion of particles, as in an implicit solvent simulation. A coarse-graining scheme usually starts with a short simulation in which every particle is explicitly included; this is used to generate descriptive “data” about the exact system. This simulation is followed by a matching procedure in which the coarse-grained potential is created and applied to a system involving a reduced number of particles to reproduce the data of the underlying all-particle simulation. Various types of data have been used for CG potential fitting, including forces<sup>1-7</sup>, reversible work<sup>8</sup>, radial distribution functions<sup>9-12</sup> (RDFs), and experimental results<sup>13-15</sup>. Additionally, a wide variety of fitting procedures have been proposed, including simplex optimization<sup>16-21</sup>, radial distribution function inversion<sup>22-31</sup>, wavelet transform<sup>32,33</sup>, and semi-grand canonical Monte Carlo<sup>13</sup>.

The value of a coarse-grained potential is determined in large part by its utility, or “transferability”, to situations outside of the one to which it was fit, because

transferability is the only way that the upfront cost of performing the fit can be recouped. As suggested by Johnson *et al.*<sup>34</sup>, transferability can be classified as one of two types: “observable transferability”, and “state point transferability”. Observable transferability (called “representability” by Johnson *et al.*) refers to the ability of a CG potential that is fit to one set of simulation observables to reproduce accurately another set of simulation observables. State point transferability, by contrast, refers to ability of a CG potential that is fit at one thermodynamic state point (temperature, density, system size, and composition) to predict the same observables at other thermodynamic state points. Both types of transferability have been previously addressed in the literature.

Lyubartsev<sup>35</sup> examined the observable transferability of a CG potential for lipid molecules in water. The potential utilized RDF matching, and was subsequently shown to reproduce reasonably well the density profile of a lipid bilayer. Noid *et al.*<sup>6</sup> showed that for isotropic, homogeneous materials a force-matched potential will also reproduce the system RDF. Frequently, however, CG algorithms are used to simulate systems on time scales that are inaccessible via all-atom simulation<sup>35</sup>. In these cases, demonstrating observable transferability is impossible, since the all-atom simulation cannot be performed for purposes of comparison. Further, theoretical study by Louis<sup>36</sup> demonstrated that observable transferability breaks down even in simple cases; for example, he showed that potentials fit to system RDFs do not reproduce system energy, and *vice-versa*. As a result, he suggested that only more complex potential forms, incorporating density dependence or many-body interactions, may overcome these representability problems. Indeed, Merabia and Pagonabarraga<sup>37</sup> have demonstrated that

density dependent potentials avoid some theoretical representability problems. In previous work<sup>38</sup> we demonstrated that a coarse-grained implicit solvent potential containing density dependent one-body and two-body interactions can replicate both the system RDF and a measure of system energy across a range of solute concentrations.

State point transferability has been studied many times as well. Henderson's<sup>39</sup> seminal theoretical work indicates that the pairwise potential that generates a given RDF is unique to within an additive constant for systems with only pairwise interactions. However, the RDF of a system depends on the state point; it does not follow that the potential obtained by RDF inversion at one state point will generate the correct RDF at other state points. Recent simulations of polymers<sup>40</sup> by this approach have exhibited limited state-point transferability of such potentials. Louis<sup>36</sup> has suggested that state point transferability may be improved in systems containing a local density dependence, since these systems contain an extra parameter based on the local environment that a CG particle experiences. It is worth noting that improving one type of transferability is no guarantee of improving the other.

This chapter examines the state point transferability of the density-dependent implicit solvent (DDIS) potentials reported in Chapter 2 (This work is also published in Reference 38). There, the relative merits of density-dependent potentials parameterized on either the local or global solute density were explored; in this chapter, we focus primarily on the local density-dependent implicit solvent potential, abbreviated as DDIS unless noted otherwise. The fitting process used in generating the DDIS potentials guarantees the

transferability across a range of solute concentrations. This chapter examines their performance in other situations. Specifically, it examines transferability to two other cases – mixtures involving more than one type of solute particle and solutes comprising chains particles – in implicit solvent. These cases are of particular relevance for the study of surfactants, because transferability would imply that one need parameterize CG potentials only for the head and tail solute particles individually, thereby greatly extending their utility. Therefore, we examine here the limits of transferability of density dependent potentials for these cases.

## **3.2 THEORY**

### **3.2.1 Density Notation Conventions**

In this chapter,  $\rho_S$  refers to the *total* density of solute particles, where the subscript “S” stands for solute, and may include contributions from different solute types.  $\rho_A$ ,  $\rho_B$ , and  $\rho_C$ , respectively, are the densities of A, B, and C-type solute particles only. Finally,  $\rho$  without subscript refers to the state point density, considering all particles (solute and solvent). A superscript  $L$  indicates that the density is the *local* density of solute particles; otherwise, the density refers to the global average density.

### **3.2.2 DDIS Potential Review**

In Chapter 2, we proposed a density-dependent, implicit solvent (DDIS) potential with the form

$$E_i = \mu(\rho_{S,i}^L) + \frac{1}{2} \sum_{j \neq i} V^{EFF}(r_{ij}, \rho_{S,i}^L) \quad (1)$$

where  $E_i$  is the effective energy of particle  $i$ ,  $V^{EFF}$  is a pairwise potential between solute particles that is an explicit function of solute particle density in the vicinity of particle  $i$ ,  $\rho_{S,i}^L$ , and  $\mu$  is a “self-interaction” term that is also a function of solute particle density.

We also proposed a method to fit such potentials, such that the solute-solute RDF and solute excess chemical potential,  $\mu_A^{ex}$ , are reproduced across all solution compositions.

We present here only a brief sketch of the solution algorithm. For further details, the reader is referred to Chapter 2.<sup>38</sup>

The algorithm first involves generating pairwise potentials as a function of the global solute density by performing RDF-inversion<sup>22</sup> for a number of solute compositions. Assuming that the distribution of local densities is centered about the global density in these first simulations, the RDF-inversion potential obtained for a given global solute density can then be taken as representative of the potential to be applied for a particle that experiences a comparable local solute density, regardless of the actual global density of the system in which the particle is found. Once the pairwise potentials are determined, an iterative procedure is used to determine the self-interaction as a function of local solute density such that the solute excess chemical potential is reproduced across all compositions. The method does not guarantee fit to an arbitrary accuracy.

The measure of error in the RDF is given by the solute-solute energy, defined as:

$$E_{L,AA} = \rho_A \int_0^r_C V_{AA}(r) g_{AA}(r, \rho_A, \rho_B) 4\pi r^2 dr \quad (2)$$

where  $\rho_A$  is the density of solute type A,  $V_{AA}$  is the exact interaction potential between A particles, and  $g_{AA}$  is the A-A RDF. The difference between  $E_L$  for a coarse-grained system and that for the all-atom simulation provides a measure of the error in  $g(r, \rho_A, \rho_B)$ , relative to the exact  $g(r, \rho_A, \rho_B, \rho)$  for the explicit system, in units of energy. The error in the excess chemical potential,  $\mu_A^{ex}$ , is the difference (in units of  $\epsilon_Z$ , the Lennard-Jones parameter of the solvent) between the target (all-atom) and measured (coarse-grain) values. In this text, both error measurements are presented in terms of the standard error over all simulations. Additionally, we supply figures where appropriate comparing the all-atom and coarse-grained chemical potentials and RDFs, as these offer an intuitive sense of the closeness of fit.

### 3.2.3 Mixing Rule

Simulations of solute particle types A and B in solvent Z require a mixing rule to govern A-B interactions. We propose here a simple yet physically intuitive mixing rule for the two-body portion of the DDIS potential, which has the benefit of requiring no additional simulation. The one-body portion of the potential is left unmodified. There are certainly more complex mixing rules that one could propose; some of these choices are discussed in Appendix A. We begin by decomposing the A-A coarse-grained potential,  $V_{AA}^{EFF}$ , into two terms:

$$V_{AA}^{EFF}\left(\frac{r}{\sigma_A};\rho_A^L\right) = V_{AA}\left(\frac{r}{\sigma_A}\right) + \Delta V_{AA}^{EFF}\left(\frac{r}{\sigma_A};\rho_A^L\right) \quad (3)$$

where  $V_{AA}$  is the all-atom potential, and  $\Delta V_{AA}^{EFF}$  incorporates all modifications to the all-atom potential. A similar equation can be written for the B-type solute. We then assume that the modification of the A-B potential upon coarse-graining follows a simple mixing rule, such that

$$V_{AB}^{EFF}\left(\frac{r}{\sigma_{AB}};\rho_S^L\right) = V_{AB}\left(\frac{r}{\sigma_{AB}}\right) + \frac{\varepsilon_{AB}}{2} \left( \frac{\Delta V_{AA}^{EFF}\left(r/\sigma_A;\rho_S^L\right)}{\varepsilon_A} + \frac{\Delta V_{BB}^{EFF}\left(r/\sigma_B;\rho_S^L\right)}{\varepsilon_B} \right) \quad (4)$$

where  $\rho_S^L$  is the local solute density comprising both A and B, and  $V_{AB}$  is the A-B all-atom potential. Whereas in the single solute system,  $V_{AA}^{EFF}$  and the one-body term are both parameterized on the local density of A particles, in the combined system both are now assumed to be parameterized based on the total solute density  $\rho_S^L$ . Equation (4) includes explicit reference to  $V_{AB}$ , and therefore does not assume a particular mixing rule for the all-atom potential.

### 3.2.4 Limits of Transferability

Even if the mixing rules presented above were to produce the correct RDFs for every system composition, this does not in general guarantee the transferability with respect to the excess chemical potential. To see this, we extend the analytical framework of Chapter 2, demonstrated in Figure 3-1, to mixtures of two solute particles. Here, the

excess chemical potential of particle type A in an implicit solvent simulation with global solute densities  $\rho_A$  and  $\rho_B$  ( $\rho_S = \rho_A + \rho_B$ ), is given by:

$$\mu_A^{ex}(\rho_A, \rho_B) = \Delta F_{1,A}(\rho_A, \rho_B) + \Delta F_{2,A}(\rho_A, \rho_B) \quad (5)$$

where  $\Delta F_{1,A}$  is the free energy change associated with pairwise interactions between the inserted particle(s) and the system,

$$\begin{aligned} \Delta F_{2,A}(\rho_A, \rho_B) = & \mu_A(\rho_S^L = \rho_S) + (n_A - 1) \frac{\partial \mu_A(\rho_S^L)}{\partial n_A} + n_B \frac{\partial \mu_B(\rho_S^L)}{\partial n_A} + \\ & \frac{(n_A - 1)^2}{2} [AA(\rho_S^L)] + (n_A - 1)n_B [AB(\rho_S^L)] + \frac{n_B^2}{2} [BB(\rho_S^L)] \end{aligned} \quad (6)$$

$n = n_A + n_B$ , and

$$\left[ AA\left(\rho_S^L = \rho_S = \frac{n}{V}\right) \right] = \frac{1}{V} \int_0^{r_c} g_{AA}(r, \rho_A, \rho_B) \frac{\partial V_{AA}^{EFF}(r, \rho_S^L)}{\partial n} 4\pi r^2 dr \quad (7)$$

Here, as in Chapter 2, Equations (6) and (7) create a linkage between the excess chemical potential, a global property of the system depending on the global solute density, and the implicit solvent potential, which depends on the local solute density. We do so by assuming that the distribution of local solute densities is tightly peaked around the global solute density. This assumption generally holds for a large cut-off radius for the calculation of the local density and for solute particles that disperse fully in the solvent.



For systems with an uneven distribution of solute particles (for example, in micellar systems), this assumption can be relaxed by measuring the average local solute density for a given global solute density.

The excess chemical potentials calculated by particle insertion for both the all-atom and coarse-grained potentials contain free energy contributions associated with the pairwise interactions between the inserted solute particle and the other particles in the system (denoted  $\Delta F_1'$  and  $\Delta F_1$ , respectively, in Figure 3-1). The correction term, Equation (6), is unique to the DDIS framework, and arises because insertion of a solute particle changes the density (and hence energetic interactions) of the system. This small change in solute density introduces a differential change in the interactions between existing particles in the system, which nonetheless produces a significant impact on the excess chemical potential for high solute densities. Because Equation (6) has a dependence on  $\mu_B^{ex}(\rho_S)$ , transferability cannot be assured, since the A potential was fit without knowledge of B-type particles.

There are, however, situations in which transferability with respect to  $\mu^{ex}$  can be expected, as will be discussed next. These special cases cover a large number of relevant simulations.

### 3.2.5 Special Cases

Transferability with respect to excess chemical potential is more likely for solutions at low total solute concentration. Equation (6) for  $\Delta F_{2,A}$  shows that as  $n_B$  becomes small,

those terms involving B-type particles will vanish. As these terms vanish,  $\mu_A^{ex}$  depends only on terms involving A particles. Since these are the conditions under which the A potential was fit,  $\mu_A^{ex}$  should be accurately reproduced. A similar argument can be made for B-type particles.

A second special case arises when A-type and B-type particles are inserted simultaneously, *in proportion to their existing ratios in the system*. The most common example of such a system is insertion of an  $A_iB_j$  chain into a system of identical chains. We assume here that the free energies of insertion are additive. In that case, the average chemical potential of each inserted particle is:

$$\mu^{ex} = \frac{n_A}{n} \mu_A^{ex} + \frac{n_B}{n} \mu_B^{ex} \quad (8)$$

where  $n_A$  and  $n_B$  are the number of A-type and B-type particles in the system, respectively, and  $n = n_A + n_B$ . By analogy to Equation (5)-(7), Equation (8) can be written as:

$$\mu^{ex} = \frac{n_A}{n} \Delta F_{1,A}(\rho_A, \rho_B) + \frac{n_B}{n} \Delta F_{1,B}(\rho_A, \rho_B) + \frac{n_A}{n} \Delta F_{2,A}(\rho_A, \rho_B) + \frac{n_B}{n} \Delta F_{2,B}(\rho_A, \rho_B) \quad (9)$$

with

$$\Delta F_{2,A}(\rho_A, \rho_B) = \left[ \mu_A(\rho_S) + (n_A - 1) \frac{\partial \mu_A(\rho_S)}{\partial n} + n_B \frac{\partial \mu_B(\rho_S)}{\partial n} \right] + \left[ \frac{(n_A - 1)^2}{2} \left[ AA\left(\frac{n}{V}\right) \right] + (n_A - 1)n_B \left[ AB\left(\frac{n}{V}\right) \right] + \frac{n_B^2}{2} \left[ BB\left(\frac{n}{V}\right) \right] \right] \quad (10)$$

and

$$\Delta F_{2,B}(\rho_A, \rho_B) = \left[ \mu_B(\rho_S) + (n_B - 1) \frac{\partial \mu_B(\rho_S)}{\partial n} + n_A \frac{\partial \mu_A(\rho_S)}{\partial n} \right] + \left[ \frac{(n_B - 1)^2}{2} \left[ BB\left(\frac{n}{V}\right) \right] + (n_B - 1)n_A \left[ AB\left(\frac{n}{V}\right) \right] + \frac{n_A^2}{2} \left[ AA\left(\frac{n}{V}\right) \right] \right] \quad (11)$$

If we assume first that  $AA(n/V)$  and  $BB(n/V)$  are not greatly different from their pure component values, and second that  $AB(n/V)$  is an average of  $AA(n/V)$  and  $BB(n/V)$ , then Equations (9)-(11) can be simplified to

$$\begin{aligned} \mu^{ex} = & \frac{n_A}{n} \Delta F_{1,A}(\rho_A, \rho_B) + \frac{n_B}{n} \Delta F_{1,B}(\rho_A, \rho_B) + \\ & \frac{n_A}{n} \left[ \mu_A(\rho_S) + (n-1) \frac{\partial \mu_A(\rho_S)}{\partial n} + \frac{(n-1)^2}{2} \left[ AA\left(\frac{n}{V}\right) \right] \right] + \\ & \frac{n_B}{n} \left[ \mu_B(\rho_S) + (n-1) \frac{\partial \mu_B(\rho_S)}{\partial n} + \frac{(n-1)^2}{2} \left[ BB\left(\frac{n}{V}\right) \right] \right] \end{aligned} \quad (12)$$

The bracketed terms in Equation (12) are simply the correction terms derived for pure solute system,  $\Delta F_{2,A}(\rho_S)$  and  $\Delta F_{2,B}(\rho_S)$ . Therefore, in this special case, the average excess chemical potential will be correct.

### 3.2.6 Solute Enhancement Ratio

One way to characterize the local environment is to calculate the local number of solute particles,  $\langle N_L \rangle$ , defined as:

$$\langle N_L \rangle = \rho_s \int_0^{R_c} g(r; \rho_s) 4\pi r^2 dr \quad (13)$$

We define the solute enhancement ratio as the number of solute particles within a sphere of radius  $R_c$  divided by the average number of particles in such a volume. A ratio near 1.0 indicates that the local solute environment is very similar to the global average environment. We find this metric particularly useful in the discussion of chain molecules.

### 3.3 SIMULATION PROTOCOL

Simulation protocols for all-atom and coarse-grained simulations of monomeric solute are given in Chapter 2. As in that paper, all simulation were carried out in the  $NVT$  ensemble at  $T^* = k_B T / \epsilon_Z = 1.35$  and  $\rho^* = \rho \sigma_Z^3 = 0.55$ , where  $\epsilon_Z$  and  $\sigma_Z$  are the Lennard-Jones parameters of the solvent  $Z$ . Each simulation consisted of 1000 total particles. The cutoff radius for interactions was set to  $5\sigma_Z$  unless otherwise indicated.

All-atom simulations of chains were carried out in the same manner as the monomeric all-atom simulations, with the exception of intramolecular degrees of freedom for the chain simulations. Bond lengths and angles were held fixed at the minimum energy values using the LINCS<sup>41</sup> algorithm, while torsion angles were allowed to rotate freely. Chains of length two and four were simulated.

Implicit solvent simulations were carried out in the same manner as the monomeric density dependent, implicit solvent simulations. Nearest bonded neighbor particles were included in the calculation of local density. In addition to the chain translation moves used for monomeric simulations, simulations of dimers included rigid body rotation moves as well, in the proportion 80% translation:20% rotation. For tetramers, we also included rotation moves about individual bonds, in the proportion 50% translation:20% rigid body rotation:30% bond rotation. Simulations of monomeric solutes were equilibrated for  $10^4$  cycles, followed by sampling for  $3 \times 10^4$  cycles. Sampling of dimers was performed for twice as many cycles, or  $6 \times 10^4$ , and chains of length four or longer were sampled for  $1.2 \times 10^5$  MC cycles. Free energy sampling was performed every two MC cycles, which was sufficient to generate statistically independent samples, as determined by the autocorrelation function of the measured free energy. RDF sampling was performed every 100 MC cycles.

### 3.4 RESULTS AND DISCUSSION

#### 3.4.1 Particle Types Used in this chapter

In the all-atom simulations described in this chapter, particles interact via the truncated and shifted Lennard-Jones potential, where the potential between particles  $i$  and  $j$  is described by:

$$V_{ij}^{TS}(r_{ij}; \epsilon_{ij}, \sigma_{ij}, R_{c,ij}) = \begin{cases} V_{ij}^{LJ}(r_{ij}; \epsilon_{ij}, \sigma_{ij}) - V_{ij}^{LJ}(R_{c,ij}; \epsilon_{ij}, \sigma_{ij}) & r_{ij} \leq R_{c,ij} \\ 0 & r_{ij} > R_{c,ij} \end{cases} \quad (14)$$

with

$$V_{ij}^{LJ}(r_{ij}; \epsilon_{ij}, \sigma_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (15)$$

where  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the Lennard-Jones parameters for  $ij$  interactions. Equations (14) and (15) allow for the possibility of different cut-off radii ( $R_{C,ij}$ ) for interaction between different particle types.  $\epsilon$  and  $\sigma$  values of unlike particles are governed by Lorentz-Berthelot mixing rules.

In Chapter 2, we examined the behavior of a single solute type (hereafter referred to as “A”) in solvent (“Z”) at reduced temperature and density  $T^*=1.35$  and  $\rho^*=0.55$ . Because the A-type particles were identical to the solvent Z, the behavior of A-Z mixtures was identical for all compositions of A. For the work presented here, we find it useful to introduce two additional solute types, which display “solvent-philic” and “solvent-phobic” behavior, respectively.

A solute’s relative preference for the solvent can be measured by the free energy of transfer,  $\Delta G_S(Z \rightarrow S)$ , defined as the free energy change associated with transferring a single solute particle from a bath of solvent particles Z to a bath of solute particles S. The free energy of transfer can be calculated from the excess chemical potential of solute particles:

$$\Delta G_s(Z \rightarrow S) = \mu_s^{ex} \left( \frac{\rho_s}{\rho} = 1 \right) - \mu_s^{ex} \left( \frac{\rho_s}{\rho} = 0 \right) \quad (16)$$

By varying particle  $\epsilon_{ii}$ 's and  $R_{C,ij}$ 's, we created a “solvent-philic” particle (type B) and a “solvent-phobic” particle (type C), with interaction parameters given in Table 3-1. Solute type B is distinguished by its reduced  $\epsilon_B$  of 0.5, compared to  $\epsilon_Z$  of 1.0. Solute type C interacts with the solvent Z via a reduced cutoff radius of  $1.84\sigma_Z$ , with C-C and Z-Z interactions maintaining the usual  $5.0\sigma_Z$  cutoff.

Table 3-1 shows the solute enhancement ratio for all three types of solute particles at  $\rho_s/\rho=0.5$ . The specific  $\epsilon_{ii}$ 's and  $R_{C,ij}$ 's used were selected such that B-type and C-type solute enhancement ratios were close to 1.0 for all compositions. This indicates that both solute types, while expressing relative preference for like or unlike interactions, are completely miscible in solvent at all compositions, and validates the use of Equations (6)-(7).

### 3.4.2 DDIS Potentials

Chapter 2 previously reported the DDIS potential for the coarse-graining of A-type particles in solvent Z, where the A and Z particles were identical. Here, we followed the same fitting procedures to produce DDIS potentials for B-type and C-type particles in the same solvent Z. Figure 3-2 shows the fitted values of the excess chemical potential compared to the all-atom simulations. Figure 3-3 shows the worst fit RDFs for each system. In general, the accuracy of fit achieved for the B-type particles is comparable to

the results for A-type particles reported in Chapter 2, and both are superior to that obtained for the C-type potentials. The standard error in  $\mu^{ex}$  for B-type particles was  $0.04\epsilon_Z$ , versus  $0.11\epsilon_Z$  for C-type particles. Similarly, the  $E_L$  standard error for B-type particles was  $0.005\epsilon_Z$ , versus  $0.06\epsilon_Z$  for C-type particles. The worst case RDFs for the B-type and C-type particles are shown in Figure 3-3. The lower quality of fit attained for the C-type particles is perhaps because of the large discrepancy in cut-off radii between like and unlike particles. In both cases, however, the visual examination suggests the fit is quite good.

Figure 3-4 compares the two-body portion of the coarse-grained potentials for the A-type, B-type, and C-type potentials at a local density of  $\rho_s^L/\rho=0.5$ . All three potentials display the same general form, with a secondary local maximum in the potential. The “solvent-philic” B-type particles show a shallower attractive well when compared to the A-type particles, while the “solvent-phobic” C particles have a deeper well. This is consistent with the form of the pairwise implicit solvent potentials used in the Effective Energy Function<sup>42</sup>.

Figure 3-5 shows the one-body portion of the DDIS potential for A-, B-, and C-type particles. As with the two-body portion, all the potentials share certain general characteristics. As solute density goes to zero, the value of the one-body term approaches  $\mu_s^{ex}(\rho_s/\rho=0)$ . The profile is relatively flat for low solute densities, before rising rapidly as the solute fraction approaches 1.



### 3.4.3 Mixtures

We investigated the behavior of A/B and A/C mixtures in implicit solvent Z. The analysis of this system is simplified by the fact that the A and Z particles are identical, which means that the system can be expressed solely as a function of  $\rho_B$  for the A/B system, and of  $\rho_C$  for the A/C systems. Considering A/X/Z systems, where X=B or C, is equivalent to considering A/X systems with partially coarse-grained A.

In this chapter we performed simulations in increments of 10% in solute concentration. As a result, there were 55 possible combinations of A/X/Z mixtures. To reduce computational time, we simulated only a representative fraction of these systems. That subset consisted of systems in which the total solute density,  $\rho_S$ , was 20%, 50%, or 90% of the total particle density, and constitutes 13 of the 55 mixture fractions possible. Additionally, we measured only  $\mu_X^{ex}(\rho_A, \rho_X)$  (and not  $\mu_A^{ex}(\rho_A, \rho_X)$ ) for each system, to further reduce computational time.

The ability of the DDIS potential to reproduce the all-atom RDF's for both systems was very good. For the A/B/Z mixtures, the standard errors in  $E_L$  were  $E_{L,AA}=0.013\epsilon_Z$  and  $E_{L,BB}=0.005\epsilon_Z$ , which compare favorably with the pure-component errors of  $E_{L,AA}=0.015\epsilon_Z$  and  $E_{L,BB}=0.005\epsilon_Z$ . In other words, the RDF accuracy of the coarse-grained mixtures is comparable to the accuracy of the coarse-grained pure systems. Results were similar for the A/C/Z mixtures. Here, the standard errors were  $E_{L,AA}=0.010\epsilon_Z$  and  $E_{L,CC}=0.06\epsilon_Z$ , compared to the pure component errors of  $E_{L,AA}=0.015\epsilon_Z$  and  $E_{L,CC}=0.06\epsilon_Z$ . Figure 3-6 shows the worst reproductions of all-atom RDFs by the

DDIS potentials for the A/B/Z and A/C/Z potentials, which demonstrates visually the similarity of RDFs.

Figure 3-7 examines the transferability with respect to excess chemical potential for the A/B/Z and A/C/Z mixtures at the three total solute concentrations. As expected, the transferability with respect to excess chemical potential is not as good as that with respect to the RDF, for the reasons laid out in the theory section. The standard error in  $\mu_B^{ex}(\rho_A, \rho_B)$  for the A/B/Z system is  $0.47\epsilon_Z$ , and the standard error in  $\mu_C^{ex}(\rho_A, \rho_B)$  for the A/C/Z system is  $0.63\epsilon_Z$ , which are both substantially higher than the pure-component errors of  $0.04\epsilon_Z$  and  $0.11\epsilon_Z$ . Nevertheless, the performance of the DDIS potentials at low solute concentration is quite good. The chemical potential errors for  $\rho_S/\rho=0.2$  are  $0.01\epsilon_Z$  and  $0.21\epsilon_Z$ , which are comparable to the error in the pure component systems.

### 3.4.4 Chains

To study the transferability of DDIS potentials to systems of chains, we examined the behavior of dimers and tetramers of A, B, and C solutes in solvent Z. These simulations were performed over a range of solute densities from  $\rho_S/\rho=[0.1, 1.0]$  in increments of 0.1, which were the same solute densities at which the potential was fit. Bond lengths were held fixed at  $1.22\sigma_Z$ , which is equal to the average separation between particles in the monomeric simulations. We selected this bond length so that the total local density of particles (solute and solvent) remained close to the fitting density. For shorter chain lengths, total local density was higher than in the monomeric case. Angles were held fixed at 112 degrees. Errors in  $\mu^{ex}$  were measured on a per-particle basis for chain

simulations, to permit comparison to the errors measured in monomeric solute simulations.

We first examined the all-atom and DDIS errors in  $E_L$  for monomers, dimers and tetramers of A, B, and C. These results are summarized in Table 3-2. For all three particle types, the errors in the reproduction of all-atom RDFs increased with increasing chain length. It appears that, for the systems studied here, each doubling of the chain length results in an approximate doubling in the error of  $E_L$ , although the progression is not entirely smooth. For example, the accuracy appears nearly equal for dimers and tetramers of B. Figure 3-8 shows the worst case reproductions of all-atom RDFs by the DDIS potentials for tetramers for each particle type. At the tetramer level, one can begin to see visual disagreement between the all-atom and coarse-grained RDFs. In general, the locations of the peaks in the RDF remain correct. Errors arise in maxima and minima of  $g(r)$  that are systematically too high or too low.

Figure 3-9 compares the all-atom and DDIS results for  $\mu^{ex}$  per particle for monomers, dimers and tetramers of A, B, and C. The standard errors are given in Table 3-3. For all particle types, the errors increased with increasing chain length. As is the case in mixtures, chemical potential transferability is better at low solute concentrations. The reason that transferability breaks down at higher solute concentrations is related to the two-body correction factor, Equation (7), which contains the solute-solute RDF. As chain length increases, the solute-solute RDF becomes less similar to the RDF at the

fitting conditions, and as a result transferability degrades. At low solute densities, the influence of the two-body correction is smaller, and  $\mu^{ex}$  transferability is better.

In Chapter 2, we demonstrated that a potential with local density dependence performed better in reproducing the RDF behavior of a system with solute aggregation than did a potential with a global density dependence. We hypothesized that this was because the local density dependence permitted the potential to be responsive to local variations in the environment around the solute at every time point of the simulation. In this chapter, we tested if this result was applicable to other systems with solute aggregation, namely dimers and tetramers of C-type solute in solvent Z. As discussed previously, C-type particles are completely miscible in solvent Z in the monomer state. However, longer chains of C-type particles aggregate substantially. At  $\rho_s/\rho=0.5$ , the solute enhancement ratio for monomers is 1.01, indicating a nearly homogeneous dispersion. For dimers, the solute enhancement ratio rises to 1.09, and to 1.21 for tetramers. This solute enhancement cannot be attributed to the presence of solute particles held in close proximity by bonding, as chains of A-type particles do not experience an equivalent increase in solution enhancement ratio. Instead, the local composition is more concentrated in solute than is the system in which the coarse-graining was performed due to chain aggregation. Aggregation is expected because increasing the chain length reduces molecular entropy, and thus the driving force for solute particle dispersion in solution.

Figure 3-10 compares the performance of the local DDIS potential of C-type solutes dispersed in Z to the equivalent global density-dependent potential. The global potential was tuned to fit the RDF of monomeric solute particles at a given solute density, and then tested on chains at the same solute density. The local potential shows superior RDF transferability when compared with the global potential. The slope of the standard error line as a function of chain length is smaller for the local potential, and for tetramers the absolute value of the error is smaller than for the global potential.

Figure 3-11 examines the all-atom RDF at a given solute density, and compares the results to those obtained using the local and global potentials. The results demonstrate that the global density-dependent potential does not correctly capture the solute aggregation effect. The value of  $g(r)$  is systematically too high for all  $r$ . For the composition shown in Figure 3-11, the all-atom ratio is 1.47, and the local potential yields a ratio of 1.43. By contrast, the global potential yields a solute enhancement ratio of 1.76.

### **3.4.5 Chains of Mixed Composition**

Finally, we examined a system of chains of mixed composition. This system invokes a combination of mixture and chain transferability. The surfactants studied here are  $A_1B_1$  molecules in solvent Z. The bond length is set to  $1.22\sigma_Z$ , as in the foregoing section on chains, and is held constant throughout the simulation.

Transferability with respect to  $\mu^{ex}$  was quite effective, with a standard error in  $\mu^{ex}$  of  $0.13\epsilon_Z$ , as shown in Figure 3-12. This value is substantially below the errors found in mixtures of monomeric A and B ( $0.47\epsilon_Z$ ) and validates our predictions from the theoretical section. The RDF transferability is also good, with standard errors in  $E_{L,AA}$  and  $E_{L,BB}$  of  $0.012\epsilon_Z$  and  $0.008\epsilon_Z$ , respectively. These results are comparable to the standard error at the state point of fit.

### 3.5 CONCLUSION

In this chapter, we developed coarse-grained DDIS potentials for pure solute particles, and investigated their transferability to mixtures and chains. For mixtures, transferability with respect to the solute-solute RDF is very good. For the mixture systems studied in this chapter, the RDF standard errors were equal to those at the state point of fitting, indicating no loss of accuracy. The transferability with respect to  $\mu^{ex}$  in mixtures is good for low solute concentrations, but at high solute concentrations the transferability breaks down. This breakdown is due to the unique nature of the DDIS potential, in which particle insertion causes a change in local density for some fraction of the existing system particles. The parameters of the DDIS potential are carefully tuned to account for this energy change at the state point of fitting. At other compositions, the energy change is different; as a consequence, the predictions of excess chemical potential are less accurate.

For the systems of chains studied, we found that errors in RDF fitting increased linearly with chain length. Increasing chain length was also associated with increases in the  $\mu^{ex}$  error. Additionally, we find that the local DDIS potential produced superior RDF

transferability when compared to comparable global density-dependent implicit solvent potentials for a chain system with solute aggregation. The most likely reason for this improved transferability is that the local density dependence captures solute enhancement effects more accurately than the global potential. Finally, we found that chains of mixed composition had  $\mu^{ex}$  transferability that was superior to mixtures of monomers of equivalent composition. This is because the simultaneous insertion of *both* particle types removes the biases inherent in mixed systems.

The transferability characteristics described in this chapter indicate that DDIS potentials can function effectively for a wide variety of systems, including mixed composition chain molecules at low concentrations. Given these results, we feel confident in suggesting that the potentials would be useful in the simulation of surfactant systems. We intend to report results for such systems in the near future.

### 3.6 APPENDIX

The performance of the basic mixing rule described in this chapter is generally good. The lone exception is in its performance with respect to  $\mu^{ex}$  in systems of mixtures. We discuss here a more complex mixing rule that can improve transferability of potentials.

Equations (5)-(7) provide a framework to predict the excess chemical potential,  $\mu_A^{ex}(\rho_A, \rho_B)$ , based on the self-interaction energies,  $\mu_A$  and  $\mu_B$ . Therefore, Equations (5)-(7) can be used to optimize the values of the self-interaction energies based on any number of parameterizations. Here, we show that  $\mu_A(\rho_A^L, \rho_B^L)$  and  $\mu_B(\rho_A^L, \rho_B^L)$  can be

parameterized as a two-dimensional function of both  $\rho_A^L$  and  $\rho_B^L$ . To do so, we propose adding a cubic correction term to the pure component one-body terms, so that

$$\mu_A(\rho_A^L, \rho_B^L) = \mu_A^{pure}(\rho_A^L + \rho_B^L) + \mu_A^{cub}(\rho_A^L, \rho_B^L) \quad (\text{A1})$$

with

$$\mu_A^{cub}(\rho_A^L, \rho_B^L) = c_1 \rho_A^L + c_2 \rho_B^L + c_4 (\rho_A^L)^2 + c_4 \rho_A^L \rho_B^L + c_5 (\rho_B^L)^2 + c_6 (\rho_A^L)^3 + c_7 (\rho_A^L)^2 \rho_B^L + c_8 \rho_A^L (\rho_B^L)^2 + c_9 (\rho_B^L)^3 \quad (\text{A2})$$

A similar equation can be written for the  $\mu_B$ 's. If values of  $\mu_{ex}^A$ ,  $\mu_{ex}^B$ ,  $\Delta F_{I,A}$ , and  $\Delta F_{I,B}$  can be measured or estimated across a range of densities ( $\rho_A, \rho_B$ ); then the resulting system of 18 unknowns can then be optimized without the use of additional molecular simulations.

We elected to use a cubic correction instead of the simpler quadratic because optimization over Equations (5)-(7) indicated a slightly superior fit with the cubic correction. However, quadratic correction would likely provide adequate results in many cases.

We examined the performance of this new algorithm in A/B/Z mixtures, and compared the results to the original mixing rule. A subset of the results is shown in Figure 3-A1. It indicates that the two-dimensional density dependence improves transferability substantially. The standard error with respect to  $\mu^{ex}$  is  $0.18 \varepsilon_Z$  for the 13 points studied, compared to  $0.47 \varepsilon_Z$  for the original mixing rule.

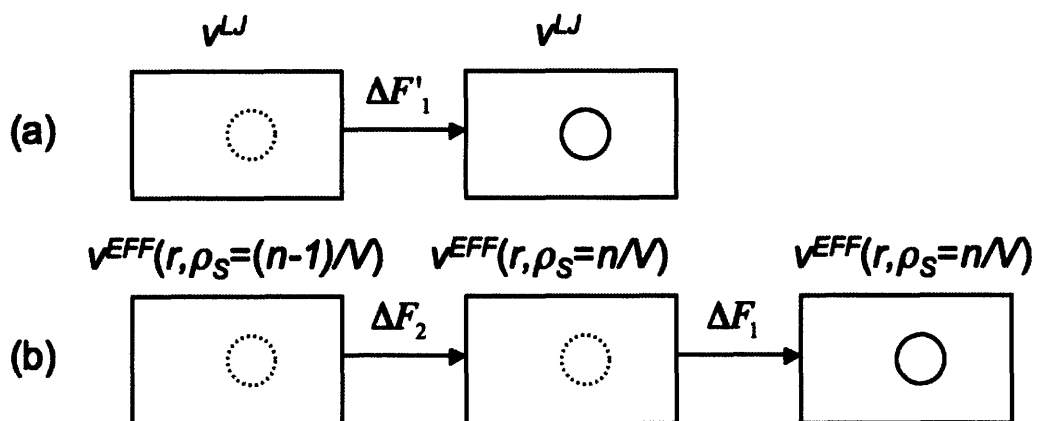


### 3.7 REFERENCES

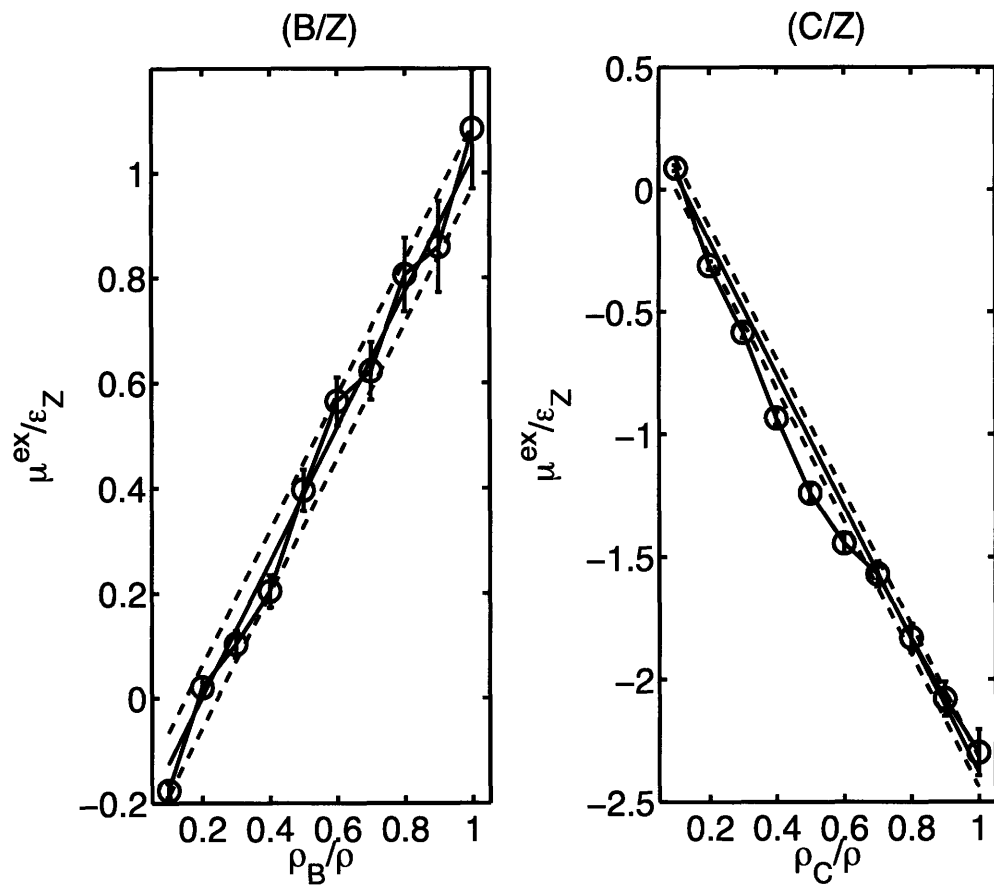
1. S. Izvekov and G. A. Voth, *J. Chem. Phys.* **123**, 134105 (2005).
2. Q. Shi, S. Izvekov, and G. A. Voth, *J. Phys. Chem. B* **110**, 15045 (2006).
3. S. Izvekov and G. A. Voth *J. Chem. Phys.* **125**, 151101 (2006)
4. G. S. Ayton, W. G. Noid, and G. A. Voth, *Mrs Bull.* **32**, 929 (2007).
5. P. Liu, S. Izvekov, and G. A. Voth, *J. Phys. Chem. B* **111**, 11566 (2007).
6. W. G. Noid, J. W. Chu, G. S. Ayton, and G. A. Voth, *J. Phys. Chem. B* **111**, 4116 (2007).
7. J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth, *Biophys. J.* **92**, 4289 (2007).
8. J.D. McCoy and J.G. Curro, *Macromolecules* **31**, 9362 (1998).
9. K. Kremer and F. Muller-Plathe, *Mol. Simulat.* **28**, 729 (2002).
10. N. B. Wilding, *J. Chem. Phys.* **119**, 12163 (2003).
11. M. Bathe and G. C. Rutledge, *J. Comput. Chem.* **24**, 876 (2003).
12. M. Bathe, G. C. Rutledge, A. J. Grodzinsky, and B. Tidor, *Biophys. J.* **88**, 3870 (2005).
13. G. C. Rutledge, *Phys. Rev. E* **63** 021111 (2000).
14. S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein, *J. of Chem. Phys.* **119**, 7043 (2003).
15. W. Shinoda, R. Devane, and M. L. Klein, *Mol. Simulat.* **33**, 27 (2007).
16. R. Faller, H. Schmitz, O. Biermann, and F. Muller-Plathe, *J. Comput. Chem.* **20**, 1009 (1999).

17. H. Meyer, O. Biermann, R. Faller, D. Reith, and F. Muller-Plathe, *J. Chem. Phys.* **113**, 6264 (2000).
18. D. Reith, H. Meyer, and F. Muller-Plathe, *Macromolecules* **34**, 2335 (2001).
19. D. Reith, M. Putz, and F. Muller-Plathe, *J. Comput. Chem.* **24**, 1624 (2003).
20. G. Milano, S. Goudeau, and F. Muller-Plathe, *J. Polym. Sci. Pol. Phys.* **43** (8), 871 (2005).
21. G. Milano and F. Muller-Plathe, *J. Phys. Chem. B* **109**, 18609 (2005).
22. W. Schommers, *Phys. Lett. A* **43**, 157 (1973).
23. W. Schommers, *Phys. Rev. A* **28**, 3599 (1983).
24. D. Levesque, J.J. Weis, and L. Reatto, *Phys. Rev. Lett.* **54**, 454 (1985).
25. G. Toth and A. Baranyai, *Mol. Phys.* **97** (3), 339 (1999).
26. G. Toth and A. Baranyai, *J. Mol. Liq.* **85**, 3 (2000).
27. G. Toth and A. Baranyai, *J. Phys.: Condens. Matter* **17**, S159 (2005).
28. A. P. Lyubartsev and A. Laaksonen, *J. Phys. Chem.* **100**, 16410 (1996).
29. A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E* **55**, 5689 (1997).
30. A. P. Lyubartsev and A. Laaksonen, *Comput. Phys. Comm.* **122**, 57 (1999).
31. A. P. Lyubartsev and A. Laaksonen, *Chem. Phys. Lett.* **325**, 15 (2000).
32. A.E. Ismail, G.C. Rutledge, G. Stephanopoulos, *J. Chem. Phys.* **118**, 4414 (2003).
33. A.E. Ismail, G. Stephanopoulos, G.C. Rutledge, *J. Chem. Phys.* **118**, 4424 (2003).
34. M.E. Johnson, T. Head-Gordon, and A.A. Louis, *J. Chem. Phys.* **126**, 144509 (2007).
35. A. P. Lyubartsev, *Eur. Biophys. J. Biophys.* **35**, 53 (2005).
36. A.A. Louis, *J. Phys.: Condens. Matter* **14**, 9187 (2002).
37. S. Merabia and I. Pagonabarraga, *J. Chem. Phys.* **127**, 054903 (2007).

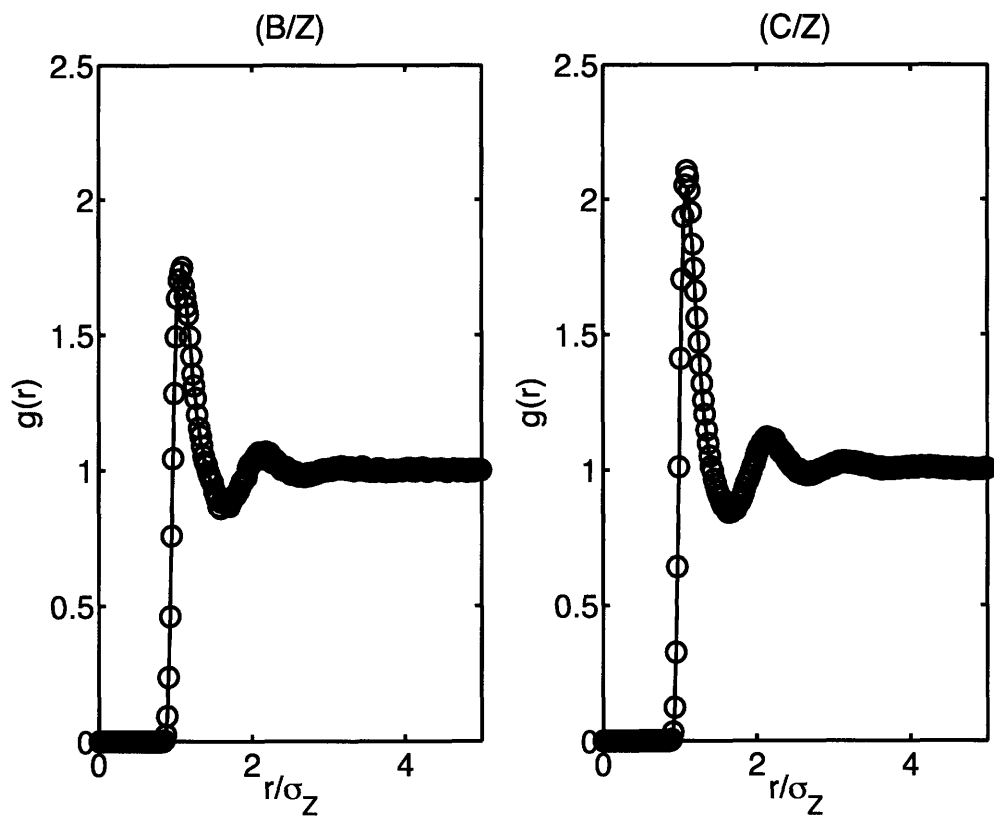
38. E.C. Allen and G.C. Rutledge, *J. Chem. Phys.* **128**, 154115 (2008).
39. R.L. Henderson, *Phys. Lett. A* **49**, 197 (1974).
40. P.G. Bolhuis, A.A. Louis, J.P. Hansen, and E.J. Meijer, *J. Chem. Phys.* **114**, 4296 (2001).
41. B. Hess, H. Bekker, H.J.C. Berendsen, and J.G.E.M. Fraaije, *J. Comp. Chem.* **18**, 1463 (1997).
42. T. Lazaridis and M. Karplus, *Prot. Struct. Funct. Genet.* **35**, 133 (1999).



**Figure 3-1:** Comparison of free energy changes upon particle insertion in the (a) all-particle and (b) density-dependent, implicit solvent cases. The density-dependent potential introduces a secondary free energy change due to the change in energy models associated with a change in global average solute density of the system.



**Figure 3-2:** Comparison of excess chemical potential in all-atom (solid line) and coarse-grained (circles) simulations. Left: B-type particles in solvent Z. Right: C-type particles in solvent Z. The dashed lines demarcate errors of  $0.06\epsilon_Z$ , and are provided as a guide to compare the relative errors between the two coarse-grained potentials.



**Figure 3-3:** Comparison of worst-fit solute RDF in all-atom (solid line) and coarse-grained (circles) simulations. Left: B-type particles in solvent Z. Right: C-type particles in solvent Z.

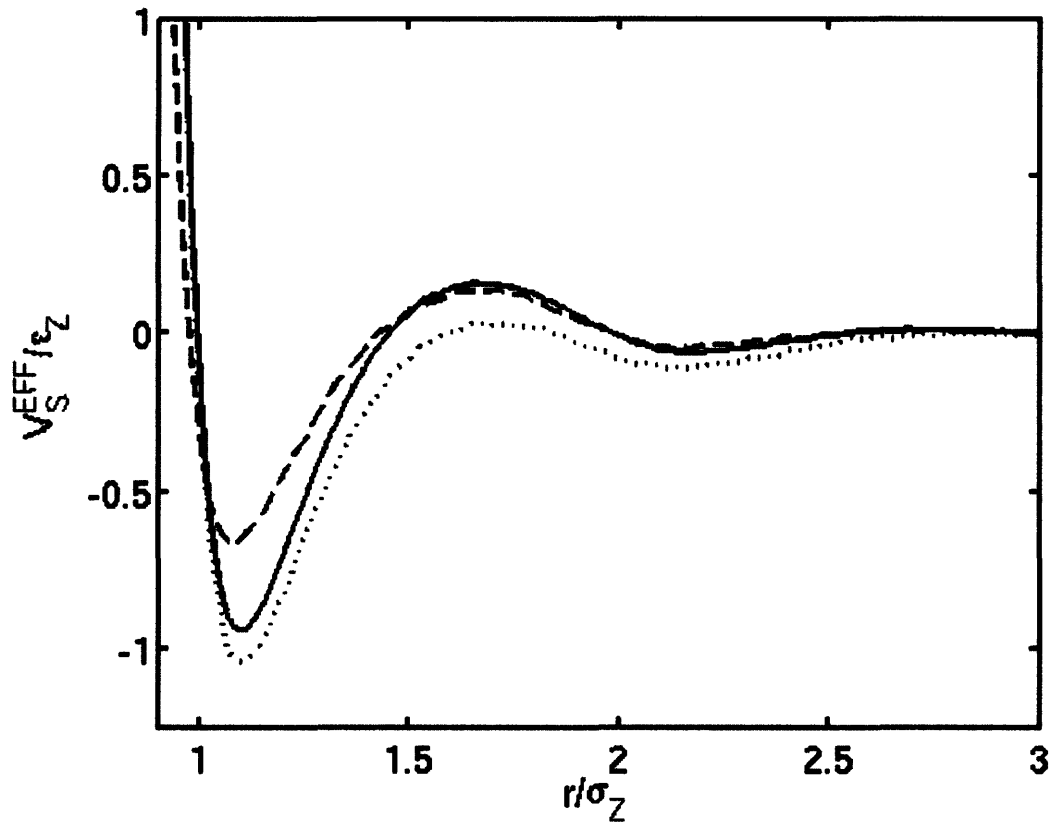
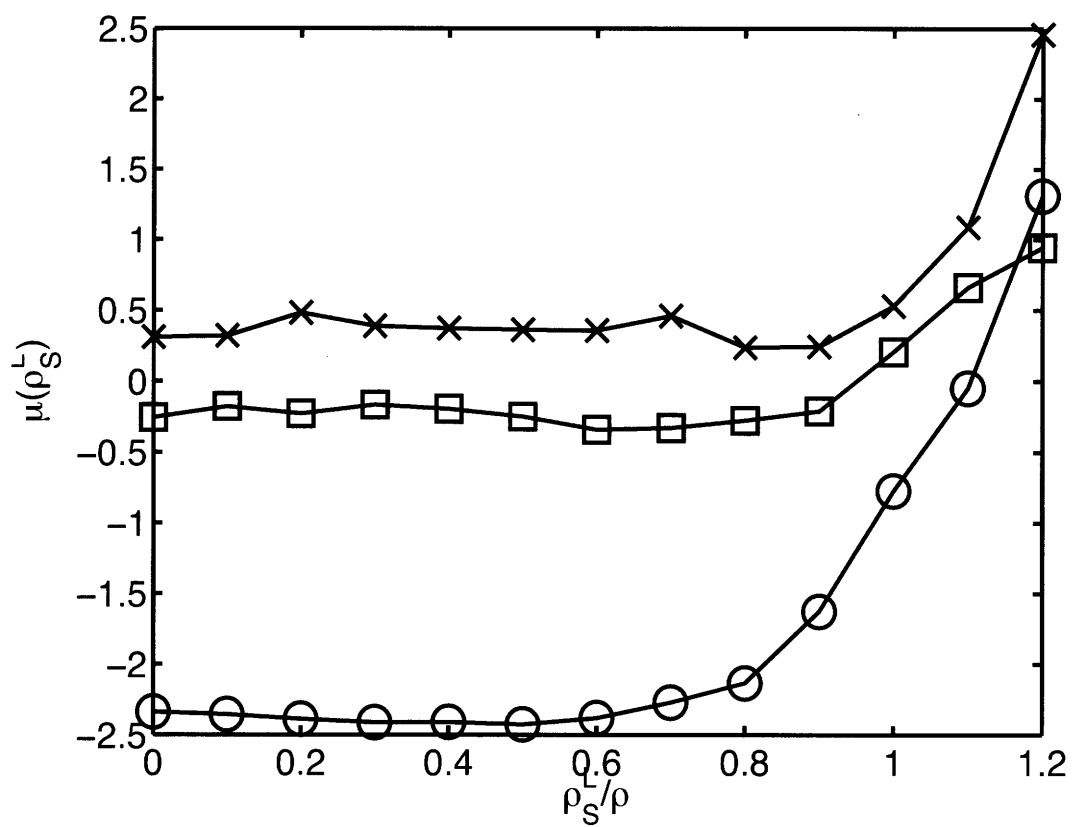
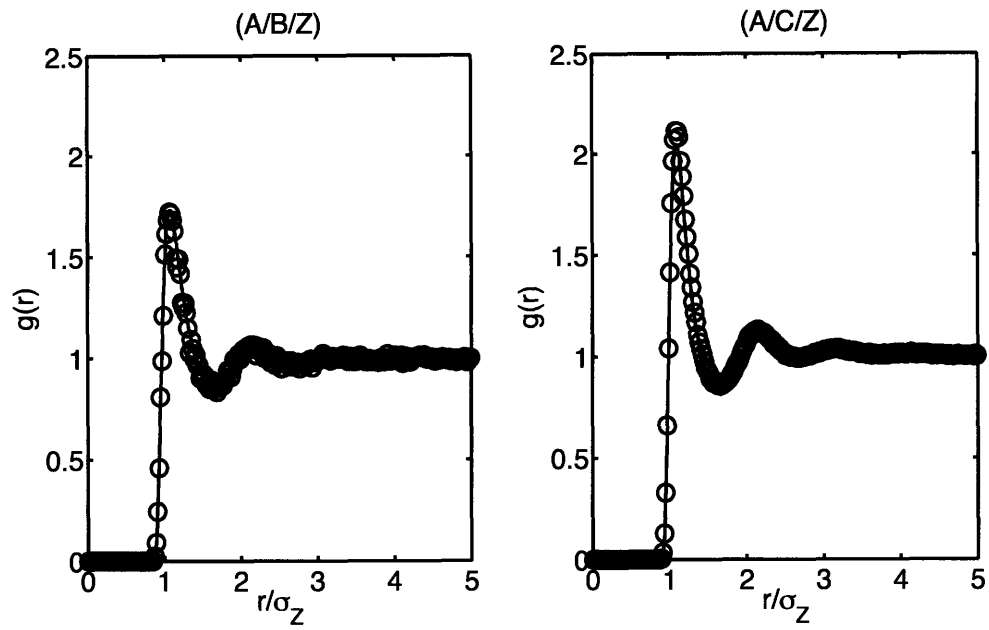


Figure 3-4: Coarse-grained two-body term for local solute density  $\rho_s^L/\rho=0.5$  (—A,--B,···C)

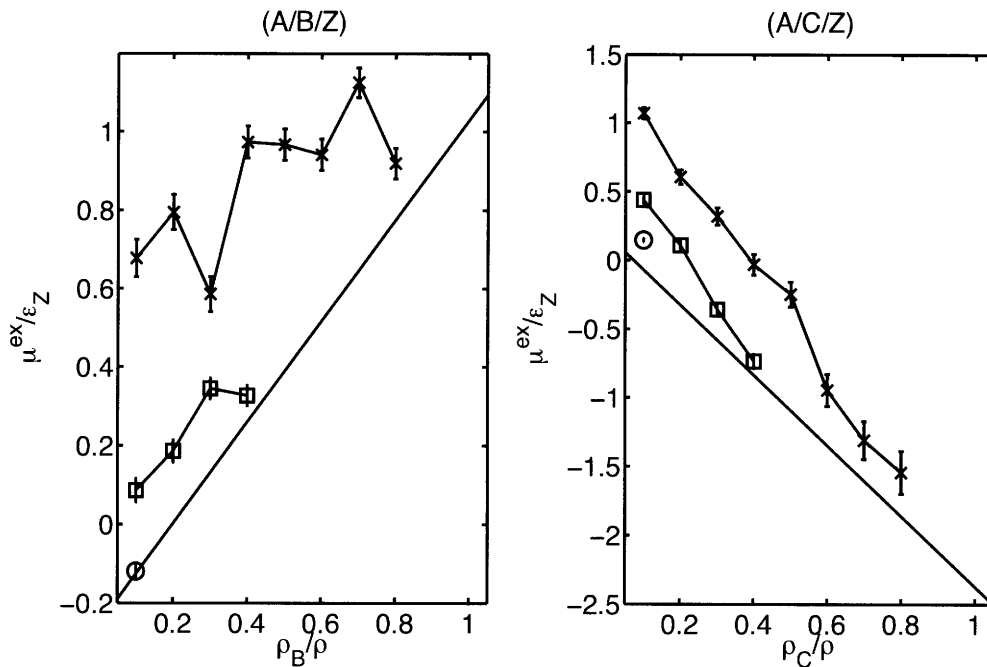


**Figure 3-5:** Coarse-grained one-body term as a function of local solute density (Circles: A-type particles, Squares: B-type particles, Crosses: C-type particles)

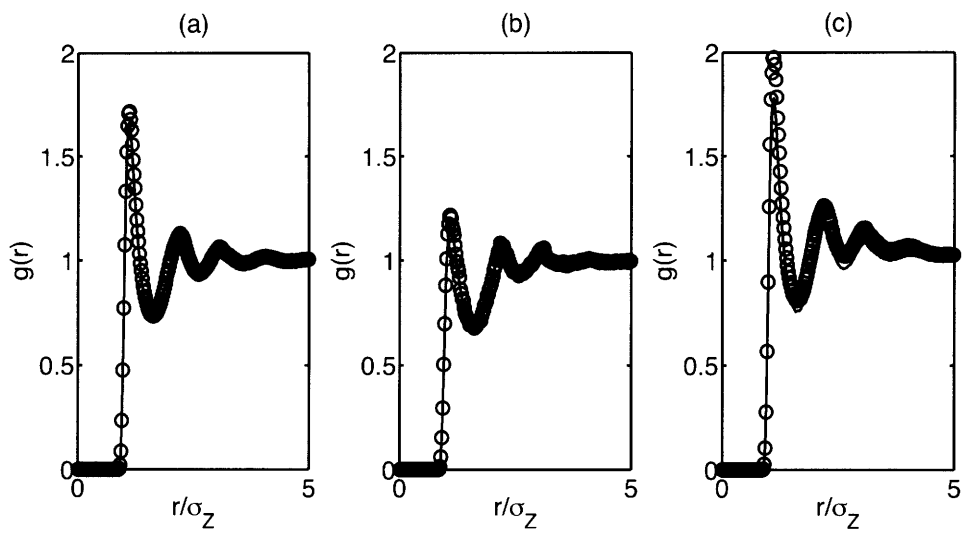




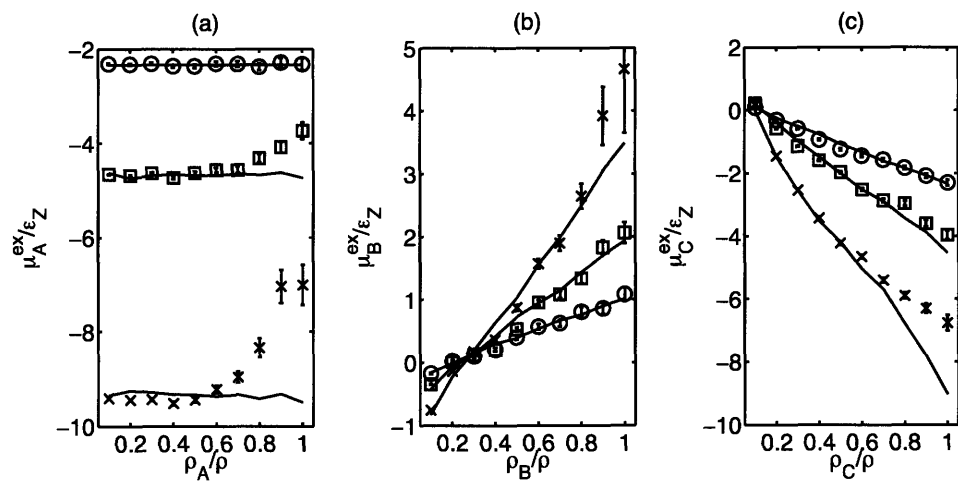
**Figure 3-6:** Worst case RDF reproduction for mixtures. Left Side: A/B/Z mixture, Right Side: A/C/Z mixture (Solid Line – all-atom results; Circles –DDIS results).



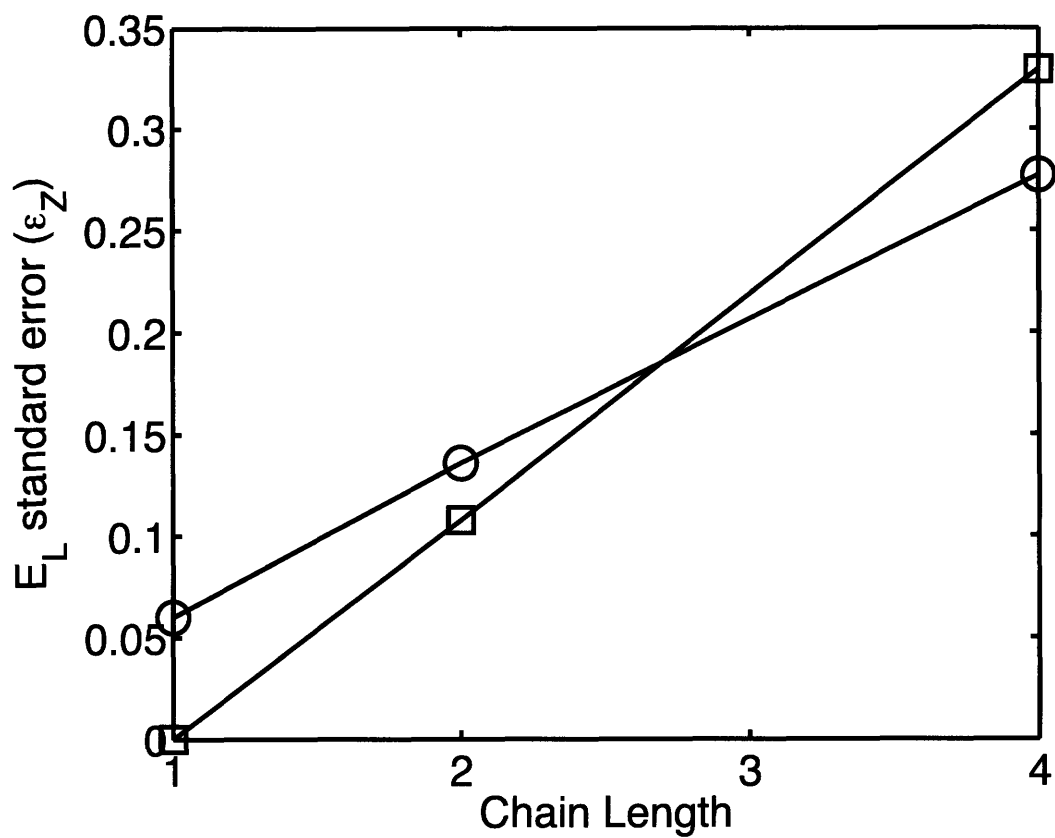
**Figure 3-7:** Comparison of coarse-grained and all-atom excess chemical potential for mixtures, broken down by total solute density,  $\rho_S$ . Left Side:  $\mu_B^{ex}(\rho_A, \rho_B)$  in A/B/Z mixtures. Right Side:  $\mu_C^{ex}(\rho_A, \rho_C)$  in A/C/Z mixtures. (Solid Line – All-atom values; Circles –  $\rho_S/\rho=0.2$ ; Squares –  $\rho_S/\rho=0.5$ ; Crosses –  $\rho_S/\rho=0.9$ ).



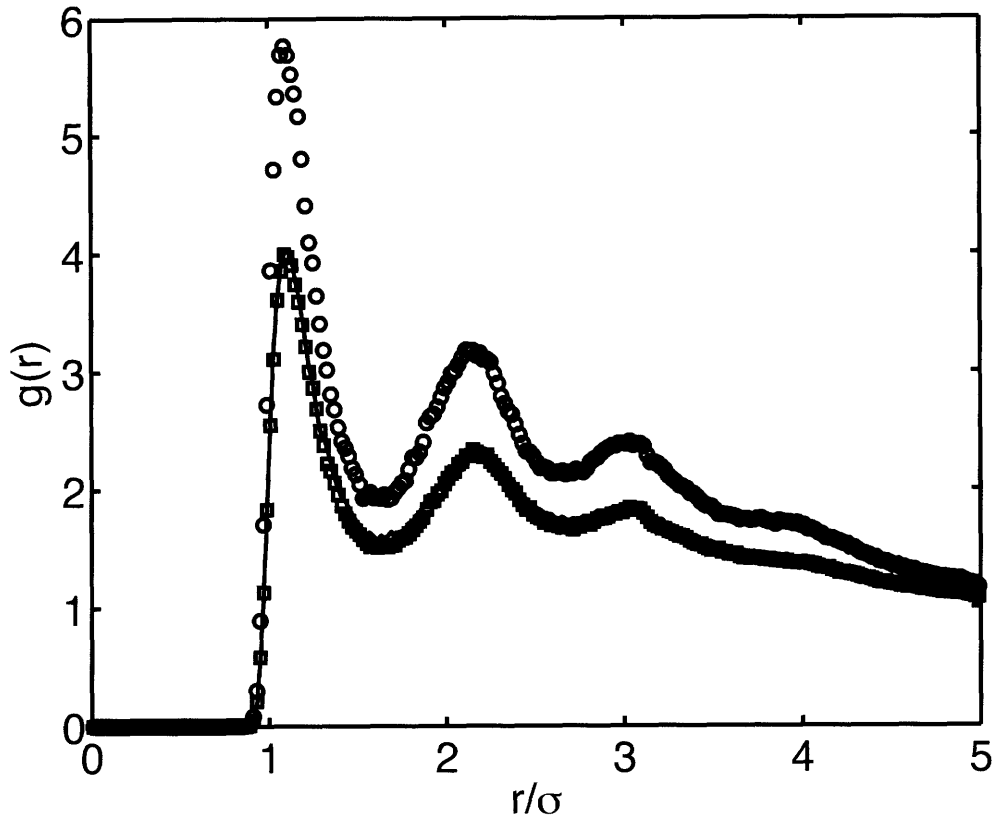
**Figure 3-8:** Worst case RDF reproduction for tetramers of solute particles (a) A-type, (b) B-type, (c) C-type. (Solid Line – All-atom; Circles – DDIS results).



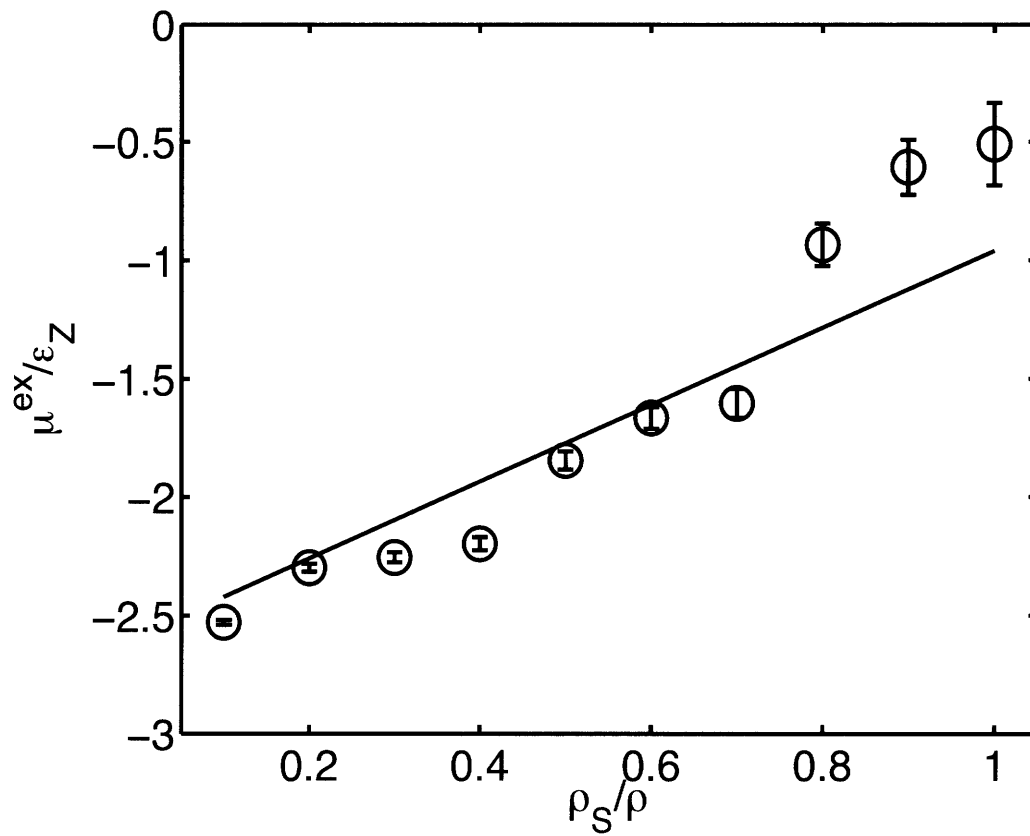
**Figure 3-9:** Comparison of coarse-grained and all-atom excess chemical potential for monomers, dimers, and tetramers as a function of solute density,  $\rho_S$ . (a) A-type, (b) B-type, (c) C-type. (Solid Line – All-atom; Circles – Monomers; Squares – Dimers; Crosses – Tetramers).



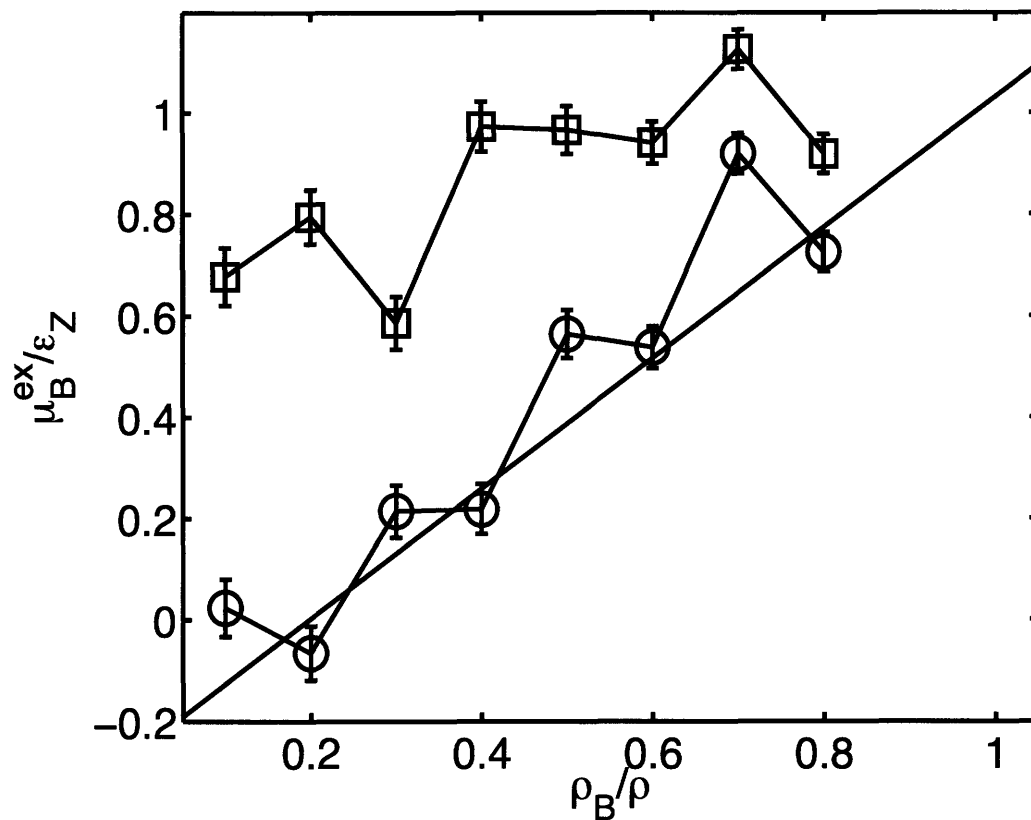
**Figure 3-10:**  $E_L$  standard error (units of  $\epsilon_Z$ ) as a function of chain length for local DDIS and global density-dependent potentials. (Circles – Local potential; Squares – Global potential).



**Figure 3-11:** Comparison of C-C particle RDFs for all-atom, local potential, and global potential for C-type tetramers in solvent Z at  $\rho_C/\rho=0.2$  (Solid Line – all-atom; Circles – global potential; Squares – local potential).



**Figure 3-12:** All-atom (solid line) and coarse-grained (circles)  $\mu^{ex}(\rho_S)$  for  $A_1B_1$  molecule.



**Figure 3-13:** Comparison of measured  $\mu_{ex}^B(\rho_A, \rho_B)$  transferability with and without mixture correction function for  $\rho_s/\rho=0.9$ . Results indicate that parameterizing the self-interaction energy on the density of both particle types improves transferability, and that allowing the pure-component values to vary improves transferability even more (Straight line – All-atom values; Squares – original mixing rule; Circles – modified mixing rule).



<i>Particle Type</i>	$\epsilon_{ii}$	$\sigma_{ii}$	$R/\sigma_{ii}$ ( <i>Like</i> )	$R/\sigma_{ij}$ ( <i>Unlike</i> )	$\Delta G_S(Z \rightarrow S)$	$\langle N_L \rangle / \langle N_L \rangle^{IDEAL}$
A	1.0	1.0	5.0	5.0	0.0	1.00
B	0.5	1.0	5.0	5.0	$1.29 \pm 0.05 \epsilon_Z$	1.00
C	1.0	1.0	5.0	1.84	$-2.71 \pm 0.08 \epsilon_Z$	1.01

**Table 3-1:** Parameters for solute types used in this chapter. B-type solute is “solvent-philic”, as indicated by the positive free energy of transfer. C-type is “solvent-phobic”, as indicated by a negative free energy of transfer.

$E_L/\epsilon_Z$	$N=1$	$N=2$	$N=4$
A	0.015	0.04	0.06
B	0.005	0.010	0.011
C	0.06	0.14	0.28

**Table 3-2:** Values of  $E_L$  for systems of monomers ( $N=1$ ), dimers ( $N=2$ ), and tetramers ( $N=4$ ) for chains of A, B, and C type solute particles.

$\mu^{ex}/\epsilon_z$	$N=1$	$N=2$	$N=4$
A	0.04	0.19	0.27
B	0.04	0.06	0.11
C	0.09	0.13	0.23

**Table 3-3:** Values of  $\mu^{ex}$  for systems of monomers ( $N=1$ ), dimers ( $N=2$ ), and tetramers ( $N=4$ ) for chains of A, B, and C type solute particles.

## **4 Simulation of Surfactant Systems using Coarse-Grained, Density Dependent Implicit Solvent Models**

### **4.1 INTRODUCTION**

A surfactant molecule is composed of a head group that is compatible with the surrounding solvent, for example water, and a tail group that is not. Above a critical concentration in solution, the surfactant molecules spontaneously aggregate into structures known as micelles, in which the head groups surround and shield the tail groups from the solvent.

Theoretical approaches<sup>1-2</sup> treat micellization as arising from a set of competing free energy effects. The transfer of hydrophobic tails from melt to solution promotes micelle formation, but is offset by a surface energy penalty and the loss of surfactant translational entropy. These competing effects create a well-defined free energy minimum as a function of aggregation number, so that the resulting micelles are of nearly uniform size. The concentration at which surfactant molecules spontaneously form micelles is referred to as the critical micelle concentration, or CMC, and in theoretical treatments is directly related to the aforementioned free energy considerations. Both the CMC and the average aggregation number are important properties that characterize the self-assembling nature of surfactants and for which we desire efficient, predictive capabilities.

The literature on particle-based surfactant simulations contains a diverse set of particle representation approaches<sup>3-22</sup>, which can be grouped according to two key design decisions: the level of detail used to describe the surfactant molecule itself, and explicit or implicit treatment of the solvent. Explicit solvent simulations<sup>3-19</sup> are computationally burdensome because of the low value of experimentally measured CMC's ( $10^{-6}$ - $10^{-2}$  M). At these concentrations, a significant majority of the simulation cell is composed of the relatively uninteresting solvent. As a result, the direct measurement of the CMC by explicit solvent simulations is impractical, even for very coarse-grained surfactant and solvent representations. Instead, the CMC and aggregation number distribution are generally determined indirectly, through the use of free energy simulations<sup>10,14,15,18</sup>.

By reducing the number of particles represented, and thus also the computational burden, implicit solvent simulations of surfactant systems allow investigation of much longer time and length scales. By this approach, Lazaridis *et al.*<sup>21</sup> studied the formation of dodecylphosphocholine (DPC) micelles using the Effective Energy Function<sup>23</sup> (EEF1) implicit solvent model. The simulated CMC was close to the reported experimental value. Von Gottberg *et al.*<sup>22</sup> studied  $A_2B_2$  surfactants in implicit solvent using stochastic dynamics (SD), in which each surfactant bead was intended to represent a Kuhn segment (approximately ten methylene groups) of a polymer chain. They successfully simulated a fully equilibrated micelle size distribution at multiple surfactant concentrations, from which they were able to determine the system CMC.

In this paper, we employ our previously developed coarse-graining algorithm to generate density-dependent, implicit solvent (DDIS) potentials<sup>24,25</sup> for a model surfactant and test their ability to describe surfactant aggregation. The DDIS potentials are specifically derived to reproduce the chemical potentials and radial distribution functions (RDFs) of an underlying all-atom simulation. The advantage of creating an implicit solvent energy model via coarse-graining is that the results have a clear correspondence to an underlying all-atom energy model that is presumed to be a more accurate representation of the system.

To investigate the applicability of DDIS potentials to surfactant systems, we create and test such potentials for a model surfactant in a Lennard-Jones solvent. The model is a derivation of Smit's<sup>11-13</sup> well-known surfactant model, for which Pool and Bolhuis recently estimated the CMC and aggregation number distribution by means of free energy calculations<sup>18</sup>. As in previous chapters, we derive the DDIS potentials from simulations of monomeric solutes in explicit solvent. Our potentials are not derived from mixtures of solute types nor chain simulations. We made this decision because it represents the simplest set of simulations from which to derive DDIS potentials, and allows us to establish a "baseline" transferability for the DDIS potentials against which alternative derivation simulations might be tested. We use the results of these previous works to test the accuracy and transferability of our DDIS potentials to this surfactant model.

## 4.2 THEORY

### 4.2.1 Density Notation Conventions

In this work,  $\rho_S$  refers to the *total* density of solute particles, where the subscript “S” stands for solute, and may include contributions from different solute types.  $\rho_I$  is the density of type  $I$  solute particles only.  $\{\rho\}$  refers to the set of  $\rho_I$  that includes every particle type in the system. Finally,  $\rho$  without subscript refers to the state point density, considering all particles (solute and solvent) explicitly. A superscript  $L$  indicates that the density is the *local* density of solute particles; otherwise, the density refers to the global average density.

Particles of different size contribute to the density in proportion to their volume, which is consistent with the Effective Energy Function<sup>23</sup> implicit solvent model. This convention is chosen because each solute particle reduces the local solvent density roughly in proportion to its volume. Thus, the solute density in a specified volume  $V$  is

$$\rho_S = \frac{1}{V} \sum_{i=1}^N \left( \frac{\sigma_i}{\sigma_W} \right)^3 \quad (1)$$

where  $N$  is the number of solute particles in the specified volume,  $\sigma_i$  is the Lennard-Jones radius for particle  $i$ , and  $\sigma_W$  is the Lennard-Jones radius of the solvent.

## 4.2.2 DDIS Potential Review

In Chapter 2, we proposed a density-dependent, implicit solvent (DDIS) potential with the form

$$E_i = \mu(\rho_{S,i}^L) + \frac{1}{2} \sum_{j \neq i} V_{IJ}^{EFF}(r_{ij}, \rho_{S,i}^L) \quad (2)$$

where  $E_i$  is the effective energy of particle  $i$ ,  $V_{IJ}^{EFF}$  is a pairwise potential between  $I$ - and  $J$ -type solute particles that is an explicit function of solute particle density in the vicinity of particle  $i$ ,  $\rho_{S,i}^L$ , and  $\mu$  is a “self-interaction” term that is also a function of solute particle density. The first term serves to capture the effects of solute-solvent and solvent-solvent interactions in a mean field approximation, while the second term represents solute-solute interactions that are mediated by solvent. The dependence of both terms on local density allows the potential to be cognizant of the greater or lesser influence of solvent in the different regions of a system that is inhomogeneous on the solute length scale, such as that which arises in a system of surfactants that self-assemble into micelles. We also proposed a method to derive such potentials from simulations of solute type  $I$  in solvent. This fitting was performed such that the solute-solute RDF and solute excess chemical potential,  $\mu_i^{ex}$ , are reproduced across all solution compositions. For details of this fitting procedure, the reader is referred to Chapter 2. The information is also available in Ref. 24.



Mixtures of multiple particle types are treated by a simple mixing rule in which like particle interactions are unmodified, but the local density used is the total local solute density. DDIS interactions between dissimilar particles are derived from the interactions between pairs of identical particles through the equation

$$V_{IJ}^{EFF}\left(\frac{r}{\sigma_{IJ}}; \rho_S^L\right) = V_{IJ}\left(\frac{r}{\sigma_{IJ}}\right) + \frac{\epsilon_{IJ}}{2} \left( \frac{\Delta V_{II}^{EFF}\left(r/\sigma_I; \rho_S^L\right)}{\epsilon_I} + \frac{\Delta V_{JJ}^{EFF}\left(r/\sigma_J; \rho_S^L\right)}{\epsilon_J} \right) \quad (3)$$

where  $\epsilon_{IJ}$  and  $\sigma_{IJ}$  are the Lennard-Jones parameters for  $I$ - $J$  interactions,  $V_{IJ}(r)$  is the  $I$ - $J$  all-atom potential, and

$$\Delta V_{II}^{EFF}\left(\frac{r}{\sigma_I}; \rho_S^L\right) = V_{II}\left(\frac{r}{\sigma_I}\right) - V_{II}^{EFF}\left(\frac{r}{\sigma_I}; \rho_S^L\right) \quad (4)$$

Chapter 3 discusses the mixing rule and its applicability further, and is also available in Ref. 25.

### 4.2.3 Simulation Metrics

The measure of error in the RDF is given by the solute-solute energy, defined as:

$$E_{L,IJ} = \sqrt{\rho_I \rho_J} \int_0^{r_C} V_{IJ}(r) g_{IJ}(r, \{\rho\}) 4\pi r^2 dr \quad (5)$$

where  $\rho_I$  is the density of solute type I,  $V_{IJ}$  is the exact interaction potential between I- and J-type particles, and  $g_{IJ}$  is the I-J RDF. The difference between  $E_L$  for a coarse-grained system and that for the all-atom simulation provides a measure of the error in  $g(r, \{\rho\})$ , relative to the exact  $g(r, \{\rho\}, \rho)$  for the explicit, all-atom system, in units of energy. The error in the excess chemical potential of particle type I,  $\mu_I^{ex}$ , is the difference (in units of  $\epsilon_W$ , the Lennard-Jones parameter of the solvent) between the target (all-atom) and measured (coarse-grain) values. In this text, both error measurements are presented in terms of the standard error over all simulations. Additionally, we supply figures where appropriate comparing the all-atom and coarse-grained chemical potentials and RDFs, as these offer an intuitive sense of the closeness of fit.

A solute's relative affinity for the solvent can be measured by the free energy of transfer,  $\Delta G_I(W \rightarrow I)$ , defined as the free energy change associated with transferring a single solute particle from a bath of solvent particles W to a bath of solute particles I. The free energy of transfer can be calculated from the excess chemical potential of solute particles:

$$\Delta G_I(W \rightarrow I) = \mu_I^{ex} \left( \frac{\rho_I}{\rho} = 1 \right) - \mu_I^{ex} \left( \frac{\rho_I}{\rho} = 0 \right) \quad (6)$$

We characterize the local solute environment by calculating the local number of solute particles,  $\langle N_L \rangle$ , defined as:

$$\langle N_L \rangle = \rho_I \int_0^{R_C} g(r; \rho_I) 4\pi r^2 dr \quad (7)$$

We define the solute enhancement ratio as the number of solute particles within a sphere of radius  $R_c$  divided by the average number of particles in such a volume. A value less than one indicates a local depletion in solute density, while a value greater than one indicates a local enhancement. A value near 1.0 indicates that the local solute environment is very similar to the global average environment.

#### 4.2.4 Surfactant Theory

Statistical thermodynamic theories of micelle formation attempt to estimate the critical micelle concentration based on the free energy of micelle formation. The law of mass action<sup>1,2</sup> says that the mole fraction of micelles of size  $n$ ,  $X_n$ , is given by:

$$X_n = (X_1)^n \exp[-\beta n g_{mic}(n)] \quad (8)$$

where  $X_1$  is the mole fraction of free surfactant (i.e. surfactants not bound in a micellar state) and  $g_{mic}(n)$  is the free energy of transferring a surfactant molecule from solution to a micelle of size  $n$ . Equation (8) assumes that the translational entropy of surfactants in free solution can be equated to that of an ideal gas, and the equation is generally valid for spherical micelles at low concentrations<sup>22</sup>. The critical micelle concentration,  $X_{CMC}$ , represents the free surfactant concentration at which significant micellar aggregation occurs, and can be expressed in free energy terms as:

$$X_{CMC} = \exp[\beta g_{mic}(n^*)] \quad (9)$$

where  $n^*$  is the most probable aggregation number. In contrast to statistical thermodynamic theories, molecular simulation can be used to measure the CMC directly, and the mole fraction of free surfactant can then be used to estimate the free energy of micelle formation.

Israelachvili *et al.*<sup>26</sup> have suggested that the shape adopted by a micelle depends on the value of the dimensionless packing parameter,  $v/a_0l_c$ , where  $v$  is the volume of the solvent-phobic tail,  $a_0$  is the head-group area, and  $l_c$  is the length of a fully extended surfactant molecule. Micelles in which the value of the packing parameter is less than 1/3 generally adopt a spherical configuration, while those in which the packing parameter is greater than 1/3 but less than 1/2 adopt an extended, cylindrical shape in which the micelle can grow indefinitely along one dimension. In general, Israelachvili's theory predicts that, for a given head group, increasing the length of a surfactant molecule's tail group tends to promote the transition from spherical to cylindrical micelle growth because the tail volume  $v$  grows more quickly than the length  $l_c$ . In addition, increasing the volume of the surfactant tail group (or equivalently, decreasing the packing density of the tail group) while keeping chain length constant also tends to promote cylindrical micelle growth.

## 4.3 SIMULATION PROTOCOL

### 4.3.1 All-Atom Model

The underlying all-atom model described by Smit<sup>11-13</sup> is composed of three basic particle types: head, tail, and solvent particles, which are hereafter referred to as H, T, and W respectively. These particles interact via the truncated and shifted Lennard-Jones potential, where the potential between particles  $i$  and  $j$  is described by:

$$V_{ij}^{TS}(r_{ij}; \epsilon_{ij}, \sigma_{ij}, R_{c,ij}) = \begin{cases} V_{ij}^{LJ}(r_{ij}; \epsilon_{ij}, \sigma_{ij}) - V_{ij}^{LJ}(R_{c,ij}; \epsilon_{ij}, \sigma_{ij}) & r_{ij} \leq R_{c,ij} \\ 0 & r_{ij} > R_{c,ij} \end{cases} \quad (10)$$

with

$$V_{ij}^{LJ}(r_{ij}; \epsilon_{ij}, \sigma_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (11)$$

where  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the Lennard-Jones parameters for  $ij$  interactions.  $\epsilon$  and  $\sigma$  values of dissimilar particles are governed by Lorentz-Berthelot mixing rules. Lennard-Jones parameters for all three particle types are given in Table 1. Equations (10) and (11) allow for the possibility of different cut-off radii ( $R_{c,ij}$ ) for interactions between different particle types.  $(R_{c,ij}/\sigma_{ij})=2.5$  for W-W, H-W, and T-T interactions, and  $(R_{c,ij}/\sigma_{ij})=2^{1/6}$  for T-W, H-H, and H-T interactions. These values were chosen to give the H particles a “solvent-philic” behavior, and the T particles a “solvent-phobic” behavior. As indicated in Table 1,  $\Delta G_H(W \rightarrow H)$  is positive for H-type particles, indicating that H-type particles

prefer a solvent environment relative to one composed of other H-type particles.

Conversely,  $\Delta G_T(W \rightarrow T)$  is negative for T-type particles.

Table 1 shows the solute enhancement ratio for all three types of solute particles at  $\rho_H/\rho=0.5$ . These results indicate that H-type particles experience a local depletion of other H-type particles, while T-type particles experience a substantial local enhancement of other T-type particles. The local depletion of H-type particles arises because the radius of H-type particles ( $2\sigma_W$ ) creates a large excluded volume. As a result, the total density within the cutoff radius ( $5\sigma_W$ ) is less than the average solute density.

Surfactant molecules are composed of a single H-type particle connected to multiple T-type particles in a linear fashion, denoted  $H_1T_M$ , where M is the number of T-type particles that form the tail. Bonded interactions between I- and J-type particles are governed by a harmonic potential:

$$V_{ij}^{bond}(r_{ij}) = \frac{1}{2} k_{bond} (r_{ij} - r_{eq,IJ})^2 \quad (12)$$

where  $k_{bond}=5000 \epsilon/\sigma_W^2$  is the harmonic spring constant, and  $r_{eq,IJ}=\sigma_{IJ}$  is the equilibrium bond length. The bonded interaction applies only to nearest neighbor particles in the surfactant chain.

### 4.3.2 Simulation Details

Simulation protocols for all-atom and coarse-grained simulations of monomeric solute are given in Chapter 2. Monomeric solute simulations were carried out in the *NVT* ensemble at  $T^*=k_B T/\epsilon_W=1.0$  and  $\rho^*=\rho\sigma_W^3=0.60$ , where  $\epsilon_W$  and  $\sigma_W$  are the Lennard-Jones parameters of the solvent W.

All-atom simulations of surfactant molecules were carried out in the same manner as the monomeric all-atom simulations, and implicit solvent simulations of surfactants were carried out in the same manner as the monomeric DDIS simulations. Nearest bonded neighbor particles were included in the calculation of local density. In addition to the atom-level translation moves used for monomeric simulations, simulations of surfactants included rigid body chain translation moves, rigid body chain rotation moves, and rotation about individual bonds. The proportion of these moves was 20% atom translation:60% chain translation:10% chain rotation:10% bond rotation. Simulations were equilibrated for  $10^4$  cycles, followed by sampling for  $4 \times 10^5$  cycles. Free energy sampling was performed every two MC cycles, which was sufficient to generate statistically independent samples, as determined by the autocorrelation function of the measured free energy. RDF sampling was performed every 100 MC cycles.

Micelle aggregation number was measured by a clustering algorithm. Surfactant chains were designated as clustered if the distance between their centers of mass was less than  $1.5\sigma_T$ , where  $\sigma_T$  is the Lennard-Jones radius for tail particles.

## 4.4 RESULTS AND DISCUSSION

### 4.4.1 DDIS Potentials

DDIS potentials were generated for H- and T-type particles in implicit solvent W as described previously<sup>24,25</sup>. Figure 4-1 demonstrates the ability of the H-H and T-T DDIS potentials to reproduce the excess chemical potential of the equivalent all-atom simulations across a range of solute densities. The standard error in  $\mu^{ex}$  for H-type particles is  $0.06\epsilon$ , versus  $0.33\epsilon$  for T-type particles. The larger relative error in the solvent-phobic T-type fitting is consistent with our previous work<sup>25</sup>.

Figure 4-2 shows the worst-case reproductions of the all-atom RDFs by the DDIS potentials for the H-type and T-type potentials. As with the excess chemical potential, the level of reproduction obtained for the H-type potential is much better than for the T-type potential. The standard errors in  $E_L$  were  $E_{L,H}=0.01\epsilon$  and  $E_{L,T}=2.3\epsilon$ . The T-type DDIS potential appears to produce a level of solute aggregation that is too low relative to the all-atom potential. While the solute enhancement ratio of the all-atom and DDIS potentials is nearly identical in the case of H-type particles (0.90 and 0.92, respectively), they differ tremendously in the case of T-type particles (2.5 and 1.3, respectively). The RDFs generated by the DDIS potential contain first and second peaks at the same radii as the corresponding RDFs generated by the all-atom potential. However, in the case of T-type particles, these peaks are strongly attenuated.

The failure of the T-type DDIS potential to reproduce the all-atom RDF may be related to the small system size used in the fitting process. For the smallest system size (100 T-type



particles in 900 implicit W) a droplet of T particles forms with a radius smaller than the density cutoff of  $5\sigma_W$ . Subsequent tests on a bilayer of T particles revealed more accurate reproduction of the RDF for the same fitted potential. Furthermore, the quality of RDF reproduction improves as the T particle density increases. Figure 4-3 compares the RDFs of T-type particle in implicit solvent particles for  $\rho_T/\rho=0.1, 0.5$  and  $1.0$ . As the number of T-type particles in the simulation increases, so too does the degree of agreement between the all-atom and DDIS RDFs. Figure 4-3 also demonstrates that the DDIS potential predicts correctly that the magnitude of the first peak increases as solute density decreases, as is observed in the all-atom results. However, this increase is attenuated in the DDIS potential.

Figure 4-4 compares the two-body portion of the coarse-grained potentials for the H-type and T-type potentials at a local density of  $\rho_i^l/\rho=0.2$ . At short distances, the “solvent-philic” H-type particles experience a repulsive potential where the attractive well would typically reside. The shape of the H-type potential is reminiscent of the class of potentials referred to as “Hard Core/Soft Shoulder” (HCSS) interactions<sup>27-28</sup>. However, the system behavior generated by the repulsive soft shoulder in this potential differs from that of previous HCSS studies. There, the potential was used at low temperatures, and the repulsive plateau promoted local particle aggregation. In this work, the soft shoulder models the first neighbor shell of implicit solvent, and actually encourages local particle depletion.

In contrast to the H-type particles, the T-type particles experience an attractive well deeper than the all-atom Lennard-Jones potential, which is responsible for increasing solute aggregation as the density of T-type particles decreases. These trends are consistent with Chapter 3.

Figure 4-5 shows the one-body portion of the DDIS potential for H- and T-type particles. As solute density goes to zero, the value of the one-body term approaches  $\mu_l^{ex}(\rho_l/\rho=0)$ . The profile is relatively flat for low solute densities, changing rapidly as the solute fraction approaches 1. The one-body term increases for H-type particles as density increases, indicating that it is energetically unfavorable to achieve high local densities. In contrast, the one-body term for T-type particles decreases, further promoting aggregation.

#### 4.4.2 Surfactant Results

We performed multiple simulations of 216 surfactant molecules across a range of densities. These simulations demonstrate the true benefit of the implicit solvent approach, since the equivalent all-atom simulation is not computationally feasible. From these simulations, we attempt to determine the system CMC and average micelle aggregation number. Pool and Bohuis<sup>18</sup> estimated the CMC's for H<sub>1</sub>T<sub>4</sub> and H<sub>1</sub>T<sub>5</sub> surfactants using semi-grand canonical simulations and found values of  $5 \times 10^{-6} \sigma_W^{-3}$ , and  $6 \times 10^{-7} \sigma_W^{-3}$ , respectively. These results are given in Table 2. Due to the presence of explicit solvent, Pool and Bolhuis were not able to directly observe these values, but required a thermodynamic framework to evaluate the results. Given the differences in simulation type, direct comparison of their results and ours is not possible, but the results provide reasonable guidelines.

Figure 4-6 shows the number density of free surfactant molecules,  $\rho_1$ , as a function of the total number density of surfactants,  $\rho_S$ . The solid black line is the 45° line corresponding to no aggregation. As Figure 4-6 demonstrates, the free surfactant density initially increases along the 45° line, but deviates from it as total surfactant density is increased. This deviation indicates the formation of micellar aggregates. The presence of micelles can be visually confirmed, as is shown in Figure 4-7. The decrease in  $\rho_1$  above a certain  $\rho_S$  level is consistent with the findings of von Gottberg *et al.*<sup>22</sup>, who demonstrate that such a decrease is due to excluded volume effects.

To calculate critical micelle concentration, we follow the convention of von Gottberg *et al.*<sup>22</sup>, who define the CMC as the maximum free surfactant concentration across all simulations. The values obtained for H<sub>1</sub>T<sub>4</sub> and H<sub>1</sub>T<sub>5</sub> surfactants by this method are  $2 \times 10^{-4} \sigma_W^{-3}$ , and  $1.9 \times 10^{-5} \sigma_W^{-3}$ , as shown in Table 2. These CMC values are 20-30 times greater than the values reported by Pool and Bolhuis. We attribute the discrepancy to the approximate nature of the fitting conditions used to generate the DDIS potentials, which did not explicitly include either chains or mixtures. Nevertheless, Table 2 shows that, while the absolute magnitudes of the CMCs obtained using the DDIS potentials are too high, the ratio of CMCs for H<sub>1</sub>T<sub>4</sub> and H<sub>1</sub>T<sub>5</sub> is correct. Thus, even under these conditions, the DDIS potentials have utility for comparative CMC prediction.

Figure 4-8 shows the H<sub>1</sub>T<sub>4</sub> micelle size distribution for a total surfactant number density of  $3 \times 10^{-4} \sigma_W^{-3}$ . The profile shows that a large fraction of the surfactant exists as free

surfactant, with the secondary maximum indicating the presence of stable micellar structures. We calculate the average aggregation number as the number weighted average of all aggregates of size two or larger. For the  $H_1T_4$  surfactant system, the average aggregation number is 26, which is larger than the size of  $\sim 20$  reported by Pool and Bolhuis. This finding implies a looser packing of aggregates using the DDIS potential than with the corresponding all-atom potential. We will examine this point further in the next section.

For  $H_1T_5$  surfactants, the DDIS potential produces cylindrical, or “worm-like”, micelles. The polydisperse distribution of cylindrical micelles makes estimation of the average aggregation number difficult. By contrast, Pool and Bolhuis report that the all-atom potential creates spherical micelles. This finding is also consistent with the formation of less densely packed micellar aggregates. The micelle theory of Ischreavelli<sup>26</sup> states that increasing the volume of a surfactant chain for a constant length and headgroup area will promote formation of cylindrical aggregates. Thus, the transition to cylindrical growth of the  $H_1T_5$  surfactants using the DDIS potential is consistent with the finding that the potential produces higher volume (lower density) micellar structures.

#### **4.4.3 Analysis of Transferability**

Having demonstrated that the DDIS potentials capture micelle formation at least qualitatively, we now explore the reasons that the DDIS potential failed to reproduce the underlying all-atom potential quantitatively. To do so, we consider a few special cases that are intermediate between the systems used to derive the DDIS potentials and the

micellar systems to which they are ultimately applied here. We compare the behavior of the systems using both the all-atom and DDIS potentials.

First, we examine the behavior of monomeric H/T/W mixtures using the DDIS potentials. Figure 4-9 shows the reproductions of all-atom RDFs by the DDIS potentials for mixtures of H and T-type particles at total solute density  $\rho_S/\rho=0.3$ . Visually, the agreement is quite good. This is true even for the T-T RDF, for which good matching was not achieved at the state point of fitting. The error in  $E_L$  for the H-H, H-T, and T-T interactions were  $0.01\epsilon$ ,  $0.02\epsilon$ , and  $0.09\epsilon$  respectively. The middle panel, displaying the H-T RDF, indicates that the mixing rule used here is effective despite its simplicity and lack of additional fitting.

The values of  $\mu^{ex}$  obtained from the DDIS potential do not match the corresponding all-atom values, however. The  $\mu^{ex}$  values from the all-atom simulation are  $-4.00\pm 0.02\epsilon$  and  $0.77\pm 0.02\epsilon$  for H-type and T-type particles, respectively. The DDIS values, by contrast, are  $-2.38\pm 0.02\epsilon$  and  $2.18\pm 0.02\epsilon$ . In Chapter 3, we suggested that the quantity  $\mu_H^{ex}+4\mu_T^{ex}$  should be the same in both simulations, since the ratio of H-type to T-type particles is 1:4. However, the simulations differ in this quantity by over  $7\epsilon$ . The reason for this discrepancy is attributed to an unusual feature of the Smit potential: H-T interactions in the Smit model have a different cutoff radius from both the H-H and T-T interactions. The analysis of transfereability of DDIS potentials presented previously in Chapter 3 assumed implicitly that the H-H and T-T RDFs would be largely unaffected when moving from pure component to mixture simulations. It further assumed that the H-T

RDF would be intermediate in shape to the H-H and T-T RDFs. In the Smit system, neither of these assumptions is true, and as a result, the chemical potential assumptions do not hold. These issues could likely be ameliorated using a more traditional interaction model.

We next performed simulations of single micellar aggregates, in both all-atom and implicit solvent form. We used a simulation cell of twenty  $H_1T_4$  surfactant molecules, for a chain number density of  $5 \times 10^{-3} \sigma_W^{-3}$ , three orders of magnitude above the CMC reported by Pool<sup>18</sup>. The all-atom simulation also included 3418 explicit W particles, for a total molecule number density of  $0.6 \sigma_W^{-3}$ .

The RDFs for H-H, H-T, and T-T interactions are given in Figure 4-10. Visual examination reveals that the density dependent implicit solvent model reproduces the underlying all-atom simulation. The energy error for the H-H, H-T, and T-T RDFs are  $0.08\epsilon$ ,  $0.08\epsilon$ , and  $0.18\epsilon$  respectively. These errors, though small, are significantly higher than the errors in mixtures of monomers. Additionally, the DDIS potentials have solute enhancement ratios significantly lower than the corresponding all-atom simulations, as shown in Table 3, indicating a less densely packed micelle than the corresponding all-atom simulation. These results confirm that the packing of a micellar aggregate of the size 20 is less dense using the DDIS potential than the corresponding all-atom potential. This result may arise from the weaker aggregation found in fitting the T-type potentials, and borne out by Figure 4-3. We note that the DDIS potential is sufficient to maintain the micellar aggregate as a coherent object despite the lack of explicit solvent particles

applying an external force. However, it appears that adding molecular connectivity contributes to a decrease in the RDF transferability.

We also measured  $\mu^{ex}$  for surfactant molecule insertion in both the all-atom and DDIS cases. The  $\mu^{ex}$  measured here represents the weighted average excess chemical potential of a surfactant molecule added to either the micellar aggregate or the surrounding free solution. For the all-atom case, we measured  $\mu^{ex}=3.28\pm 0.05\epsilon$ , while the DDIS case yielded  $6.25\pm 0.02\epsilon$ , a difference of approximately  $3\epsilon$ . This discrepancy is smaller than that seen in the monomeric mixture simulations. Thus, the transferability of  $\mu^{ex}$  does not degrade upon the addition of molecular connectivity. This finding is consistent with the findings of Chapter 3, which showed that chain molecules at low density demonstrated good excess chemical potential transferability.

## 4.5 CONCLUSIONS

In this work, we have created coarse-grained, density dependent implicit solvent (DDIS) potentials based on underlying all-atom simulations of a truncated and shifted Lennard-Jones model that has previously been shown to demonstrate micellization behavior<sup>11-13</sup>. Coarse-grained potentials were generated as we have previously outlined<sup>24,25</sup>. Potentials were fit to simulations of monomeric (i.e. single repeat units) solute particles in solvent, without explicit consideration of mixtures as inputs to the process. These simple fitting simulations were used to test the transferability of DDIS potentials for simulation environments far from the state point of fitting.

We performed large simulations of surfactant solutions at solute densities near the CMC. The ratio of CMCs between  $H_1T_4$  and  $H_1T_5$  surfactants was the same as estimated by Pool and Bolhuis<sup>18</sup>, indicating the utility of the DDIS potentials for comparative CMC studies. However, the measured DDIS CMCs were higher than the all-atom estimates by at least an order of magnitude. We attribute the discrepancy to the approximate nature of the fitting conditions used to generate the DDIS potentials, which did not explicitly include either chains or mixtures. The average aggregation number generated by the DDIS potential was larger than the value estimated by free energy methods for both  $H_1T_4$  and  $H_1T_5$  surfactants. Additionally,  $H_1T_5$  surfactant solutions displayed cylindrical micelle growth behavior using the DDIS potential, while Pool and Bolhuis predicted spherical micelle growth. These results indicate a less dense packing of the micelles using the DDIS potential compared to the all-atom results.

To further analyze the transferability, we performed a series of simulations with the resulting DDIS potentials that progressed from more to less similar to the state point of fitting. We first simulated monomeric mixtures of solute particles resulting in reproduction of the RDF with accuracy comparable to that at the state point of fitting. By contrast, the accuracy of reproduction in  $\mu^{ex}$  of the mixture was not as good. This degradation in accuracy can be attributed to the treatment of cutoff lengths in the Smit model, where H-T interactions are fundamentally different than the H-H and T-T interactions. We then simulated an  $H_1T_4$  surfactant system at a concentration well above the CMC, which revealed that the DDIS potential produced a micellar aggregate of lower density than the corresponding all-atom simulation, and an excess chemical potential too



high by approximately  $3\varepsilon$ . These results indicate that a loss in RDF transferability occurs upon the creation of molecular connectivity, but that the corresponding loss in the accuracy of reproduction in  $\mu^{ex}$  is small. This is consistent with the findings of our previous work<sup>25</sup>.

The design choice of fitting DDIS potentials to monomeric solute data is evaluated here as a baseline with which to compare alternate fitting simulations. It appears that transferability degrades by degree as simulations become less similar to the original fitting simulations. A significant loss of reproduction occurs in the transition from monomeric particles to chain molecules. In addition, the marginal cost of fitting DDIS potentials to simulations of pure chains in solvent is negligible. This is because fitting to chain data does not introduce any of the combinatorial issues introduced by fitting to mixture data. Based on these facts, a potential fruitful area of future research may be to fit DDIS potentials to systems of chains, rather than systems of monomers. This choice would also be consistent with the Effective Energy Function<sup>23</sup>, which is fit to data on short chain molecules.

## 4.6 REFERENCES

1. S. Pubbada and D. Blankschtein, *J. Chem. Phys.* **92**, 3710 (1990).
2. L. Maibaum, A.R. Dinner, and D. Chandler, *J. Phys. Chem. B* **108**, 6778 (2004).
3. B. Jonsson, O. Edholm, and O. Teleman, *J. Chem. Phys.* **85**, 2259 (1986).
4. K. Watanabe, M. Ferrario, and M.L. Klein, *J. Chem. Phys.* **92**, 819 (1988).
5. A.D. MacKerell, *J. Phys. Chem.* **99**, 1846 (1995).
6. D.P. Tielman, S.J. Marrink, and H.J.C. Berendsen, *Biochim. Biophys. Acta* **1331**, 235 (1997).
7. M. Tarek, S. Bandyopadhyay, and M.L. Klein, *J. Mol. Liq.* **78**, 1 (1998).
8. S.J. Marrink, D.P. Tieleman, and A.E. Mark, *J. Phys. Chem. B* **104**, 12165 (2000).
9. D.P. Tielman, D. van der Spoel, and H.J.C. Berendsen, *J. Phys. Chem. B* **104**, 6389 (2000).
10. N. Yoshii, K. Iwahashi, and S. Okazaki, *J. Chem. Phys.* **124**, 184901 (2006).
11. B. Smit, *Phys. Rev. A* **37**, 3431 (1988).
12. B. Smit, P.A.J. Hilbers, K. Esselink, L.A.M. Rupert, N.M. van Os, and A.G. Schlijper, *J. Phys. Chem.* **95**, 6361 (1991).
13. B. Smit, K. Esselink, P.A.J. Hilbers, N.M. van Os, L.A.M. Rupert, and I. Szleifer, *Langmuir* **9**, 9 (1993).
14. P.G. Bolhuis and D. Frankel, *Physica A* **244**, 45 (1997).
15. M. Kinoshita and Y. Sugai, *Chem. Phys. Lett.* **313**, 685 (1999).
16. S.O. Nielsen, C.F. Lopez, G. Srinivas, and M.L. Klein **119**, 7043 (2003).
17. S.J. Marrink, A.H. de Vries, and A.E. Mark **108**, 750 (2004).

18. R. Pool and P.G. Bolhuis, *J. Chem. Phys. B* **109**, 6650 (2005).
19. M. Hatakeyama and R. Faller, *Phys. Chem. Chem. Phys.* **9**, 4662 (2007).
20. H. Sinto, S. Marisada, M. Miyahara, and K. Higashitani, *Langmuir* **20**, 2017 (2004).
21. T. Lazaridis, B. Mallik, and Y. Chen, *J. Phys. Chem. B* **109**, 15098 (2005).
22. F.K. von Gottberg, K.A. Smith, and T.A. Hatton, *J. Chem. Phys.* **106**, 9850 (1997).
23. T. Lazaridis and M. Karplus, *Prot. Struct. Funct. Genet.* **35**, 133 (1999).
24. E.C. Allen and G.C. Rutledge, *J. Chem. Phys.* **128**, 154115 (2008).
25. E.C. Allen and G.C. Rutledge, *J. Chem. Phys.* *in press*.
26. J.N. Israelachvili, D.J. Mitchell, and B.W. Ninham, *J. Chem. Soc., Faraday Trans. 2* **72**, 1525 (1976).
27. G. Malescio and G. Pellicane, *Nature Mater.* **2**, 97 (2003).
28. M.A. Glaser, G.N. Grason, R.D. Kamien, A. Kosmrlj, C.D. Santangelo, and P. Ziherl, *Europhys. Lett.* **78**, 46004 (2007).

<i>Particle</i>	$\epsilon_{ii}$	$\sigma_{jj}$	$\Delta G_I(W \rightarrow I)$	$\langle N_L \rangle / \langle N_L \rangle^{IDEAL}$
<i>Type</i>				
H	1.0	2.0	$13.6 \pm 0.1 \epsilon_W$	0.95
T	1.0	1.0	$-7.04 \pm 0.05 \epsilon_W$	1.41
W	1.0	1.0	0	1.00

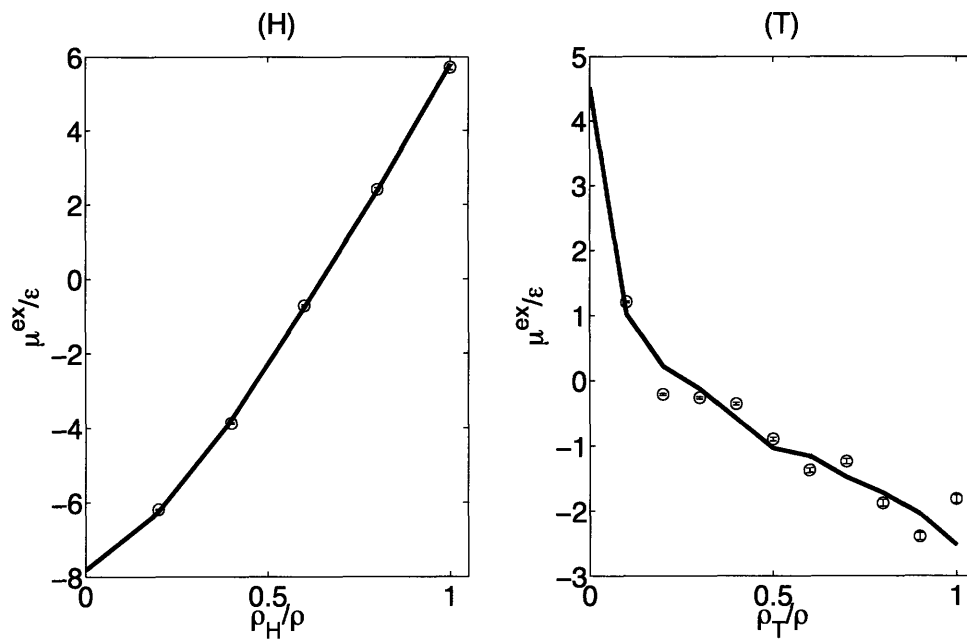
**Table 4-1:** Key parameters for all-atom particle types. H-type particles are “solvent-philic”, as indicated by the positive free energy of transfer. T-type are “solvent-phobic”, as indicated by a negative free energy of transfer and high solute enhancement ratio.

<i>Simulation</i>	<i>CMC (<math>\sigma_w^{-3}</math>)</i>			<i>N<sub>AGG</sub></i>	
	<i>H<sub>1</sub>T<sub>4</sub></i>	<i>H<sub>1</sub>T<sub>5</sub></i>	<i>CMC</i>	<i>H<sub>1</sub>T<sub>4</sub></i>	<i>H<sub>1</sub>T<sub>5</sub></i>
			<i>Ratio</i>		
All-Atom	$5 \times 10^{-6}$	$6 \times 10^{-7}$	0.12	~20	~30
DDIS	$1.9 \times 10^{-4}$	$2.0 \times 10^{-5}$	0.11	26	Cylindrical

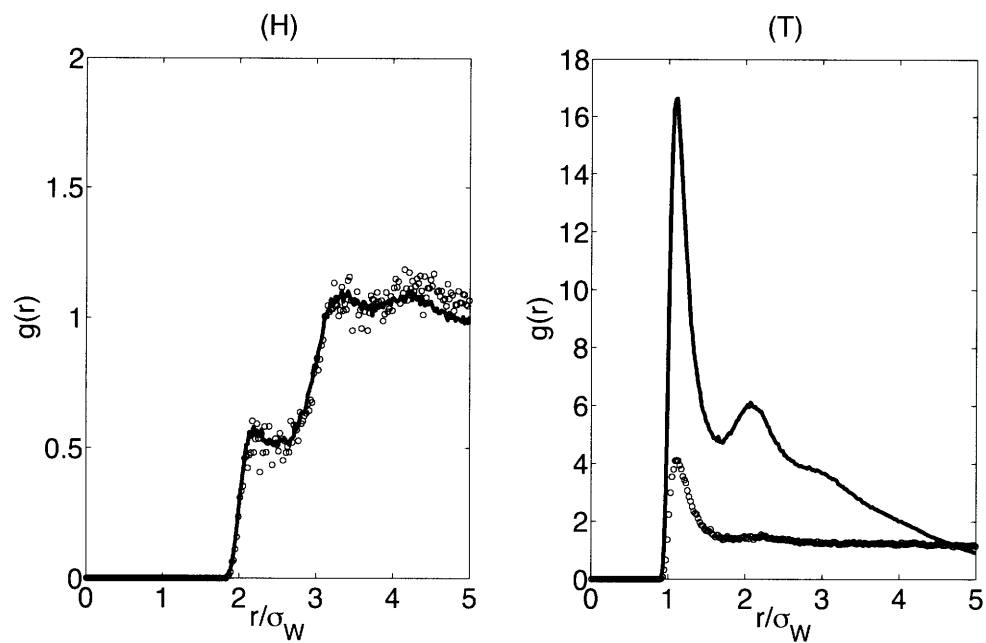
**Table 4-2:** Behavior of H<sub>1</sub>T<sub>4</sub> and H<sub>1</sub>T<sub>5</sub> solutions using the all-atom and DDIS potentials.

<i>Interaction Type</i>	<i>SER AA</i>	<i>SER DDIS</i>
H-H	3.4	3.2
H-T	6.2	5.6
T-T	9.5	8.5

**Table 4-3:** Solute enhancement ratios (SERs) for interactions in single micellar aggregate simulations. “AA” denotes all-atom simulation, “DDIS” density dependent implicit solvent.

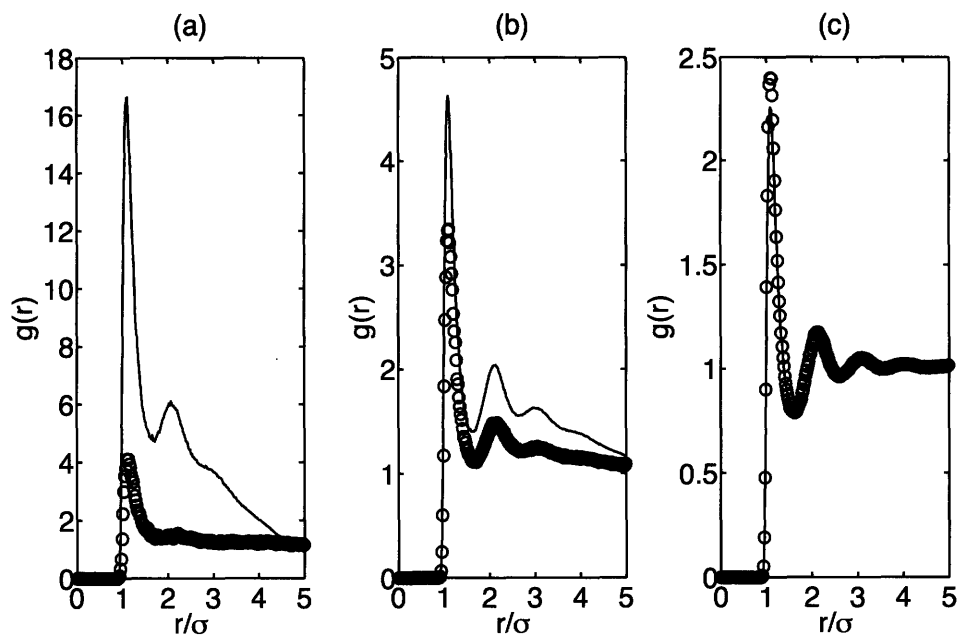


**Figure 4-1:** Comparison of excess chemical potential in all-atom (solid line) and coarse-grained (circles) simulations. Left: H-type particles in solvent W. Right: T-type particles in solvent W.

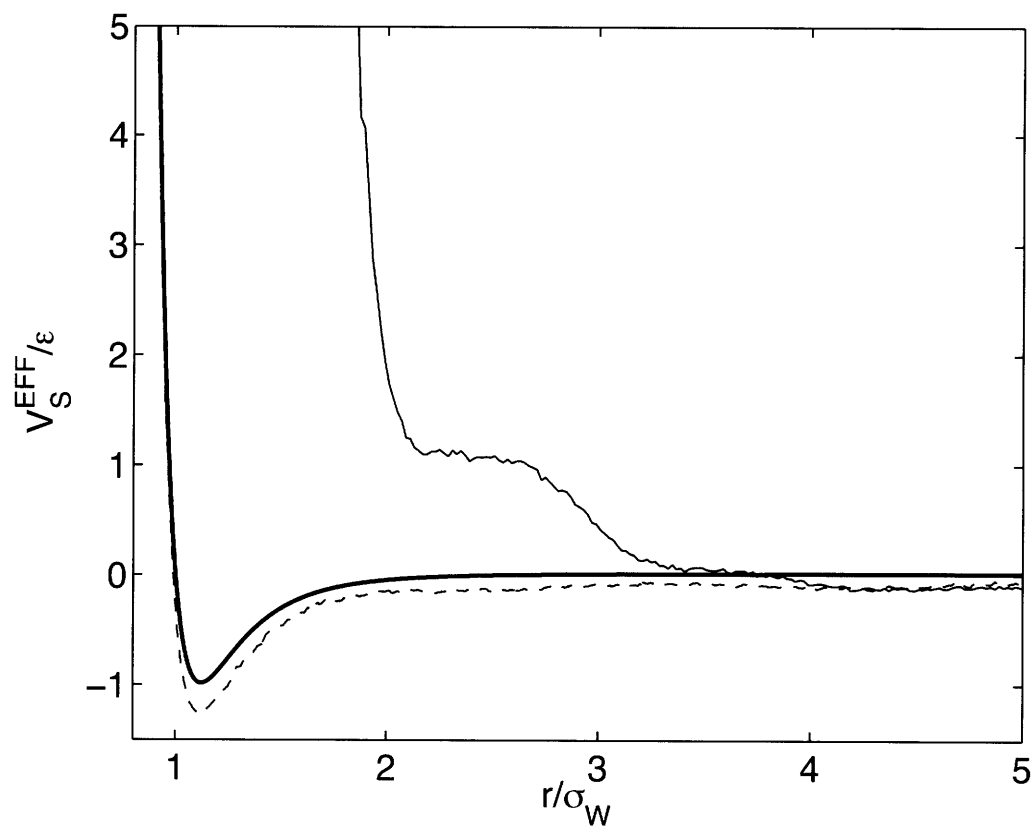


**Figure 4-2:** Comparison of worst-case fits of solute RDF in all-atom (solid line) and coarse-grained (circles) simulations. Left: H-type particles in solvent W at  $\rho_H=0.2$ . Right: T-type particles in solvent W at  $\rho_T=0.1$ .



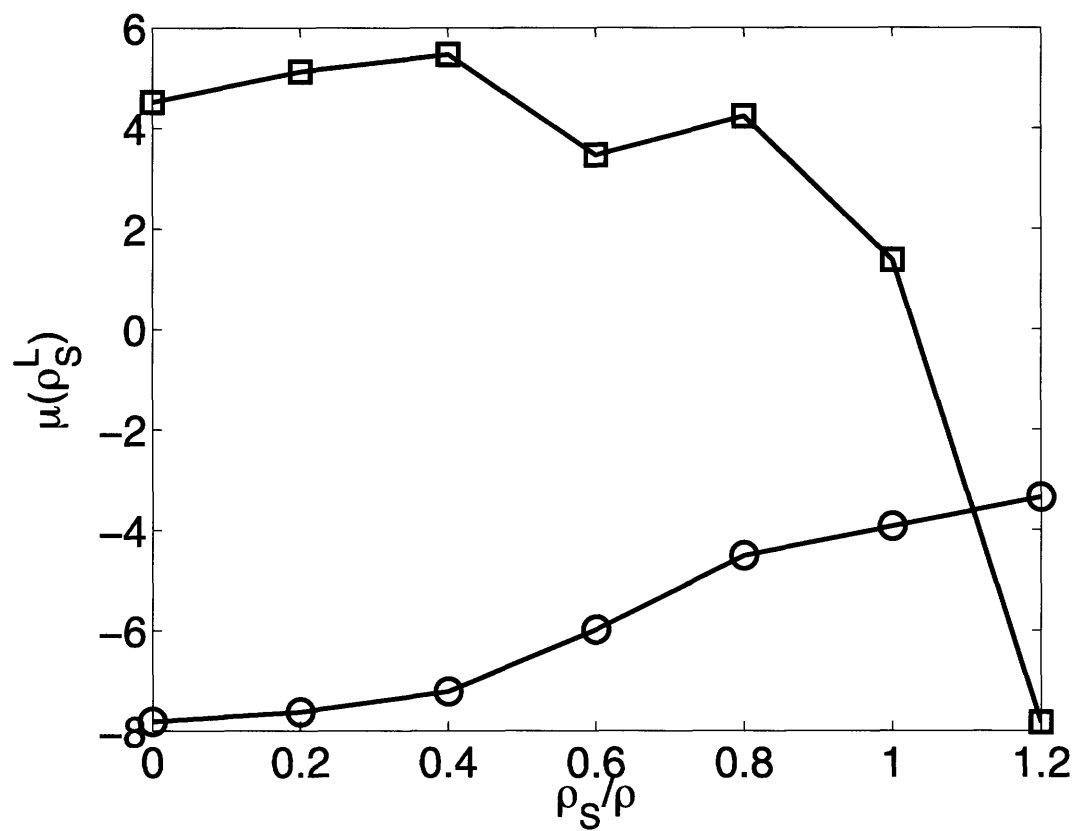


**Figure 4-3:** Comparison of RDFs in all-atom (solid line) and coarse-grained (circles) simulations. (a):  $\rho_T/\rho=0.1$ . (b):  $\rho_T/\rho=0.5$  (c):  $\rho_T/\rho=1.0$ .

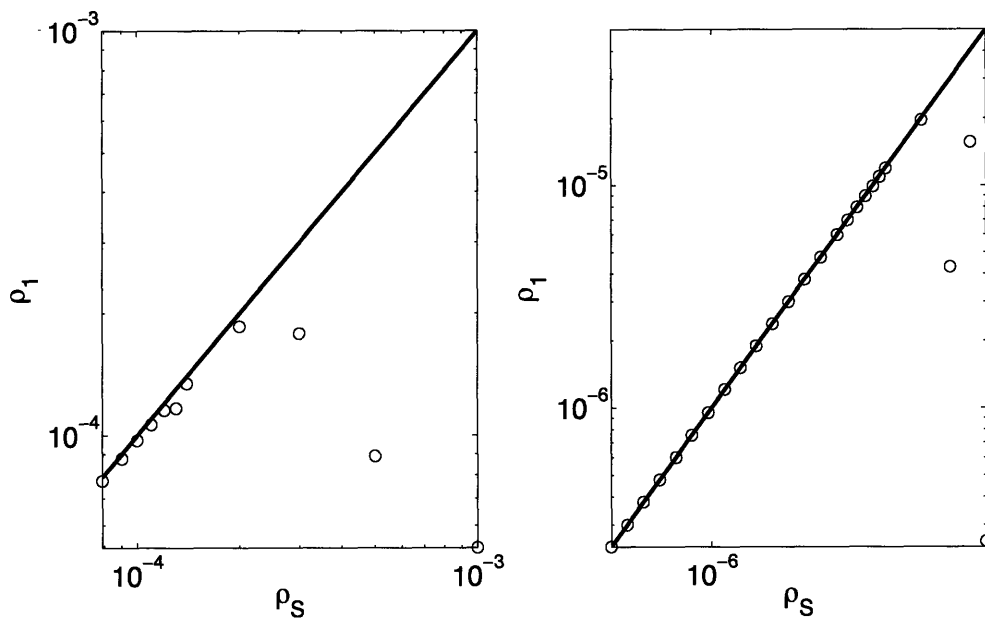


**Figure 4-4:** Coarse-grained two-body term for local solute density  $\rho_s^L/\rho=0.2$  (—H,--T).

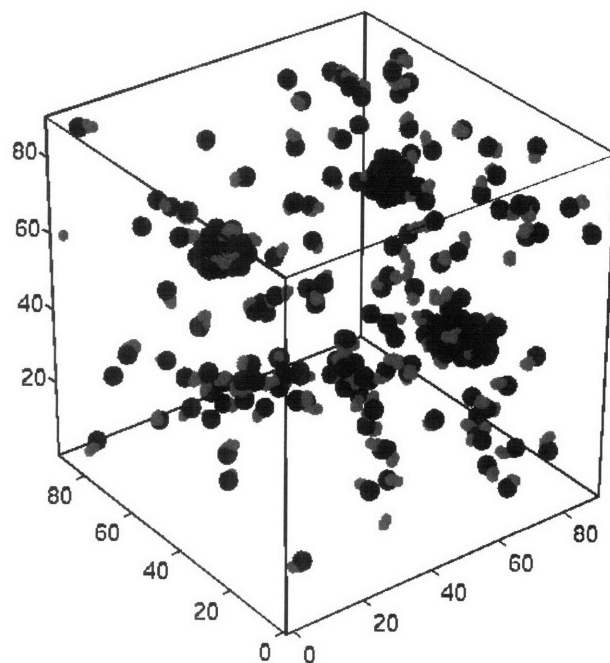
The dark line shows the T-type Lennard-Jones interaction for comparison.



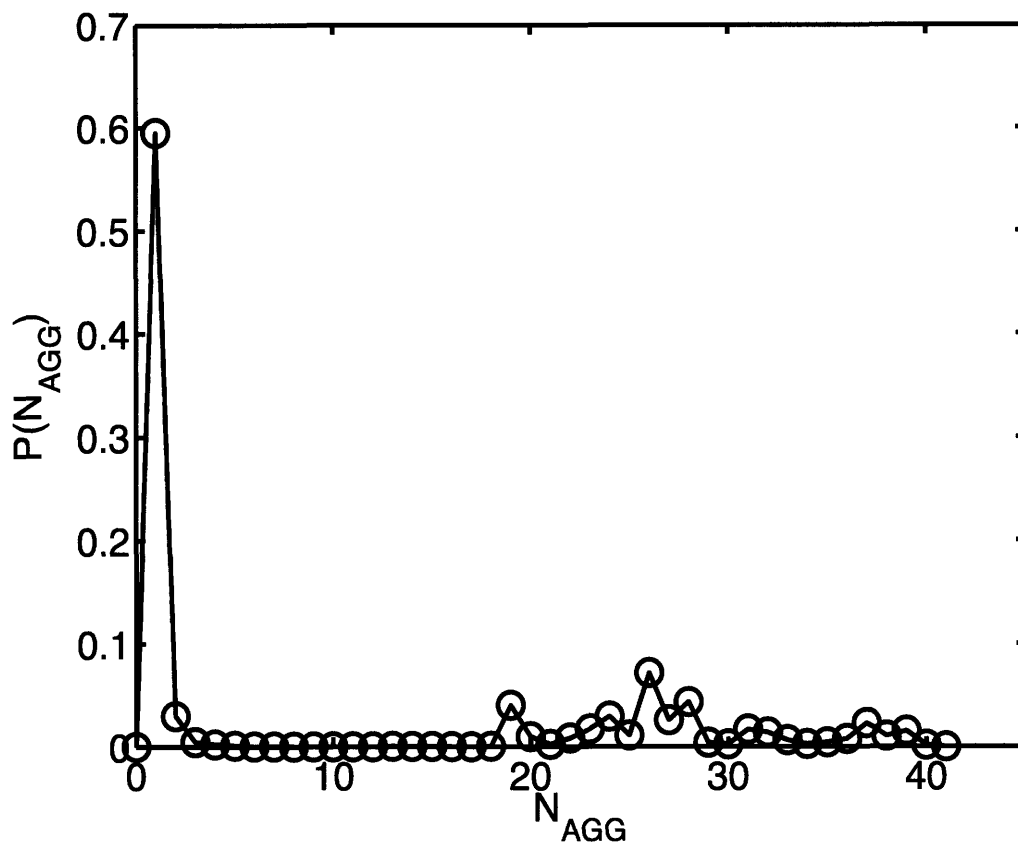
**Figure 4-5:** Coarse-grained one-body term as a function of local solute density (Circles: H, Squares: T)



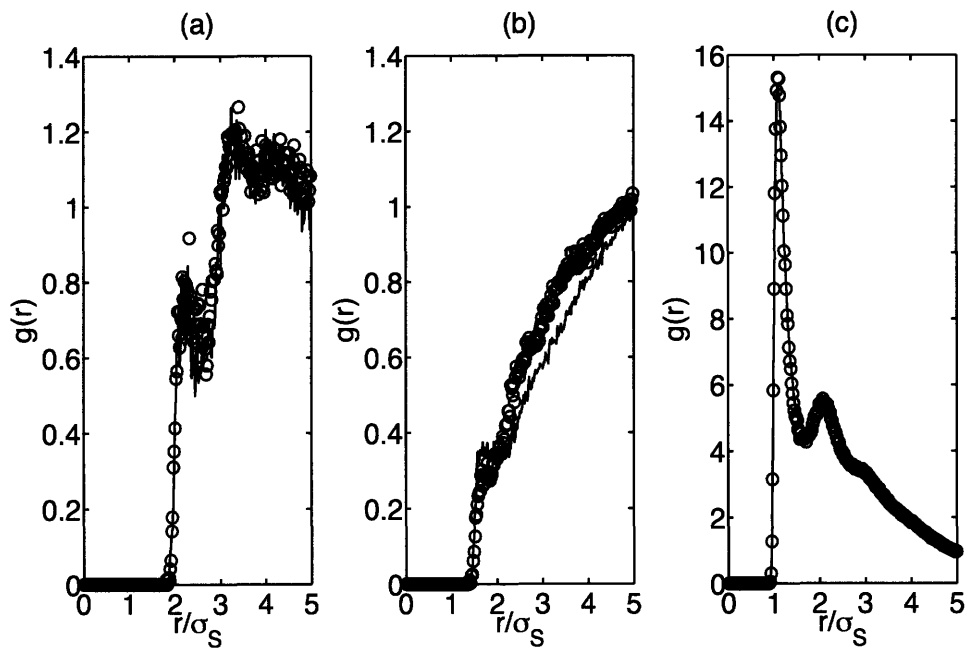
**Figure 4-6:** Free surfactant density (circles) as a function of total surfactant density. Solid line is the 45° line corresponding to a condition of no micellar aggregates. Left:  $H_1T_4$ ; Right:  $H_1T_5$ .



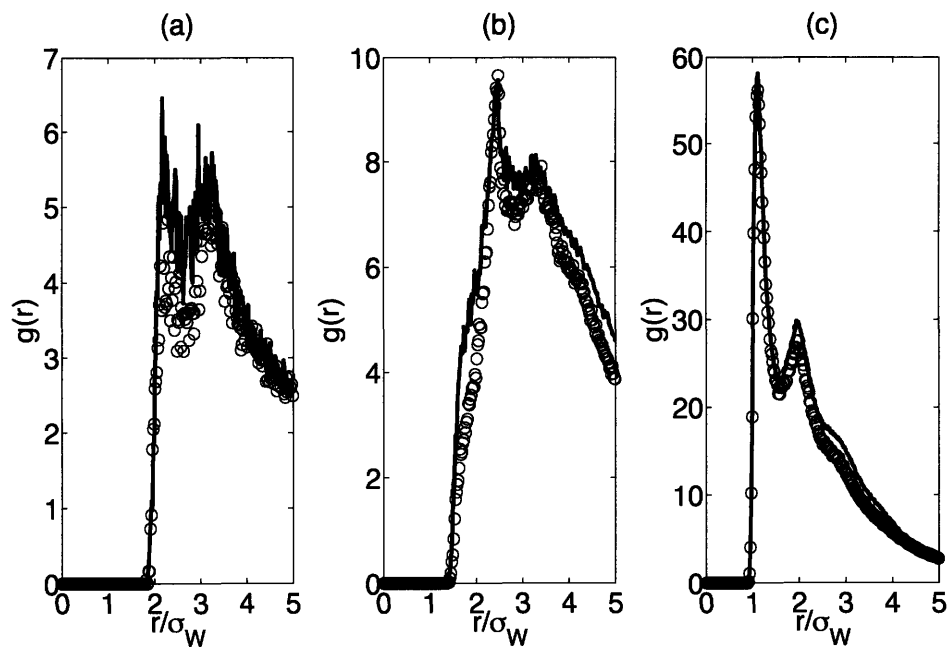
**Figure 4-7:** Simulation snapshot of 216 H<sub>1</sub>T<sub>4</sub> surfactant molecules in implicit solvent showing the formation of micellar aggregates. H-type particles are black, T-type particles are grey.



**Figure 4-8:** Micelle aggregation number distribution for H<sub>1</sub>T<sub>4</sub> surfactants.



**Figure 4-9:** Comparison of all-atom (solid lines) and implicit solvent (circles) RDFs in a 1:4 H:T mixture at  $\rho_S/\rho=0.3$ . (a): H-H RDF; (b) H-T RDF; (c) T-T RDF.



**Figure 4-10:** Comparison of all-atom (solid lines) and implicit solvent (circles) RDFs for a single micelle composed of 20  $H_1T_4$  surfactants in 3418 solvent molecules W. (a): H-H RDF; (b) H-T RDF; (c) T-T RDF.



## 5 Conclusions and Future Directions

### 5.1 CONCLUSIONS

This thesis has studied the issue of coarse-grained molecular simulation within the specific application of surfactant systems. We have focused on utilizing CG techniques to create an implicit solvent model between solute particles. This thesis has succeeded in accomplishing each of the three objectives given in the introduction. These objective are considered here in turn:

- **Develop a coarse-graining methodology to generate DDIS potentials from all-atom simulations of monomeric solute in solvent.**

Chapter 2 introduced a CG methodology to create DDIS potentials that reproduced both the solute-solute radial distribution function (RDF), and the excess chemical potential ( $\mu^{ex}$ ) of solute particle insertion. These properties were chosen for two reasons: 1) they are identified as having relevance in statistical thermodynamic models of micelle formation, and 2) they are general to all molecular simulations and measurable in a computationally reasonable amount of time. Approaches that address only one of these factors in describing micellization are likely to be incomplete. If a CG model only matches the excess chemical potential while ignoring the RDF, for example, it will likely incorrectly predict the solute packing density and therefore the average micelle aggregation number. By contrast, if a CG model only matches the RDF while ignoring

the excess chemical potential, the free energy driving force for micellization will likely be incorrect, and the value of the CMC will be incorrect.

We demonstrated that a DDIS potential could reproduce both properties simultaneously with the following functional form:

$$E_i = \mu(\rho_{S,i}^L) + \frac{1}{2} \sum_{j \neq i} V_{IJ}^{EFF}(r_{ij}, \rho_{S,i}^L)$$

where  $E_i$  is the effective energy of particle  $i$ ,  $V_{IJ}^{EFF}$  is a pairwise potential between  $I$ - and  $J$ -type solute particles that is an explicit function of the solute particle density in the vicinity of particle  $i$ ,  $\rho_{S,i}^L$ , and  $\mu$  is a “self-interaction” term that is also a function of solute particle density. The first term serves to capture the effects of solute-solvent and solvent-solvent interactions in a mean field approximation, while the second term represents solute-solute interactions that are mediated by solvent.

Although we investigated parameterizing the energy function by both the global and local solute density, we suggest that a parameterization based on local solute density is likely to be more effective in systems with high solute heterogeneity. The dependence of both terms on local density allows the potential to be cognizant of the greater or lesser influence of solvent in the different regions of a system that are inhomogeneous on the solute length scale, such as that which arises in a system of surfactants that self-assemble into micelles.

- **Examine the transferability of the generated DDIS potentials to other state points without the use of further optimization.**

In Chapter 3, we derived our coarse-grained DDIS potentials from simulations of pure solute particles, and investigated their transferability to mixtures and chains. The key findings here were that the transferability of behavior was quite effective at low concentrations for mixtures, chains, and chains of mixed composition. One of the biggest successes that paved the way for surfactant simulation was that the DDIS potentials predicted that chains of solvent-phobic particles would aggregate in solution, even though the monomers of the solvent-phobic particles did not. Additionally, we found that the local DDIS potential produced superior RDF transferability when compared to an analogous global density-dependent implicit solvent potential for the same system. The DDIS potential demonstrated a behavior quite different from the behavior of the system from which it was derived, while the globally density dependent potential did not. The most likely reason for this improved transferability is that the local density dependence captures solute enhancement effects more accurately than the global potential.

- **Evaluate the performance of the DDIS potentials to predict thermodynamic properties of surfactant solutions, including critical micelle concentration.**

In Chapter 4, we created coarse-grained, DDIS potentials based on underlying all-atom simulations of a truncated and shifted Lennard-Jones model (the “Smit model”) that has previously been shown to demonstrate micellization behavior. Potentials were fit to simulations of monomeric (i.e. single repeat units) solute particles in solvent, and no

solute mixtures were used as fitting inputs. These simple fitting simulations were used because we wished to establish the transferability of DDIS potentials to simulation environments far from the state point of fitting.

The DDIS potentials were able to accurately predict the trends in critical micelle concentration (CMC) for two surfactant types, meaning that the DDIS potentials could have utility for comparative CMC studies. However, the absolute values of the predicted CMC were an order of magnitude higher than previously established estimates for the same surfactants using atomistic simulations. Additionally, the formed micelles were less densely packed than the corresponding all-atom micelles, leading to a larger average aggregation number. In sum, the performance of the DDIS potentials can be considered a success, given the simplicity of the simulations from which they were derived. However, the accuracy did not meet the standards we came to expect from Chapter 3.

To further investigate the transferability of the DDIS potentials, we performed a series of simulations that progressed from more to less similar to the state point of fitting.

*Mixture Simulations:* We first simulated monomeric mixtures of solute particles, which resulted in RDF reproduction accuracy at least equal to the state point of fitting. By contrast, the  $\mu^{ex}$  reproduction accuracy of the mixture simulations was quite poor. This poor reproduction can be attributed to the highly non-ideal nature of the Smit model, where H-T interactions are fundamentally different than the H-H and T-T interactions. We expect that for more traditional mixing rules, the DDIS potentials would provide

more reliable property reproduction, as is consistent with the results of Chapter 3. Thus, a large part of the misbehavior of the DDIS potentials can be attributed to a quirk of the Smit potential.

*Single Micelle Simulations:* We then simulated an H<sub>1</sub>T<sub>4</sub> surfactant system at a concentration well above the CMC. The measured  $\mu^{ex}$  of the chain system was actually closer to the all-atom value than it was in the mixture simulations. We therefore conclude that the transferability of  $\mu^{ex}$  does not degrade upon the addition of molecular connectivity. This finding is consistent with the findings of Chapter 3, which showed that chain molecules at low density demonstrated good excess chemical potential transferability. By contrast, the RDF reproduction in the single micelle simulations was considerably worse than in the chain simulations. This is also consistent with Chapter 3, in which we demonstrated that adding molecular connectivity reduced RDF reproduction for systems of Lennard-Jones particles.

The results of these simpler simulations demonstrate that the discrepancy between the all-atom and CG behavior in surfactant solutions can be attributed to two factors:

1. Improper reproduction of  $\mu^{ex}$  in mixtures, caused by the unique nature of the Smit model.
2. Improper reproduction of the RDF in chains. This weakness could perhaps be ameliorated in future research by deriving DDIS potentials directly from simulations of solute chains in solvent, rather than the current practice of solute monomers in solvent.

## 5.2 FUTURE RESEARCH

While I feel this thesis has made significant strides towards a method for creating DDIS potentials, there are numerous avenues for continued research. Broadly, these ideas can be categorized into topics relating to methodology and topics relating to applications. We consider both in turn:

### 5.2.1 Methodological Improvements

- The design choice of fitting DDIS potentials to monomeric solute data was evaluated as a baseline with which to compare alternate fitting simulations. It appears that transferability degrades by degree as simulations become less similar to the original fitting simulations. A significant loss of reproduction occurs in the transition from monomeric particles to chain molecules. In addition, the marginal cost of fitting DDIS potentials to simulations of pure chains in solvent is negligible. This is because fitting to chain data does not introduce any of the combinatorial issues introduced by fitting to mixture data. Based on these facts, a potential fruitful area of future research may be to fit DDIS potentials to chain, rather than monomeric, data. This choice would also be consistent with the Effective Energy Function, which is fit to data on short chain molecules.
- Examining the shape of the density function at more depth. Particles within the cutoff radius,  $r_D$ , contribute to the local density through some weighting function  $w(r_D)$ . In Chapter 2, we examined a constant weighting ( $w(r_D)=1$ ), a linear

weighting ( $w(r_D)=1-r/r_D$ ), and a quadratic weighting ( $w(r_D)=(1-r/r_D)^2$ ). In that Chapter, we found little difference in the behavior of Monte Carlo simulations to the choice of weighting function. However, those simulations represented a special case of identical solute and solvent particles. It is likely that the choice of weighting function will impact the transferability of the derived DDIS potential to other state points. Whether there exists an optimum weighting that is applicable to all state points, or is rather situation dependent, is still unknown.

- Examining the effect of bonded interactions in more depth. In all our simulations, we kept the bonded interactions unchanged between all-atom and CG simulations. For nearest-neighbor bonds, this is a good approximation. The length scale of interaction is too short, and the strength of interaction too strong, to be significantly affected by the solvent environment. However, the local solvent environment could certainly affect bonded interactions beyond nearest neighbor (angle and dihedral interactions). Although the all-atom potentials utilized in this model contained no angle or dihedral terms, an optimized CG potential might. As with the non-bonded potential, we would anticipate that the bonded potentials would be density dependent.
- Related to the bonded interaction effects, further research should be done to investigate the optimum density weighting for bonded particles in the local environment. In our work, we assumed that the density weighting for bonded particles was identical to that of non-bonded particles. This assumption may well prove to be too simplistic.

- Finally, I believe the power of DDIS potentials for surfactant simulation will truly become apparent when they are incorporated into a Molecular Dynamics (MD) or Brownian Dynamics (BD) code. For this work, we utilized Monte Carlo (MC) simulation because I had previously created MC software that I was comfortable modifying as necessary. However, MC is not the ideal computational tool for DDIS potentials. The reason for this is that the MC move of a single particle causes changes in the local density of a very large number of neighboring particles. As a result, the calculation of energy for each time step can be computationally intensive relative to a density independent potential. By contrast, the coordinates of every particle change at every time step in MD and BD. As such, the computational load of DDIS potentials is fixed at twice a comparable density independent potential. This scaling arises from the fact that the DDIS potential requires two pairwise interaction loops: one to calculate the local density of each particle, and one to calculate the energy of each particle. In addition to the energy calculation issue, MC simulation of surfactant simulations also suffers from a sampling issue. The transfer of surfactants into and out of micellar structures is very slow in MC, because the micelle structure is highly energetically favorable. As a result, it is very difficult to generate an equilibrated aggregation number distribution from MC. By contrast, the random force of BD allows for more rapid transfer of surfactant molecules, and could further speed computation. BD simulations are also of interest from a theoretical standpoint, as the random force itself should be considered density dependent in a completely general formulation of the problem.



## 5.2.2 Applications

In designing the DDIS protocol described in the thesis, I very deliberately made design decisions that allowed for the eventual simulation of chemical complex systems. The key point is that the DDIS potentials are derived from simulations of a single solute type in solvent. In Chapter 3, we show that these “pure solute” potentials are transferable to solute mixtures using a very simple mixing rule. This setup is advantageous, in that simulating an  $n$  component system  $n$  DDIS potentials rather than the  $n(n-1)/2$  that would be required in the completely general case.

Given the relatively robust transferability demonstrated by the DDIS potentials, one could consider creating a large number of DDIS potentials relating to commonly simulated solutes. These DDIS potentials could be derived once, and subsequently used in a wide variety of chemical environments, much as all-atom potentials GROMOS, TraPPE, and AMBER are used today. With a suitable library, DDIS potentials could be used to study any solvated system in which local heterogeneity occurs and solvent constitutes a large fraction of the system. In addition to surfactant simulations, DDIS potentials could be used to study:

- Protein folding. The minimum energy configuration of a protein is determined by the pattern of hydrophilic and hydrophobic charge groups along the chain, which produce local ordering. The DDIS potential method is perfect for dealing with the wide range of solute densities experienced by different amino acids.

- Predicting the solubility of drug compounds. There is active research into the uptake of drug molecules into micellar structures. A thermodynamic equilibrium partitions the drug compounds between the solvent and micelle core. Once again, we see a situation in which the solute (the drug molecule) experiences a wide range of local density environments.
- Simulate lipid bilayers and other common biological systems. The formation of different biological structures is driven by the interplay of free energy ( $\mu^{ex}$ ) and packing effects (RDF).
- Crystallization studies. Many polymers form crystalline and semi-crystalline phases in solution. The behavior of these systems would be in many ways analogous to surfactant systems in that they contain small aggregates of high local density.

## **Appendix A      *Ab initio* study of the binding strength of POSS-cation complexes**

(Note: Portions of this work were published in E.C. Allen and K.J. Beers, *Polymer* **46**, 569 (2004). As it does not fit with the main thrust of this thesis, I include it here in the appendix)

### **A.1 INTRODUCTION**

Polyhedral oligomeric silsesquioxane (POSS) is a “cage-shaped” molecule comprised of a silicon and oxygen core (Figure A-1a) of chemical formula  $R_8Si_8O_{12}$ . The R groups can be short oligomeric units or long polymer chains, and are varied to affect POSS solubility/dispersion in polymer matrices<sup>1</sup>.

Among its applications, POSS has emerged as viable filler in high-performance nanocomposites<sup>2</sup>. POSS-polymer nanocomposites show increased toughness vs. the unfilled polymer.<sup>3</sup> In addition, POSS-PEO based polymer electrolytes have been developed<sup>4</sup>, with the POSS molecule acting as an inhibitor to PEO crystallization.

The four oxygen atoms on a single “face” of the POSS cage structurally resembles a crown ether molecule (Figure A-1b), and introduces the possibility that the POSS molecule, like crown ether, may contain an electron rich free center that attracts and binds cations<sup>5</sup>. The first motivation to investigate these binding phenomena is obvious: to study the role of POSS in POSS-PEO polymer electrolytes. The second, less obvious,

motivation is to introduce the possibility of reducing POSS agglomeration issues through the introduction of electrostatically-bound surfactants.

In both cases, an estimate of the magnitude of this binding strength is important to establish the feasibility of POSS-cation complex formation within polymeric media. If the binding energies are more than a few  $kT$ , one expects POSS-cation binding to persist with little regard to the surrounding environment.

The existence of POSS-metal ion complexes is not in question. Baker *et al.* generated POSS/ $\text{Na}^+$  ions in the gas phase through electrospray ionization (ESI) and matrix-assisted laser adsorption/ionization (MALDI)<sup>6,7</sup>. A molecular mechanics study by the same group predicted conformations in which the attached  $\text{Na}^+$  ion resides just outside a face of the POSS cage, attached by electrostatic interactions with the four nearest oxygens. They did not, however, present an estimate of the strength of this binding.

This paper comprises a feasibility study of POSS binding to  $\text{Li}^+$  and other cations through the use of *ab initio* quantum chemical calculations, with particular focus on the questions:

- 1) Is a POSS-cation complex strongly energetically favorable, and what is the magnitude of the binding strength of their interaction?
- 2) Is the presence of such complexes identifiable by IR?
- 3) Can cations be passed among POSS molecules?

## A.2 METHODOLOGY

This paper makes use of the GAUSSIAN98<sup>TM</sup> program, using both the semi-empirical UFF potential and DFT. Initial geometries were generated by molecular mechanics optimization using the UFF potential. The initial structure was further optimized via DFT calculations using the BLYP functional and the 6-311g(d) basis set. The 6-311g(d) basis set was chosen because it generated energies within ~5 kT of those obtained with larger basis sets (Table A-1) that are significantly more computationally intensive to use. Binding energies *in vacuo* were defined as  $E = - [E(\text{POSS-cation complex}) - E(\text{POSS}) - E(\text{Li}^+)]$ .

Geometry optimization of the POSS-cation complex requires an initial guess of the cation location with respect to the POSS molecule. Multiple trial cation positions were studied, including the interior of the cage. For the case of unsubstituted POSS and Li<sup>+</sup>, the minimum energy conformation finds the lithium ion located outside the cage, equidistant from the four oxygen atoms that define a single face (Figure A-2). Baker *et al.* predict a similar cation location for POSS-Na<sup>+</sup> through AMBER force field MD<sup>7</sup>. Therefore, subsequent calculations involving different cations or substituted POSS molecules utilized an initial cation position consistent with these findings.

When simulating corner ligand-substituted POSS with R = C<sub>6</sub>H<sub>5</sub> (e.g. Figure A-5), the R group (excluding the first carbon attached to Si) was modeled using the UFF force field, and the total energy computed via the ONIOM method.

The calculations that follow are presented without regard to entropic contributions to the free energy. In general, these entropy considerations will tend to decrease the Li<sup>+</sup> preference to bind to the POSS molecule. An order of magnitude estimate of the entropic effects can be generated by noting that, in the absence of Li<sup>+</sup> enthalpic preference to the polymer or POSS molecule, the fraction  $f$  of Li<sup>+</sup> bound to the POSS molecule should be approximately equal to the POSS volume fraction (~5%). From statistical mechanics, the entropy difference is then approximately:

$$T\Delta S \approx -\log\{f\} = 3.0 \text{ kT} \quad (1)$$

Therefore, if the Li<sup>+</sup> enthalpic preference for POSS is  $\gg 3 \text{ kT}$ , the entropic contributions can be considered negligible.

### A.3 RESULTS AND DISCUSSION

The effect of basis set size on binding energy calculations was first investigated on the unsubstituted POSS molecule and a single Li<sup>+</sup>. As shown in Table A-1, the binding energy in the limit of the very large basis set 6-311g+(2d,p) converges to ~78 kT. For subsequent calculations, we elected to use the 6-311g(d) basis set, since it provides accuracy of  $\pm 5 \text{ kT}$  at far less computational cost.

Clearly, BLYP-DFT predicts a significant POSS-Li<sup>+</sup> binding energy. Whether such an affinity is expected to persist within a polymer matrix depends upon the predicted *polymer*-cation interaction. For example, polyethylene oxide is quite effective at binding lithium ions, due to the documented ability of PEO to donate five or six coordinating oxygen to each ion. Johansson *et al.*, using quantum chemical calculations, estimate the

PEO-Li<sup>+</sup> binding strength at  $\sim 250 \text{ kT}$ <sup>7</sup>. In such a system, one expects negligible POSS-Li<sup>+</sup> complex formation at low POSS volume fractions.

By contrast, polystyrene (PS) is a relatively non-polar polymer. Using the Onsager reaction field model<sup>9</sup> to estimate the impact of the polystyrene solvent (dielectric constant  $\epsilon = 2.5$ ), simulation of POSS-Li<sup>+</sup> produces results nearly identical to those achieved *in vacuo*.

The minimum energy conformation places the Li<sup>+</sup> outside the POSS cage, equidistant from the four O atoms that define a face (Figure A-2). A profile of the system energy versus Li<sup>+</sup>/POSS face center distance (defined as the plane created by the four oxygen atoms) provides an energy profile. This energy profile is included as Figure A-3. Note that Figure A-3 indicates a large ( $\sim 30 \text{ kT}$ ) energy barrier to Li<sup>+</sup> migration to the POSS cage center, and that the local minimum inside the cage is significantly disfavored compared to that external to the face.

The strength of the POSS/Li<sup>+</sup> binding significantly distorts the POSS structure, drawing the Li<sup>+</sup>-adjacent oxygen atoms closer together than in the unbound state. Oxygen atoms on opposite sides of the same face are separated by  $3.88 \text{ \AA}$  in the unbound state, but for those nearest the Li<sup>+</sup>, this value decreases to  $3.62 \text{ \AA}$ .

This deformation of the POSS molecule breaks the symmetry present in the native structure, and has a significant affects the IR spectrum. Figure A-4 shows estimates of

the IR spectrum for native POSS and the POSS-Li<sup>+</sup> complex obtained from normal mode analysis. The results indicate the presence of a new peak in the bound state at 1000  $cm^{-1}$ , that corresponds to Si-O bond stretching. Figure A-4 was generated by assuming that the calculated IR peaks would be somewhat diffuse. Each peak was assumed to be normally distributed about the calculated value, with a standard deviation of 20  $cm^{-1}$ .

Now that the feasibility of POSS/Li<sup>+</sup> binding has been established, we investigate the effects of structural modifications of the POSS or cation. The results from all systems studied are provided in Table A-2 and discussed in turn below.

### **A.3.1 Addition of cyclohexyl side chains**

In polymer/POSS systems of interest, the terminal hydrogen atoms are replaced with an organic group to facilitate dispersion in the polymer matrix. One of the most common ligands is the cyclohexyl group. As described in the methodology section, this binding energy calculation was performed via the ONIOM method. The results indicate that the presence of the cyclohexyl side chains increases the binding strength from 80 kT to 110 kT.

### **A.3.2 Presence of a Cl<sup>-</sup> counterion**

Since lithium is invariably added to any polymer system as a salt, an anion must be present for each lithium cation. However, depending on the solvating abilities of the polymer matrix, the lithium and corresponding anion may or may not dissociate. POSS



association with LiCl *in vacuo* therefore provides a “worst case scenario” for binding.

The effect of the Cl<sup>-</sup> counterion is to reduce the binding energy (defined as  $E = [E(\text{POSS}) + E(\text{LiCl}) - E(\text{POSS/LiCl complex})]$ ) to ~21 kT, a still significant amount.

### **A.3.3 Replace Li<sup>+</sup> with CH<sub>3</sub>NH<sub>3</sub><sup>+</sup>**

CH<sub>3</sub>NH<sub>3</sub><sup>+</sup> was chosen as a model ion to represent potential surfactant molecules. While Li<sup>+</sup> binding is of potential interest for applications in polymer electrolytes, the ability of POSS to bind with cations opens another avenue to POSS modification in addition to corner ligand substitution. In particular, association with cationic surfactants might help to impede POSS aggregation, which occurs at volume fractions of more than a few percent and that generally degrades the nanocomposite properties. The binding of energy of CH<sub>3</sub>NH<sub>3</sub><sup>+</sup> with an unsubstituted POSS molecule was calculated as ~25 kT, suggesting that such modification is possible.

### **A.3.4 Feasibility of Lithium transfer between two POSS molecules**

As a final line of questioning, we consider the feasibility of cation transfer from one POSS molecule to another. One hypothetical transport mechanism is that the cation “jumps the gap” between two POSS molecules as they come in close contact (Figure A-5).

To test this hypothesis, we first minimized a system of two POSS molecules with an intermediate lithium ion. The initial guess structure contained the two POSS molecules

aligned so that the cyclohexyl groups abutted. In the optimized structure, by contrast, the POSS molecules are aligned by 45 degrees to each other, to allow sterically a closer approach of the POSS faces. We then performed a series of single energy calculations in which the position of the lithium ion was moved from one POSS molecule to the other while the positions of all other atoms remained constant. The transfer barrier calculated by this method is ~80 kT, implying a very low likelihood of lithium transfer. It should be noted, however, that the method described here does not allow for favorable changes in POSS conformation during ion transfer that may result in a lower barrier. Further study using a cheaper energy model is needed to investigate these effects.

#### A.4 SUMMARY

BLYP-DFT Calculations of POSS molecule binding to various cations indicate that the strength of this association is quite substantial. In the case of POSS-Li<sup>+</sup> complex formation, the strength of binding is estimated at room temperature as ~78 kT *in vacuo*, with negligible change due to the presence of a polystyrene solvent. Even in the most unfavorable circumstances (i.e. anion Cl<sup>-</sup> remaining tightly bound to Li<sup>+</sup>), the strength of binding remains ~20 kT. Further, binding was demonstrated for the surfactant-like cation, CH<sub>3</sub>NH<sub>3</sub><sup>+</sup>, indicating a possible mechanism by which to reduce POSS agglomeration in polymer nanocomposites.

## A.5 REFERENCES

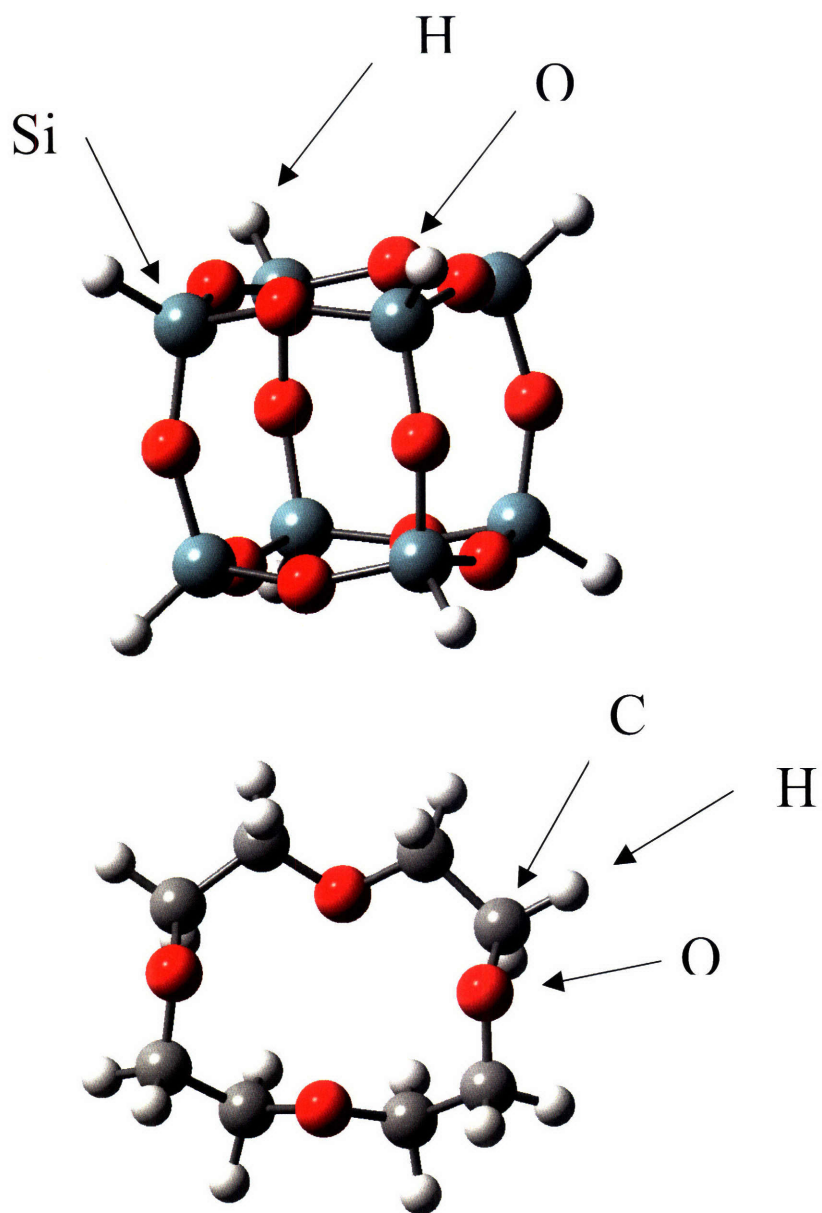
1. G. Li, L. Wang, H. Ni, and C.U. Pittman Jr., *Journal of Inorganic and Organometallic Polymers* **11**, 123 (2002).
2. B.X. Fu, B.S. Hsiao, S. Pagola, P. Stephens, H. White, M. Rafailovich, J. Sokolov, P.T. Mather, H.G. Jeon, S. Phillips, J. Lichtenhan, and J. Schwab, *Polymer* **42**, 599-611 (2001).
3. A. Romo-Uribe, P.T. Mather, T.S. Haddad and J.D. Lichtenhan, *J Polym Sci, Part B-Polym Phys* **36** (1998), 1857–1872.
4. P. Maitra, and S.L. Wunder, *Electrochemical and Solid-State Letters* **7**, A88-A92 (2004).
5. P.T. Mather, U. Conn. Materials Science. Personal communication.
6. E.S. Baker, J. Gidden, D.P. Fee, P.R. Kemper, S.E. Anderson, and M.T. Bowers, *International Journal of Mass Spectrometry* **227**, 205–216 (2003).
7. J. Gidden, P.R. Kemper, E. Shammel, D.P. Fee, S.Anderson, and M.T. Bowers, *International Journal of Mass Spectrometry* **222**, 63–73 (2003).
8. P. Johansson, J. Tegenfeldt, and J. Lindgren, *Polymer* **40**, 4399–4406 (1999).
9. M. W. Wong, M. J. Frisch and K. B. Wiberg, *J. Amer. Chem. Soc.* **113**, 4776 (1991).

<i>BASIS SET</i>	<i>Binding Energy (kT)</i>
3-21g	137
6-311g	100
6-311g (d)	85
6311g (d,p)	85
6-311g+(d)	79
6-311g+(d,p)	79
6-311g+(2d,p)	78

**Table A-1:** Energy Convergence with Respect to Basis Set

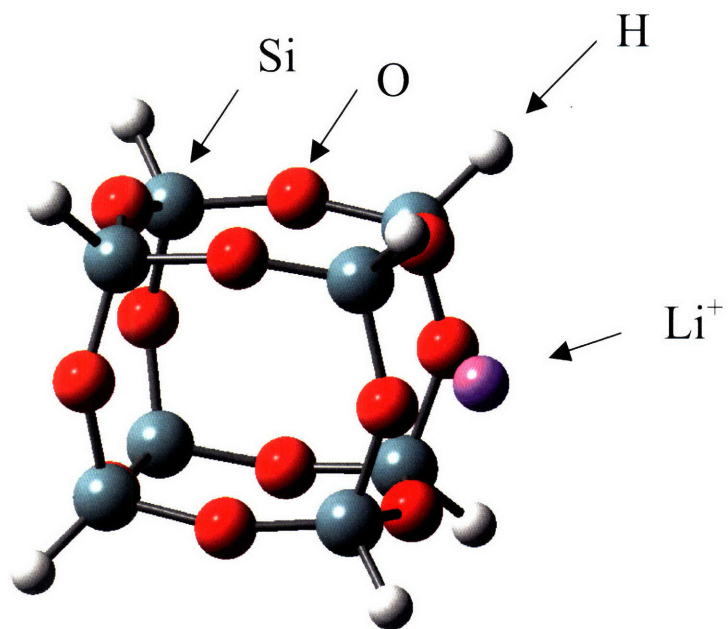
<i>POSS Terminal Groups</i>	<i>Cation</i>	<i>Counterion</i>	<i>Predicted Binding Energy (kJ)</i>
H	Li	None	78
C6H5	Li	None	120
H	Li	Cl	21
H	CH3NH3	None	25

**Table A-2:** Summary of Binding Energies for all systems studied

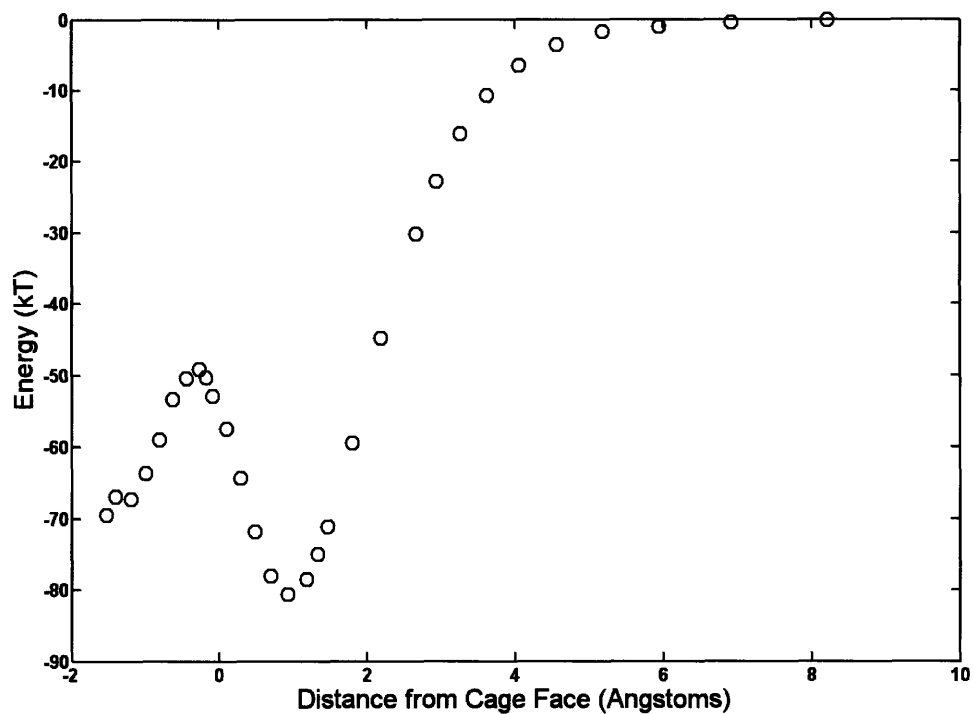


**Figure A-1:** Structural similarity of the POSS molecule to crown ether. (a) POSS (b)

Crown ether

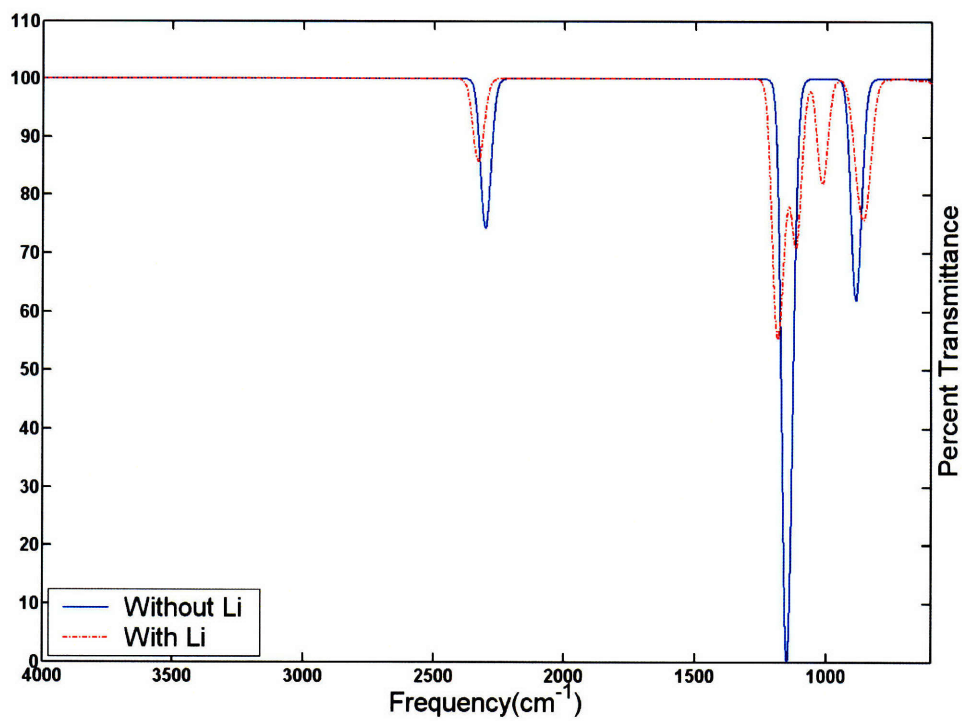


**Figure A-2:** GAUSSIAN98 output – POSS-Li<sup>+</sup> Complex

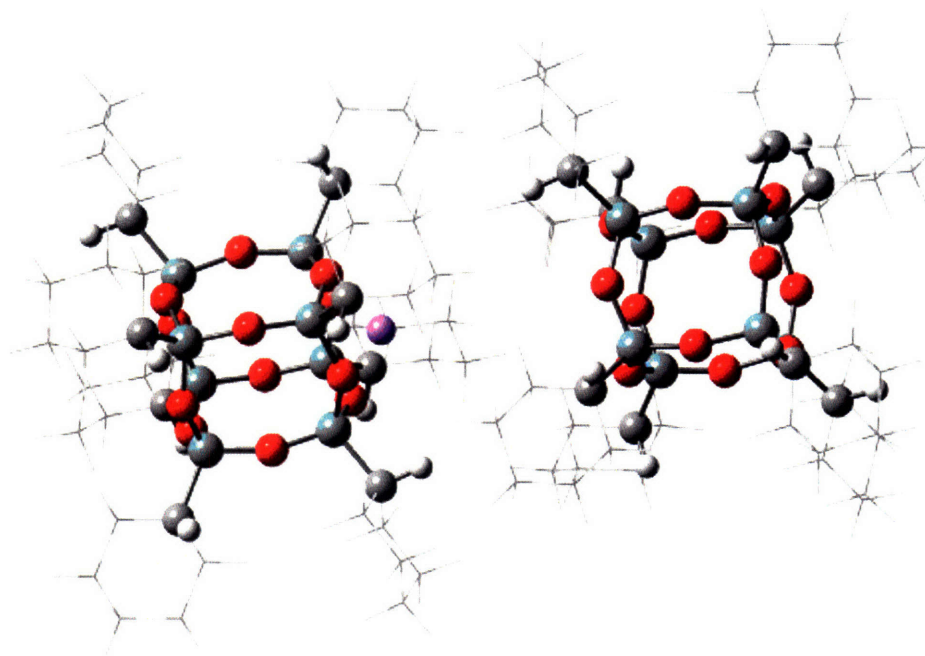


**Figure A-3:** POSS/Li<sup>+</sup> Energy as a function of lithium position. The energy minimum is achieved approximately 1 angstrom away from the POSS cage face





**Figure A-4:** Predicted IR Spectrum of POSS (solid) and POSS/Li<sup>+</sup> (dotted). DFT calculations predict the presence of an additional IR peak at ~1000cm<sup>-1</sup>



**Figure A-5:** Structure of two POSS molecules joined by attractive force of a single lithium ion. Results predicted by energy minimization in GAUSSIAN98.

## Appendix B Algorithms

### B.1 INTRODUCTION

This section provides some details regarding the calculation of density-dependent, implicit solvent (DDIS) potentials. Included are algorithms to calculate local density and system energy. For clarity, we follow the pseudo-code conventions of Frenkel and Smit's book (Daan Frenkel and Berend Smith, *Understanding Molecular Simulation*, Academic Press, 2002) in this section.

### B.2 CALCULATE LOCAL DENSITY

The calculation of the local density around each particle must be performed *prior to* the calculation of the system energy. This is because the system energy is explicitly dependent on the instantaneous local density.

#### Algorithm B-1: Calculate Local Density

```
subroutine calc_local_density
do i=1, npart-1
  do j=i+1, npart
    xr = x(i) - x(j)                ;xr is particle distance
    if (xr. gt. hbox) then          ;correct for PBCs
      xr = xr - hbox
    else if (xr. lt. -hbox) then
      xr = xr + hbox
    endif
    if (abs(xr) . lt. rD) then      ;rD is density cutoff radius
      density(i) += w(xr)          ;w(xr) is weighting function
      density(j) += w(xr)          ;for density
    endif
  enddo
enddo
```

### B.3 CALCULATE SYSTEM ENERGY

The system energy is the sum of each individual particle's energy. A key point is that, in the case that particles A and B interact in a pairwise fashion, the energy attributed to particle A is based on the interparticle interaction *at the local density of particle A*.

Similarly, the energy attributed to particle B is based on the interparticle interaction *at the local density of particle B*. This allows particles A and B to interact even if their local densities are widely divergent.

#### Algorithm B-2: Calculate System Energy

```
subroutine calc_system_energy

do i=1, npart
    E(i) = 0
enddo
;E(i) is energy of particle i
;Initially set to zero.

do i=1, npart
    E(i) += mu(density(i))
enddo
;mu is the one-body term of
;the energy, and is a
;function of the local
;density.

do i=1, npart-1
    do j=i+1, npart
        xr = x(i) - x(j)
        if (xr. gt. hbox) then
            xr = xr - box
        else if (xr. lt. -hbox) then
            xr = xr + box
        endif
        if (abs(xr) . lt. rC) then
            E(i) += 0.5*vEFF(xr,density(i))
            E(j) += 0.5*vEFF(xr,density(j))
        endif
    enddo
enddo
;rD is density cutoff radius
;vEFF is the pairwise
;potential between atoms
;i and j.

E_system = sum_i(E(i))
;Total System Energy
```