January 2019

Force Field Optimization, Advanced Sampling, And Free Energy Methods With Gpu-Optimized Monte Carlo (gomc) Software

Mohammad Soroush Barhaghi
Wayne State University

FORCE FIELD OPTIMIZATION, ADVANCED SAMPLING, AND FREE ENERGY METHODS WITH GPU-OPTIMIZED MONTE CARLO (GOMC) SOFTWARE

by

MOHAMMAD SOROUSH BARHAGHI

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2019

MAJOR: CHEMICAL ENGINEERING

Approved by:

Advisor	Date

DEDICATION

To Nasrin, Abdolhosein,

Mehrnaz, and Mehrdad

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

Experiments are the most accurate method for the determination of liquid structure and vapor−liquid equilibrium properties such as, vapor and liquid density, vapor pressure, heat of vaporization, compressibility factor, boiling point, and critical properties. Nonetheless, experimentally measuring these properties for some compounds can be extremely dangerous, expensive, and difficult. For example, hazardous compounds that require high pressure and temperature conditions make these measurements dangerous; thermal decomposition at higher temperatures and the difficulty of separating similar isomers make such measurements expensive and unreliable[1-4]. Molecular simulations provide another route to the prediction of equilibrium properties.

Given a molecular structure, and an accurate description of interactions between atoms, or groups of atoms, computer simulations may be used to calculate nearly any physical property associated with a specified molecule. While molecular simulations are computationally intensive, computer simulations are safe and much less expensive. Using simulations, it is possible to determine relationships between atomic-level interactions, nanoscale structure, and macroscopic properties. Hence, computer simulations are able to provide insight that may otherwise be impossible to obtain from experiments. Notable examples include self-assembly of nanoscale structures[5-8], understanding biological structure-function relationships[9-11], drug design[12-14], and materials design for separation and storage of gases[15-20]. Advances in the use of molecular simulation for materials design have been driven by simultaneous advances in force fields[21-29], algorithms[30-35], computer hardware, and software designed to take advantage of parallel computer architectures, such as multi-core CPUs, and graphics processing units (GPUs)[36-42].

Quantum mechanics (QM) and classical mechanic simulation are common techniques in molecular modeling. Molecular mechanics simulations, using classical force field-based,

may be divided into two categories: molecular dynamics (MD), where the system evolves according to Newton's equations of motion, and Monte Carlo (MC), where the system evolves through sequential trial moves accepted according to probabilities defined by statistical mechanics.

To simulate a equilibrium system (*e.g.* vapor-liquid equilibria, solid-liquid equilibria, and adsorption isotherm), it's often more efficient to use MC simulation, rather than MD simulation, because the MC simulation is capable of performing unrealistic moves, such as large translation, rotation, change in molecular configuration, and molecule transfer, to hop between microstates and satisfy the equilibrium convergence. However, MD algorithm is usually more efficient to simulate a system at high density, low temperature, and strong inter-molecular interaction (*e.g.* hydrogen bonds).

Examples of commonly used MD codes include NAMD[37], AMBER[43], CHARMM[42, 44], GROMACS[41], LAMMPS[38, 45] and HOOMD[40, 46]. Example of commonly used open-source MC codes include Towhee[47, 48], HOOMD[49], Etomica[50], FEASST[51], MS2[52-54], RASPA[55], Cassandra[56], and GOMC[57]. GOMC (GPU-Optimized Monte Carlo) is an object-oriented, general purpose Monte Carlo simulation engine, developed by Dr. Potoff and Dr. Schwiebert research group in department of Chemical Engineering and Computer Science at Wayne State University, respectively. GOMC is capable of performing simulations in canonical, isobaric-isothermal, and grand canonical ensembles, as well as Gibbs ensemble Monte Carlo. GOMC is designed for the simulation of large system, complex molecular topologies, and supports a variety advanced Monte Carlo algorithms and potential functions. GOMC utilize the OpenMP and NVIDIA CUDA to allow for execution on multi-core CPU and GPU architectures.

1.1   Thesis Motivation

1.1.1 Force Field Optimization

Alkynes are unsaturated hydrocarbons with at least one triple carbon-carbon bond and a general formula of C_nH_{2n-2}. Heavier alkynes are commercially synthesized from condensing acetylene (ethyne) via formaldehyde[58]. Acetylene and propylene are commercially synthesized from oxidation of methane, and thermal cracking of hydrocarbons, respectively[58]. Alkynes have various pharmaceutical and industrial applications, such as drugs for cancer treatment[59], torches, rocket fuel, and polyethylene production. Alkynes are used as base stock to synthesize acrylic acid, which is used in the manufacturing of paints, plastics, and adhesives. Alkynes are commonly used as a starting material in synthesis, since the triple bond is easily broken.

The highly reactive nature of alkynes motivates the development of computational methods for the accurate prediction of physical properties and phase behavior, and the application of computer simulation to these compounds requires high quality, transferable, intermolecular potentials. To date, however, a limited number of molecular models have been published for ethyne (acetylene), the smallest alkyne[60-62]. Parameters were optimized via corresponding states theory[60], regression to experimental data[61], and fitting to reproduce *ab initio* derived interaction energies[62]. For longer alkynes, only one force field (SPEAD), which is based on step potentials, has been published[63]. This force field does not have parameters for ethyne and propyne. The reported average absolute deviation in the liquid densities and vapor pressures were 3.0% and 7.0%, respectively, for compounds in the training set.

The lack of transferability in the developed intermolecular potential parameters for alkynes and high error in the reproduction of experimental saturated liquid density, vapor densities, and vapor-pressures, motivated this work to develop an optimized Mie potential

parameters for sp hybridized C, and CH groups and the sp^2 hybridized C group in propadiene to accurately reproduce experimental saturated liquid and vapor densities and vapor pressures. The result of this work has been published in Journal of Molecular Physics[64].

1.1.2 Advanced Monte Carlo Sampling Method

In Monte Carlo simulations in the grand canonical ensemble (GCMC), the chemical potential, volume, and temperature are fixed ($\mu VT = $ constant), while in the Gibbs ensemble Monte Carlo simulation (GEMC), the total number of molecules, volumes, and temperature are fixed ($NVT = $ constant). Sampling of phase space is achieved through a variety of trial moves, such as displacement, molecule insertion and deletion (in GCMC simulation), and molecule transfer (in GEMC simulation). Perhaps the greatest challenge with GCMC and GEMC simulations, however, is achieving a sufficient number of accepted molecular insertion/deletion or molecule transfer moves to ensure adequate sampling of phase space. Therefore, significant effort has been expended to develop algorithms that improve the acceptance percentage for molecule insertions, deletions, and transfer.

Biasing methods, such as rotational, energy, and cavity-bias, has been developed to improve the efficiency of GCMC simulations[65]. The introduction of configurational-bias Monte Carlo enabled the successful simulation of chain molecule adsorption in zeolites[66], which was followed by the coupled-decoupled[67]. These aforementioned biasing methods have greatly extended the complexity of systems that may be simulated with GCMC or GEMC simulations, however, at high densities and low temperatures, the acceptance rate for molecule transfers is still unacceptably low due to the difficulty in finding a favorably sized cavity to insert a molecule. Others have sought to address these issues through the use of cavity-bias[68-70], to identify favorable locations to attempt molecule insertions, or continuous fractional component Monte Carlo[71, 72], and expanded ensembles[73, 74], where molecules are

gradually inserted while the system is allowed to relax locally to minimize steric and energetic penalties due to molecule insertion.

For mixtures, a straightforward approach is to introduce a trial move where the identity of one molecule is changed to that of another[75]. The benefit of such a move is that steric overlaps are reduced significantly, leading to enhanced acceptance for the particle exchange. The methodology has been extended to allow for the exchange of multiple solvent molecules with a polymer chain composed of solvent monomers without changing the coordinates of either polymer or solvent[76]. While a number of publications state that an identity exchange move was used for molecular systems[77-79], a detailed description of the algorithm and the acceptance criteria have not been published to date. The previously described methods for identity exchange were generally applicable to only the special cases for which they were developed, e.g. single particle exchanges[80], a polymer composed of solvent monomers[76], or large hard particles or disks in a solvent of smaller hard particles[81, 82].

These methods are difficult to generalize to molecular systems of arbitrary molecular topology, and their computational performance is expected to be highly correlated with the type of system for which the move was originally developed. To address these issues, a generalized identity exchange move for simulations in the grand canonical and Gibbs ensemble, referred to as Molecular Exchange Monte Carlo (MEMC), is presented that works for systems of any molecular topology. The result of this work has been published in Journal of Chemical Physics[83] and Fluid Phase Equilibria[84].

1.1.3   Effect of Fluorination on the Partitioning of Alcohols

Perfluoroalkyl substances (PFAS) are a broad class of compounds where fluorine has been substituted for hydrogen on the alkyl chains. The most widely used and industrially relevant PFAS are surfactants, where fluorination of the alkyl tails renders them both hydrophobic and oleophobic, giving rise to unusual properties, such as exceptional chemical

and thermal stability and very low interfacial tension at the air-water interface[85-87]. Owing to their unique properties, PFAS are used in a broad array of consumer applications, including coatings for non-stick cookware[88], grease-resistant paper[89], and stain resistant fabrics. Industrial applications include fire-fighting foams[90] and mist-suppressants in hard chrome plating[91]. The strength of the C-F bond, which contributes to the stability of fluorinated surfactants, also makes them extremely resistant to thermal, chemical, or photo degradation; experiments have shown that perfluorinated surfactants are highly resistant to biological degradation[92]. Numerous studies have shown widespread distribution of PFAS in the environment[93, 94]. As a result, PFAS are now considered to be a significant environmental threat[95].

Environmental fate models rely on numerous physical property data, two of the most important of which are the Henry's law constant and the octanol-water partition coefficient, $\log K_{ow}$[96]. Given the breadth of PFAS chemistry and the lack of available experimental data, predictive methods are needed to fill these critical knowledge gaps. Prior work on the partitioning of fluorotelomer alcohols showed that common tools, such as EPISuite[97], CLOGP[98], SPARC[99] and COSMOTherm[100], produce a wide variety of results, with some predictions 2-5 orders of magnitude different than experiment[101]. Alternatively, atomistic computer simulations, combined with free energy methods such as thermodynamic integration[102], free energy perturbation[103, 104], or adaptive biasing force[30, 105], have been used with great success in the prediction of free energies of hydration and solvation in organic solvents for a wide variety of compounds[106-109]. While most work has focused on applications to drug[12, 13, 110] discovery, other calculations have focused on predicting the environmental fate of potentially toxic compounds, such as energetic materials[111, 112], ionic liquids[113], and fluorinated alcohols[114].

While molecular dynamics simulations are widely used for the calculation of free energies of solvation, systems with large energy barriers to configurational and/or conformational change may exhibit biased sampling, leading to incorrect free energies if care is not taken[115]. On the other hand, Monte Carlo simulations allow the system to hop between states and in some cases, may offer conformational sampling advantages over molecular dynamics. Free energies can be determined directly from Gibbs ensemble Monte Carlo simulations from the ratio of number densities of the solute in each phase[77, 84, 116].

Gibbs ensemble Monte Carlo provides a straightforward way of determining free energies of transfer as long as a sufficient number of successful exchanges of the solute between phases occurs, which usually requires the use of advanced configurational-bias sampling methods[83, 84, 116, 117]. For dense liquids with strong electrostatic interactions, obtaining adequately converged results for certain solutes may be challenging, even with state-of-the-art sampling algorithms, the molecular exchange Monte Carlo (MEMC). The fluoro-alcohol systems of interest in this work present a perfect storm of sampling problems: the hydroxyl group has strong electrostatic interactions with the solvent (water or octanol) and it is difficult to find a favorably sized cavity to insert the bulky fluorinated alkyl tail. With enough intermediate states, nearly any molecule exchange between phases is possible[118], but if free energies of transfer are the quantity of interest, it may be more effective to perform standard thermodynamic integration or free energy perturbation. Therefore, this work describes the implementation of thermodynamic integration (TI) and free energy perturbation (FEP) methods into the Monte Carlo simulation engine GOMC[57], and the application of TI and FEP to determine the air-water, air-oil, air-octanol, and octanol-water partition coefficients for eight carbon alcohols with varying degrees of fluorination. The result of this work has been published in Journal of Molecular Physics[119].

1.2   Thesis Organization

The accuracy of calculated physical properties using molecular simulation, depends on accurate description of molecular interaction and topology. In this work, classical force field has been used to describe the inter- and intra-molecular energy for all Monte Carlo simulations. Inter-molecular energy consists of Van Der Waals and coulombic interaction, while the intra-molecular energy includes, bond, angle, dihedral, and nonbonded-intra energy. Hence, Chapter 2 is focused on defining bonded and nonbonded energy functions, in the classical force field.

Van Der Waals interaction can be describe using standard Lennard-Jones potential function or Mie potential function. Mie potential function utilize the repulsion exponent as a variable, which provides additional degree of freedom for optimization. In previous works by our group, it has been shown that Mie potential parameters can be optimized to simultaneously reproduce experimental vapor pressures, liquid, and vapor densities. In Chapter 3, the detailed procedure of optimizing Mie potential parameters for sp hybridized C, and CH groups in alkynes and sp^2 hybridized C group in propadiene, is provided, followed up with prediction of vapor-liquid coexistence curve for propadiene and alkynes from ethyne to nonyne. To understand the impact of dipole and quadrupole moments, additional Mie potential parameters were optimized for ethyne and propyne that included electrostatic potential. The result of these model can be found in Appendix A.

In Chapter 4, the generalized identity exchange move for simulations in the grand canonical ensemble, referred to as Molecular Exchange Monte Carlo (MEMC), is presented. Three different approaches for the insertion of the large molecule are presented and the derivation of acceptance criteria and the algorithms for performing the MEMC move is provided for each of the three approaches. The utility of the three methods and their computational efficiency is illustrated for selected binary mixtures. The detailed computational procedure, mathematical calculations, and additional results are included in the appendix B.

In Chapter 5, MEMC methods are extended to the Gibbs ensemble simulation and the derivation of acceptance criteria and the algorithms for performing the MEMC move in GEMC are provided. The simulation details for determining the binary mixture phase diagrams and Gibbs free energies of transfer are provided, followed up with pressure-composition diagrams for the methane+n-butane and n-butane+perfluorobutane, and free energies of transfer for n-alkanes in 1-octanol, hexadecane, and 2,2,4-trimethylpentane. Additional results with their numerical values, and the schematic of ME-3 algorithm are presented in Appendix C.

Gibbs ensemble Monte Carlo provides a straightforward way of determining free energies of transfer as long as a sufficient number of successful exchanges of the solute between phases occurs. For dense liquids with strong electrostatic interactions, obtaining adequately converged results for certain solutes may be challenging, even with state-of-the-art sampling algorithms, the MEMC. The more effective way to calculate the free energies of transfer for such a system is to perform standard thermodynamic integration or free energy perturbation. Therefore, Chapter 6 describes the implementation of thermodynamic integration (TI) and free energy perturbation (FEP) methods into the Monte Carlo simulation engine GOMC, and the application of TI and FEP to determine the air-water, air-oil, air-octanol, and octanol-water partition coefficients for eight carbon alcohols with varying degrees of fluorination. The bonded parameters used in this chapter and comparison of predicted free energies of hydration/solvation using TI, BAR, and MBAR are presented in Appendix D.

The key findings of the works described in Chapters 3, 4, 5, and 6 are summarized in Conclusions chapter.

CHAPTER 2 FORCE FIELD

The total energy of the system with N molecules can be defined as summation of bonded, nonbonded, and long-range corrections (LRC) energies.

$$U_{total} = U_{nonbonded} + U_{bonded} + U_{LRC} \tag{2.1}$$

2.1   Nonbonded Energies

The nonbonded energies (intra-molecule interaction) includes Van Der Waals and coulombic interaction. Van Der Waals interaction can be described by Lennard-Jones or general version of it, the Mie potential.

$$U_{nonbonded} = \sum_{i=1}^{N} \sum_{j>i}^{N} U_{LJ}(r_{ij}) + U_{Coul}(r_{ij}) \tag{2.2}$$

where r_{ij} is the separation distance for the pair of interaction sites i and j.

The Van Der Waals interaction described by Mie potential is defined as:

$$U_{LJ}(r_{ij}) = C_n \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{n_{ij}} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \tag{2.3}$$

where ε_{ij}, σ_{ij}, and n_{ij} are the well depth, collision diameter, and repulsion exponent, respectively, for the pair of interaction sites i and j. The constant C_n is a normalization factor used such that the minimum of the potential remains at $-\varepsilon_{ij}$ for all n_{ij}.

$$C_n = \left(\frac{n_{ij}}{n_{ij} - 6} \right) \left(\frac{n_{ij}}{6} \right)^{6/(n_{ij}-6)} \tag{2.4}$$

For the 12-6 potential, C_n reduces to the familiar value of 4. Parameters governing interactions between unlike interaction sites were determined using the Lorentz-Berthelot combining rules[120, 121].

$$\sigma_{ij} = (\sigma_{ii} + \sigma_{jj})/2 \tag{2.5}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \tag{2.6}$$

To determine repulsion exponents for cross interactions, an arithmetic average was used.

$$n_{ij} = (n_{ii} + n_{jj})/2 \tag{2.7}$$

The coulombic interaction using the Ewald summation method[122] is defined as:

$$U_{Coul}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_o r_{ij}} erfc(\alpha r_{ij}) \tag{2.8}$$

where ϵ_o, α, q_i, and q_j are the permittivity of vacuum, partitioning parameter, and partial charges for the pair of interaction sites i and j, respectively.

2.2 Bonded Energies

Bonded energies includes bond, angle bending, and dihedral energies:

$$U_{bonded} = \sum_b^{bonds} U_{bond}(b) + \sum_\theta^{angles} U_{angle}(\theta) + \sum_\phi^{dihedrals} U_{dihedral}(\phi) \tag{2.9}$$

The energy associated with bond stretches can be described using harmonic potential, defined as:

$$U_{bond}(b) = k_b(b - b_0)^2 \tag{2.10}$$

where b is the measured bond, b_0 is the equilibrium bond, and k_b is the bond force constant. However, in this work, all simulation were performed using rigid bonds, with no bond energies.

Similar to the bond energies, the angle bending energies can be described using harmonic potential:

$$U_{angle}(\theta) = k_\theta(\theta - \theta_0)^2 \tag{2.11}$$

where θ and θ_0 is the measured and equilibrium angle between three bonded atoms,

respectively, and k_θ is the angle force constant.

The dihedral energies can be represented by cosine series:

$$U_{torsion} = \sum_{n=0} c_n\big(1 + \cos{(n\phi - \delta_n)}\big) \tag{2.12}$$

where ϕ is the dihedral angle, c_n are dihedral force constants, n is the multiplicity, and δ_n is the phase shift.

2.3   Long-range Correction Energies

To accelerate the simulation performance, the nonbonded potential is usually truncated at specific cut-off distance. To compensate the missing potential energy, beyond the cut-off distance (r_{cut}), the long-range correction for Lennard-Jones and coulombic interaction must be calculated and added to the total energy of the system.

$$U_{LRC} = U_{LRC(LJ)} + U_{LRC(Coulomb)} \tag{2.13}$$

For homogeneous system, the long-range correction for Mie potential can be analytically calculated:

$$U_{LRC(LJ)} = \frac{2\pi N^2}{V} \int_{r=r_{cut}}^{\infty} r^2\, U_{LJ}(r)dr$$

$$U_{LRC(LJ)} = \frac{2\pi N^2}{(n-3)V} C_n \varepsilon\, \sigma^3 \left[\left(\frac{\sigma_{ij}}{r_{cut}}\right)^{(n-3)} - \left(\frac{n-3}{3}\right)\left(\frac{\sigma_{ij}}{r_{cut}}\right)^3 \right] \tag{2.14}$$

where N and V are the number of molecule and volume of the system, respectively.

The long-range correction for Ewald summation method[122] is defined as:

$$U_{LRC(Coul)} = U_{self} + U_{correction} + U_{reciprocal} \tag{2.15}$$

$$U_{self} = -\frac{\alpha}{4\sqrt{\pi^3}\epsilon_o} \sum_i^N q_i^2 \tag{2.16}$$

$$U_{correction} = -\frac{1}{4\pi\epsilon_o}\sum_i^N\sum_a\sum_{b>a} q_{ia}q_{ib}\frac{erf(\alpha r_{ia,ib})}{r_{ia,ib}} \tag{2.17}$$

$$U_{reciprocal} = \frac{1}{2V\epsilon_o}\sum_{\vec{K}\neq 0}\frac{1}{|\vec{K}|^2}e^{-\frac{|\vec{K}|^2}{4\alpha^2}}\left[\left|\sum_i^N q_i\ \cos(\vec{K}.\vec{r_i})\right|^2 + \left|\sum_i^N q_i\ \sin(\vec{K}.\vec{r_i})\right|^2\right] \tag{2.18}$$

where $r_{ia,ib}$, q_{ib}, and q_{ib} are the separation distance, and partial charges for atom a and b in

molecule i, respectively, and $\vec{K}$ is the wave-vectors.

CHAPTER 3 OPTIMIZED MIE POTENTIALS FOR PHASE EQUILIBRIA: APPLICATION TO ALKYNES

3.1   Introduction

Alkynes are unsaturated hydrocarbons with at least one triple carbon-carbon bond and a general formula of C_nH_{2n-2}. Alkynes can be found in nature in the form of plants, fungi, bacteria, marine sponges, and corals[123], however, commercially heavier alkynes are synthesized from condensing acetylene (ethyne) via formaldehyde[58]. Acetylene and propyne are commercially synthesized from oxidation of methane, and thermal cracking of hydrocarbons, respectively[58]. Alkynes have various pharmaceutical and industrial applications, such as drugs for cancer treatment[59], torches, rocket fuel, and polyethylene production. Alkynes are used as base stock to synthesize acrylic acid, which is used in the manufacture of paints, plastics, and adhesives. Alkynes are commonly used as a starting material in synthesis, since the triple bond can be broken easily. Pure alkynes are very unstable and reactive compared to alkenes and alkanes. Alkynes are usually mixed with other compounds to form a solution that limits material degradation. For example, propyne is mixed with propadiene to form methylacetylene-propadiene propane (MAPP) gas.

The highly reactive nature of alkynes motivates the development of computational methods for the accurate prediction of physical properties and phase behavior, and the application of computer simulation to these compounds requires high quality, transferable, intermolecular potentials. To date, however, a limited number of molecular models have been published for ethyne (acetylene), the smallest alkyne[60-62]. These models are based on a variety of potential functions, such as Lennard-Jones + point quadrupole[60], two-center Lennard-Jones plus point quadrupole[61], and Morse-C6 potentials combined with point charges[62]. Parameters were optimized via corresponding states theory[60], regression to experimental data[61], and by fitting to reproduce *ab initio* derived interaction energies[62].

While a rigorous error analysis was not performed on any of the force fields, the model of Vrabec *et al.* appears to give the best reproduction of the experimental vapor-liquid coexistence curve to date[61]. For longer alkynes, only one force field (SPEAD), which is based on step potentials, has been published[63]. This force field does not have parameters for ethyne and propyne. The reported average absolute deviation in the liquid densities and vapor pressures were 3.0% and 7.0%, respectively, for compounds in the training set.

The results from the SPEAD model illustrate a typical problem in the development of intermolecular potentials, which is achieving high accuracy in both the reproduction of experimental saturated liquid densities and vapor-pressures. For united-atom force fields, the limitations of the standard Lennard-Jones potential function are now well known, even when applied to non-polar systems, such as n-alkanes[24]. Our group has shown that through the use of Mie potentials, it is possible to produce significant improvements in the simultaneous reproduction of saturated liquid densities and vapor pressures for n-alkanes[24], branched alkanes[124], alkenes[125], perfluorocarbons[24], and noble gases[126, 127]. In addition, Mie potentials shows significant improvement in predicting the viscosities for saturated and compressed liquids[128, 129].

Nonbonded potential parameters can be optimized using brute-force algorithm in grand canonical simulations, combined with histogram-reweighting methods[130, 131] to calculate vapor-liquid equilibria properties for each nonbonded parameters. In addition, nonbonded potential parameters can be optimized using the post-simulation analysis technique in canonical or grand canonical ensemble simulations, combined with the isothermal-isochoric integration (ITIC)[132] or histogram-reweighting methods, respectively, to predict the vapor-liquid equilibria properties, without performing additional simulations[133, 134].

In this Chapter, the optimized Mie parameters are developed for alkynes as well as propadiene, using brute-force algorithm in grand canonical simulations. Parameters are

introduced for sp hybridized C, and CH groups and the sp^2 hybridized C group in propadiene. These parameters are optimized to reproduce experimental saturated liquid densities and vapor pressures. Vapor-liquid coexistence curves are predicted for propadiene, and alkynes from ethyne to nonyne using grand canonical histogram-reweighting Monte Carlo simulations. The transferability of the models, and the reliability of available experimental data, are assessed. Radial distribution functions are used to provide insight on the effect of the sp hybridized carbon group on liquid phase structure. The result of this work has been published in Journal of Molecular Physics[64].

3.2   Force Field Parameters

In this work, the Mie potential parameters for unlike interaction sites were determined using, Eq. 2.5-2.7. The use of an arithmetic average of repulsion exponents for the combing rule is consistent with past optimization efforts for n-alkanes, perfluoroalkanes[24], alkenes[125] and noble gases[126] using Mie potentials. This is supported by recent work by Stiegler and Sadus, who examined the effect of combining rules on the physical properties predicted by non-identical potentials[135], and comparisons of interaction energies predicted by Mie potentials for noble gases to MP2/aug-cc-PVTZ ab initio calculations[126].

All non-bonded parameters used in this work are listed in Table 3.1. Parameters for CH$_3$ and CH$_2$ (sp^3), CH$_2$ (sp^2), and CH (sp^2) pseudo-atoms were taken from our previous work on n-alkanes[24] and n-alkenes[125] and used without modification. During the optimization process, it was recognized that it was impossible to develop a single set of transferable parameters for the C(sp) group. Therefore, unique C(sp) parameters were optimized for 1-alkynes and 2-alkynes. As shown in Table 3.1, the optimal ε for C in 1-alkynes is over 40% larger than in 2-alkynes. A single set of parameters for the CH(sp) group was optimized for use in all alkyne compounds.

Table 3.1: Non-bonded parameters for alkanes, alkenes and alkynes.

Pseudo-atom	$\varepsilon_i/k_b(K)$	σ_i (Å)	n_i
CH$_{3[24]}$	121.25	3.783	16
CH$_2$ (sp^3)[24]	61.00	3.990	16
CH$_2$ (sp^2)[125]	104.20	3.705	16
CH (sp^2)[125]	60.00	3.810	16
C (sp^2)	189.00	2.950	16
CH(sp)	148.50	3.570	28
C(sp) (1-alkyne)	206.00	2.875	16
C(sp) (2-alkyne)	118.00	3.120	16

Fixed bond lengths were used to connect pseudo-atoms, and are listed in Table 3.2. Equilibrium bond lengths were taken from previous work for alkanes[24] and alkenes[125] , except for bonds that included CH(sp) or C(sp) groups. Equilibrium bond lengths for CH≡CH, C≡CH, and CH$_x$−C(sp) were determined from geometry optimizations performed with MP2/aug-cc-PVTZ *ab initio* calculations using Gaussian 09[136]. Bond lengths predicted from *ab initio* calculations were found to be in close agreement with the microwave spectrum results[137], as well as prior *ab initio* calculations[138] and empirical force fields[62]. Equilibrium bond angles and force constants are listed in Table 3.2.

Table 3.2: Bonded parameters for alkynes.

Bond type	Bond length (Å)	Angle type	θ_0(degrees)	$k_\theta/k_b(\text{K} \cdot \text{rad}^{-2})$
CH$_2$−CH$_3$	1.54	CH$_3$−CH$_2$−CH$_2$	114	31250
CH$_2$−CH$_2$	1.54	CH$_2$−CH$_2$−CH$_2$	114	31250
C−CH$_3$	1.46	CH$_3$−CH$_2$−C	112	31250
C−CH$_2$	1.46	CH$_2$−CH$_2$−C	112	31250
CH≡CH	1.21	CH$_x$−C≡CH	180	30800
C≡CH	1.21	CH$_x$−C≡C	180	30800
C=CH$_2$	1.33	CH$_2$=C=CH$_2$	180	36200

The dihedral constants are listed in Table 3.3. Fourier constants for the CH_x—(CH_2)—(CH_2)—CH_2 dihedral were taken from the OPLS-UA force field[139, 140]. Missing Fourier constants were optimized to reproduce *ab initio* rotational barriers calculated from relaxed potential energy scans of the dihedral of interest. *Ab initio* calculations were performed with MP2 theory and the 6-31+G(d,p) basis set in Gaussian 09[136]. Rotational barriers around bonds next to a triple bond were approximately 50 K, therefore, Fourier constants for these dihedrals were set to zero.

Table 3.3: Fourier constants for alkyne dihedral potentials.

torsion	n	c_n/k_b (K)	δ_n
CH_x—(CH_2)—(CH_2)—CH_2	1	335.03	0
	2	-68.19	π
	3	791.32	0
CH_x—(CH_2)—(CH_2)—C	0	94.88	0
	1	162.00	0
	2	-205.40	π
	3	980.40	0
CH_x—(CH_2)—$(C) \equiv CH$	0	0.00	0
CH_x—(CH_2)—$(C) \equiv C$	0	0.00	0
CH_x—$(C) \equiv (C)$—CH_x	0	0.00	0

3.3   Simulation Methodology

Vapor-liquid coexistence curves, vapor pressures, and heats of vaporization were determined from histogram-reweighting Monte Carlo simulations in the grand canonical ensemble[130, 131, 141]. Simulations were performed with the development version of GPU Optimized Monte Carlo (GOMC)[57]. GOMC is an object-oriented Monte Carlo simulation engine, capable of performing simulations in canonical, isobaric-isothermal, grand canonical ensembles, as well as Gibbs ensemble Monte Carlo. GOMC is designed for the simulation of complex molecular topologies, and supports a variety of potential functions, such as Lennard-Jones and Mie potentials. Coulomb interactions are also supported via the Ewald summation

method[122]. GOMC is capable of parallel computation, either on multicore CPUs or GPUs. Using the development version of GOMC, simulations of 1-pentyne ($n_{ch,LJ(1st)}$=8, $n_{ch,LJ}$=4, $n_{ch,tor}$=10, $n_{ch,bend}$=100) for the vapor phase, near critical point bridge and liquid phase required 41, 137 and 246 seconds, respectively, per 1 million Monte Carlo steps on a single core of an Intel i5 3.30 GHz CPU.

For propadiene and short alkynes (acetylene and propyne), a cubic cell size of 25 Å was used. For butyne and pentyne, a cell size of 30 Å x 30 Å x 30 Å was used. A cell size of 35 Å x 35 Å x 35 Å was used for longer alkynes. Initial configurations were generated with Packmol[142]. Psfgen was used to generate coordinate (*.pdb) and connectivity (*.psf) files[143]. Potentials were truncated at 10 Å and analytical tail corrections were applied[144]. A move ratio of 30% displacements, 10% rotations, and 60% molecule transfers was used. The coupled-decoupled configurational-bias Monte Carlo (CBMC) algorithm was used to improve sampling efficiency during the simulation[67]. For all simulations, CBMC parameters were: 100 angle trials, 10 dihedral trials, 8 trial locations for the first site, and 4 trial locations for secondary sites. For simulations near the normal boiling point, the number of trial locations was increased to 12 and 10 for the initial and secondary sites, respectively. Acceptance rates for molecule insertions in liquid phase simulations were between 0.1% and 12%, depending on molecule type, chemical potential, and temperature.

To generate the phase diagrams predicted by each parameter set, 9 to 10 simulations were performed; one simulation to bridge the gas and liquid phases near the critical temperature, two in the gas phase, and 6 to 7 liquid simulations. For all compounds, 5×10^6 Monte Carlo steps (MCS) were used for equilibration, followed by a data production period of 2.5×10^7 steps, except for simulations near boiling points, where the data production period was increased to 4.5×10^7 steps. Histogram data were collected as samples of the number of molecules in the simulation cell and the non-bonded energy of the system. Samples were taken

on an interval of 200 MCS. Averages and statistical uncertainties were determined from five independent sets of simulations, where each simulation was started with a different random number seed.

In grand canonical Monte Carlo simulations, the pressure is related to the partition function through equation 3.1.

$$P = \frac{k_B T}{V} \ln \Xi \left(\mu, V, T\right) + C \tag{3.1}$$

The pressure is determined through integration of the area under the probability distributions extracted from the GCMC simulations. The additive constant C is determined by extrapolating the predicted partition function to very low densities where the system exhibits ideal gas behavior. In the ideal gas regime, the plot of $\ln \Xi(\mu, V, T)$ vs. particle number is linear with a unit slope. The additive constant is determined from the y-intercept[141]. For Gibbs ensemble Monte Carlo simulations, pressures were calculated from the virial expression.

The heat of vaporization (ΔH_V) was calculated from the energies and molar volumes in each phase[145]

$$\Delta H_V = (U_V - U_L) + P(V_V - V_L) \tag{3.2}$$

where P is the saturation pressure, U_V, U_L, V_V and V_L are the energy per mole and molar volumes of the vapor and liquid phases, respectively.

The critical temperature T_c and density ρ_c for each model were calculated by fitting the saturated liquid and vapor densities to the law of rectilinear diameters[146]

$$\frac{\rho_{liq} - \rho_{vap}}{2} = \rho_C + A(T - T_C) \tag{3.3}$$

and to the density scaling law for the critical temperatures[147]

$$\rho_{liq} - \rho_{vap} = B(T - T_C)^\beta \tag{3.4}$$

where $\beta = 0.325$ is critical exponent for Ising-like fluids in three dimensions[148]. *A* and *B* were constants fit to the saturated vapor and liquid densities. A two-step process was used to determine the critical temperatures. An estimate of the critical temperature predicted by the model was produced by fitting Eq. 3.3 and Eq. 3.4 to phase coexistence data on the range $0.7T_{c,\text{expt}}$ to $0.9T_{c,\text{expt}}$. The estimated critical point was used to set the range of temperatures that were used to determine the true critical point predicted by each model ($0.7T_{c,\text{est}} \leq T \leq T_{c,\text{est}} - 25$ K). This approach ensured that the temperature range used in the calculation was close enough to the true critical point to provide an accurate application of equations 3.3 and 3.4, while avoiding finite size effects near the critical point.

Critical pressures P_c and boiling points T_{NBP} were calculated by fitting vapor pressure data to the Clausius-Clapeyron equation

$$\ln P = -\frac{\Delta H_V}{RT} + C \tag{3.5}$$

where P is the vapor pressure, ΔH_V is the heat of vaporization, R is the gas constant, and C is a constant.

Pressure composition diagrams for binary mixtures of propyne with propane, propene, and propadiene were determined from NVT Gibbs ensemble Monte Carlo simulations[32]. Calculations were performed on systems containing 2000 molecules. Simulations were equilibrated for 1×10^8 MCS and production data were taken from a second 1×10^8 MCS simulation. Statistical uncertainties were determined from three independent sets of simulations, where each simulation was initiated with a different random number seed. The distribution of Monte Carlo moves was 69% displacement, 10% rigid body rotations, 1% volume exchange, and 20% molecule exchange.

Radial distribution functions were determined from isobaric-isothermal Monte Carlo simulations, which were performed on cubic boxes containing 1000 molecules. Simulations

were performed at the predicted normal boiling point and 1 bar. All simulations were equilibrated for 1×10^8 MCS and production data were taken from a second 1×10^8 MCS simulation. Radial distribution functions were calculated using VMD[143].

3.4   Results and Discussion

3.4.1   Parameter Optimization

Experimental vapor-liquid equilibrium (VLE) data for alkynes is very limited. Experimental saturated liquid and vapor densities and vapor pressures are available for ethyne[149] and propyne[150, 151]. These data were used in the optimization process. For 1-alkynes longer than propyne, and all 2-alkynes, vapor-liquid coexistence data are limited to temperatures at or below the normal boiling point[151-153]. Therefore, VLE predictions from simulation for longer alkynes were compared to correlations from the DIPPR database[154]. The DIPPR correlations were reported to have errors of < 1% compared to experimental saturated liquid data. Critical points for 1-alkynes have been determined experimentally for ethyne, propyne and 1-butyne[155], and these were used for comparison to simulation. Recommended experimental data for the critical temperatures of many alkynes were given by Owczarek and Blazej, however, the authors report that no information was available regarding the experiments used to determine T_c[156]. To maintain consistency with the saturated liquid density data, critical parameters for alkynes longer than 1-butyne were taken from the DIPPR database, which uses a modified version of the Ambrose method for their prediction[154]. The resulting critical temperatures listed in the DIPPR database have suggested maximum errors of 3-5%.

Force field parameters were first optimized for the CH group in ethyne, followed by the C group in propyne and 1-butyne. Preliminary calculations revealed that transferability of C group parameters from 1-alkynes to 2-alkynes was poor. Therefore, unique parameters were

optimized for C in 2-alkynes using simulations of 2-butyne. The same CH parameters were used for all alkynes. In addition, parameters for the C (sp^2) group found in alkenes were optimized from calculations performed on propadiene. Alkynes have a small dipole moment of approximately 0.7D, while ethyne has a quadrupole moment of 20.4x10^{-40} C-m^2[157]. To understand the impact of neglecting the small dipole and quadrupole moments on the predictive capability of the force field, additional parameters were optimized for ethyne and propyne models that included point charges. The results of these calculations are provided in the Appendix A, Figures A4-A6, and support the decision to neglect modeling multipole moments explicitly with partial charges.

A scoring function was used to evaluate the performance of specific parameters for each optimized compound in the alkyne series.

$$S_i = \frac{1}{n}\left[0.757 \sum_{i=0}^{n} Err\big(\rho_L(T_i)\big) + 0.152 \sum_{i=0}^{n} Err\big(P_V(T_i)\big) \right.$$

$$\left. + 0.076 \sum_{i=0}^{n} \frac{d\left(Err\big(\rho_L(T_i)\big)\right)}{dT} + 0.015 \sum_{i=0}^{n} \frac{d\left(Err\big(P_V(T_i)\big)\right)}{dT} \right]$$

(3.6)

where

$$Err(\langle X(T_i)\rangle) = \left| \frac{\langle X_{SIM}(T_i)\rangle - \langle X_{EXPT}(T_i)\rangle}{\langle X_{EXPT}(T_i)\rangle} \right| * 100$$

(3.7)

One key difference between the objective function used in this work and those used in other efforts is the use of unequal weights for various physical properties. By weighting the liquid density more heavily than other factors, it is possible to reliably evaluate the effect of different potential functions, or repulsion exponents, on the vapor pressure. This is especially true in cases where it is not possible to achieve accuracies to within 1% of experimental data for all properties of interest. In these cases, large errors in certain properties, such as vapor

pressures or heats of vaporization, can bias the results so that one is simply trading lower error in vapor pressure for greater error in liquid densities. The larger weight for the liquid density limits the maximum error to approximately 1%.

Preliminary calculations were used to identify optimal repulsive exponents and regions of parameter space on which an exhaustive grid-based search should be applied. For the CH(sp) group, simulations of ethyne were used to identify n_i=28 as the optimal repulsion exponent. The larger repulsion exponent increases the range of interaction, and compensates for the lack of an explicit hydrogen or electrostatic interactions. Simulations of 1-propyne and 1-butyne identified n_i=16 as the optimal repulsion exponent for the C(sp) group. With the region of search space identified and the repulsive exponent selected, an exhaustive grid-based search of the surrounding parameter space (σ_i, ε_i) was performed. For ethyne, 63 parameter sets were evaluated, spaced on 0.005 Å increments along σ_i on the range $3.550 \leq \sigma_i \leq 3.580$ Å, and spaced on 0.5 K increments along ε_i on the range $146.0 \leq \varepsilon_i \leq 150.0$ K. For propyne and 1-butyne, 105 parameter sets were evaluated over the same range of $2.825 \leq \sigma_i \leq 2.925$ Å and $190 \leq \varepsilon_i \leq 230$ K. The parameter space for these compounds was explored in increments of 0.025 Å for σ_i and 2.0 K for ε_i. For 2-butyne, 121 parameter sets were evaluated, spaced on 0.02 Å increments along σ_i on the range $3.00 \leq \sigma_i \leq 3.20$ Å, and spaced on 2.0 K increments along ε_i on the range $110 \leq \varepsilon_i \leq 130$ K. Phase diagrams were produced for each parameter set and used to calculate an average error score, S_i.

Normalized values of the scoring function for ethyne, 1-propyne, 1-butyne, 2-butyne, and propadiene are presented as heat maps in Figure 3.1. Examination of the heat maps for the individual compounds shows some interesting trends. For ethyne, a well-defined optimal solution was found, which is similar to prior work on noble gases[126]. For all other compounds, however, a broad band of optimal values was observed. This arises as a result of the unique linear geometry around the triple bond, which limits the impact of changing σ_i on

saturated liquid densities, and the slope of the saturated liquid curve. Creating a single optimal region (a well) on the heat map would require treating the C≡CH bond length as an optimizable parameter. Comparison of heat maps for propyne and 1-butyne shows overlap between optimal regions of parameter space, however, the optimal band of σ_i and ε_i values for propyne were shifted to slightly lower values of ε_i. The combined normalized scores for propyne and 1-butyne were used to identify the final optimized parameters for the C (sp) group in 1-alkynes. More weight was placed on the propyne results because of the greater availability of experimental data[150] compared to 1-butyne where most of the data were generated via correlation[154]. Optimal regions on the heat maps for 1-butyne and 2-butyne differ by approximately 90 K in ε_i and 0.2 Å in σ_i, showing clearly that C(sp) parameters are not transferable between 1-alkynes and 2-alkynes. Significant differences were observed in the electrostatic potential energy surfaces determined from HF/6-31+g(d,p) *ab initio* calculations for 1-alkanes and 2-alkynes (Appendix A, Figure A1), further confirming the need for unique Mie parameters for each class of molecules.
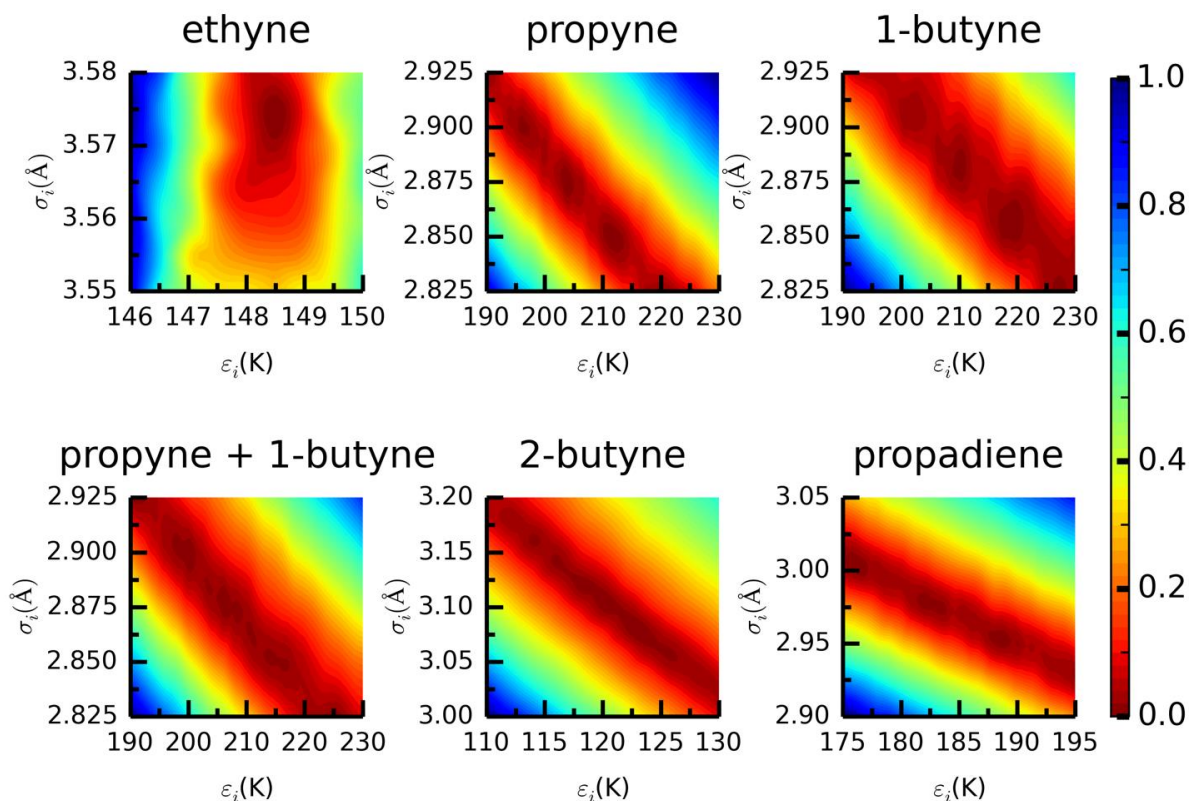
Figure 3.1: Heat map of average error scores for alkyne compounds. Red depicts the best fit to experimental data or correlation[34, 36, 39], blue depicts the worst fit.

3.4.2 Pure Fluid Vapor-Liquid Equilibria

Vapor-liquid coexistence curves for ethyne to 1-nonyne, 2-butyne, 2-pentyne, 2-hexyne, and propadiene are presented in Figure 3.2. Predictions of the two center Lennard-Jones plus quadrupole (2CLJQ) model for ethyne, propyne, and propadiene are included for comparison[61]. In Table 3.4, the average absolute deviations (AAD) from experiment and correlations are presented for predictions of the optimized Mie potentials and the SPEAD model[63]. The AAD of simulation from experiment for saturated liquid densities shows an unusual peak in the homologous series for 1-hexyne (3.0% AAD). The non-monotonic behavior of the AAD observed for saturated liquid densities suggests possible inconsistencies in the DIPPR correlations for 1-alkynes. Based on past work on *n*-alkanes, it was expected that deviations of simulation from experiment should increase monotonically with increasing number of carbon atoms[24]. The monotonic increase in error for *n*-alkanes results from the

accumulation of small errors due to the inability of the force field to account for non-additive effects, and similar behavior was expected, but not observed, for simulations of 1-alkynes.
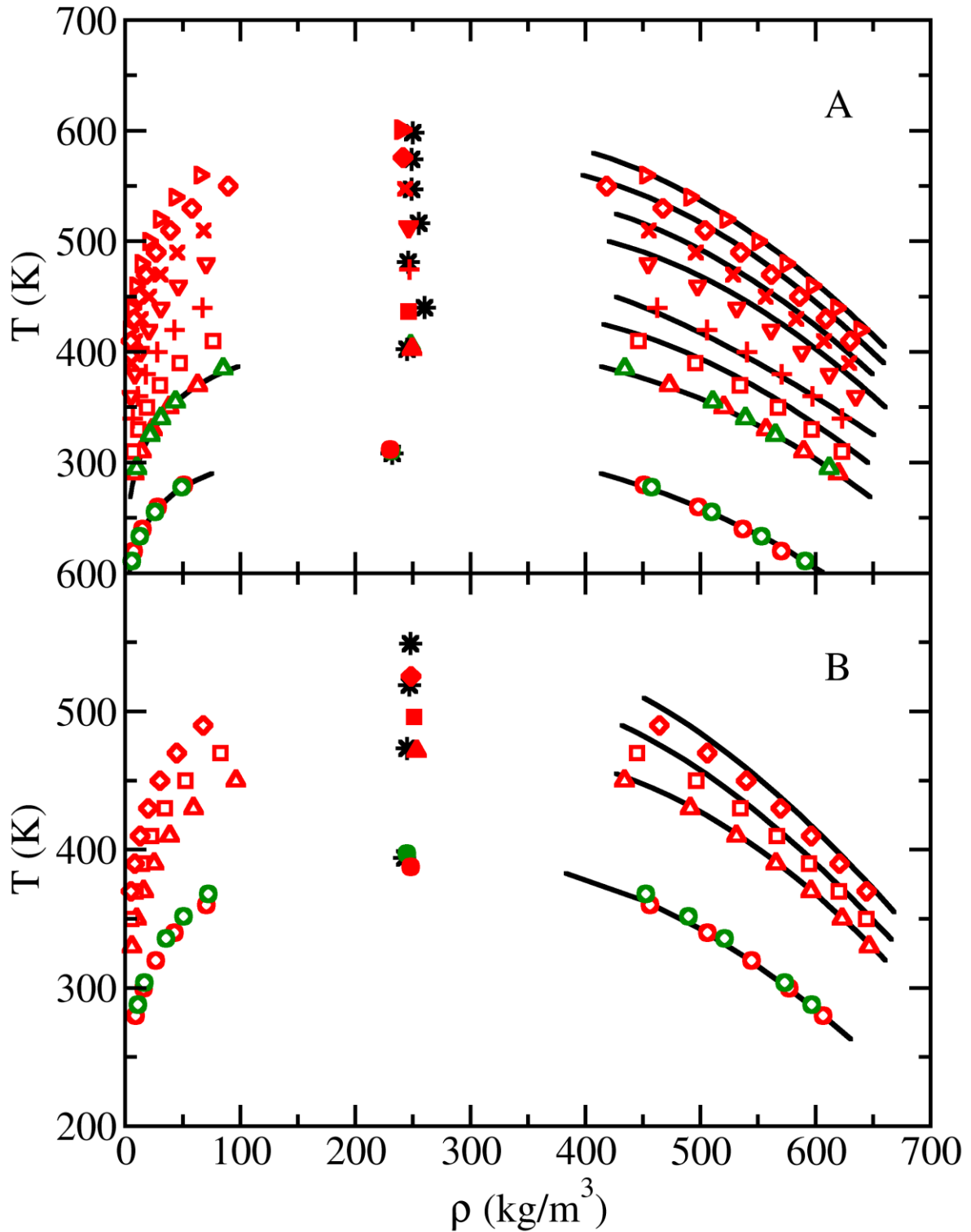
Figure 3.2: Vapor-liquid coexistence curves predicted by the optimized Mie potentials (red symbols) and 2CLJQ model (green symbols)[61] compared to experiment or correlation (solid line)[149, 151, 154] for alkynes and propadiene. Experimental critical points (black stars)[154, 155] and predictions of simulation (filled symbols). Figure 3.2A: ethyne (circles); propyne (triangles up); 1-butyne (squares); 1-pentyne (plus); 1-hexyne (triangles down); 1-heptyne (crosses); 1-octyne (diamonds); 1-nonyne (triangles right); Figure 3.2B: propadiene (circles); 2-butyne (triangles up); 2-pentyne (squares); 2-hexyne (diamonds).

The optimized Mie potentials produce significant improvements in the prediction of liquid densities, compared to SPEAD for 2-alkynes, where the Mie potentials are 3-4 times more accurate. Similar results were achieved by both models for the 1-alkynes, with differences likely due to the choice of compounds used in the optimization process. For example, the Mie potentials predicted saturated liquid densities with an AAD for 1-pentyne of 1.25%, compared to 3.98% for SPEAD. On the other hand, for 1-hexyne, the Mie potentials have an AAD of 3.01%, compared to 0.52% for SPEAD.

Claussius-Clayperon plots are presented in Figure 3.3. Vapor pressures for all 1-alkynes are predicted with an AAD of 3.4% or less, except 1-pentyne, which had an AAD of 7.57%. Vapor pressures for 2-butyne were in close agreement with experiment, while 2-pentyne and 2-hexyne show significant errors, with 21.5 and 23.8% AAD, respectively. Comparing these results to the predictions of the SPEAD model highlights some of the tradeoffs of using an objective function with equal weights vs. the one used in this work. While errors in the vapor pressures for 2-pentyne are reduced from 21.5% (Mie) to 5% (SPEAD), SPEAD trades improved vapor pressures for significantly worse reproduction of saturated liquid densities[63]. For example, for 2-pentyne the AAD in the liquid densities was 2.15% (Mie) vs. 9.99% (SPEAD).
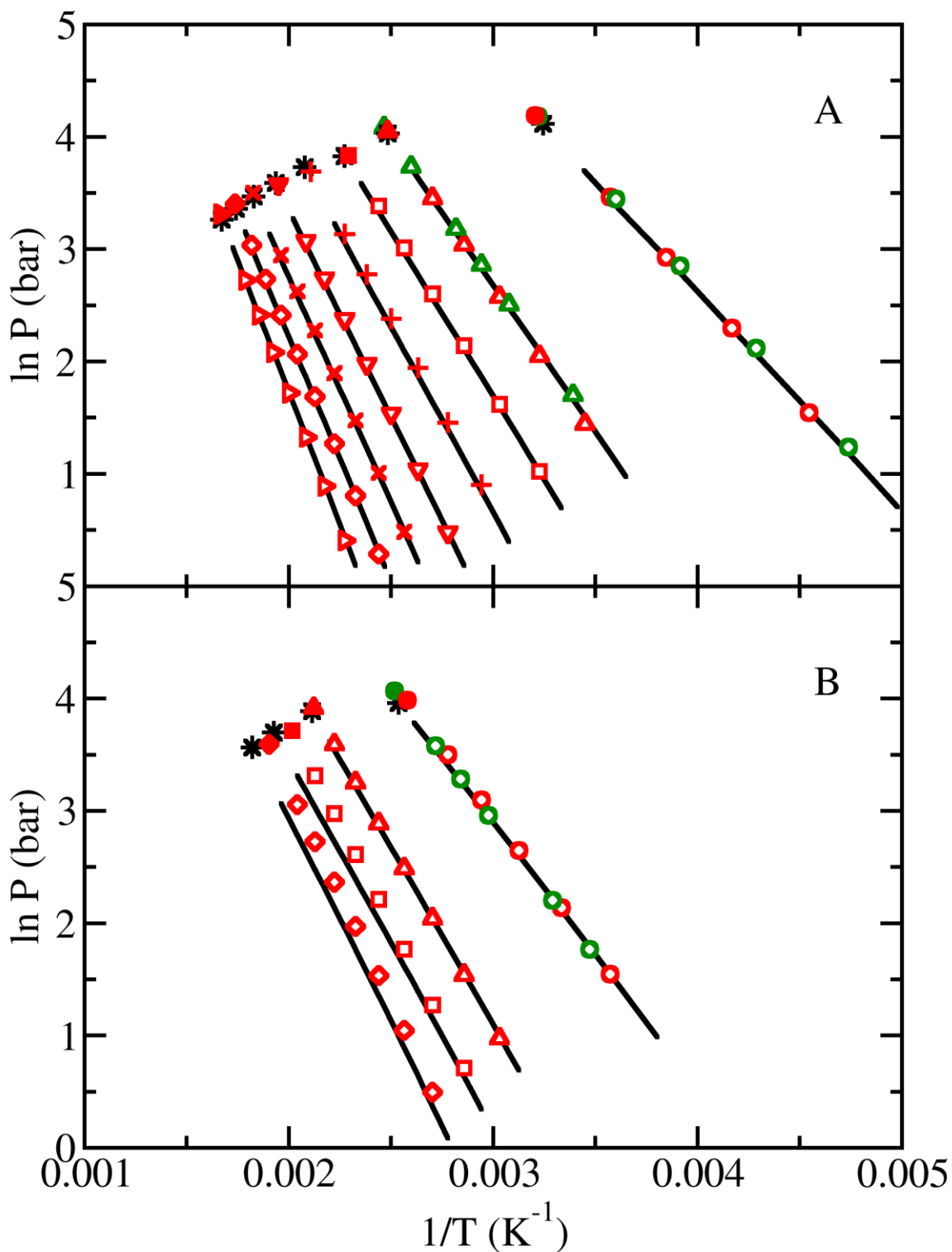
Figure 3.3: Clausius-Clapeyron plots predicted by the optimized Mie potentials (red symbols) and 2CLJQ model (green symbols)[61] compared to experiment or correlation (solid line)[149, 151, 154] for alkynes and propadiene. Experimental critical points (black stars)[154, 155] and predictions of simulation (filled symbols). Figure 3.3A: ethyne (circles); propyne (triangles up); 1-butyne (squares); 1-pentyne (plus); 1-hexyne (triangles down); 1-heptyne (crosses); 1-octyne (diamonds); 1-nonyne (triangles right); Figure 3.3B: propadiene (circles); 2-butyne (triangles up); 2-pentyne (squares); 2-hexyne (diamonds).

Critical temperatures T_c, pressures P_c, densities ρ_c, and normal boiling points T_{NBP} predicted by the optimized Mie potentials are listed in Table 3.5. For ethyne, propyne, and 1-butyne, experimental data were used for comparison[155, 158], while predictions of group contribution methods listed in the DIPPR database were used for all other compounds[154]. For all compounds, normal boiling points were predicted to within 1.1% of experiment[158]. Critical temperatures were reproduced to within 1.4% of reported values for all 1-alkynes. For 2-alkynes, the critical temperature of 2-butyne was with 0.4% of the DIPPR value, but T_c for 2-pentyne and 2-hexyne were under-predicted by about 4.4%. Critical temperatures and densities predicted by the Mie potentials for 1-alkynes, 1-alkenes, and n-alkanes as a function of number of carbon atoms are shown in Figure 3.4. As expected, the critical temperatures predicted by simulation for all three classes of molecules increase monotonically and converge for C6 and larger molecules. Simulations also correctly reproduced the ordering of critical temperatures for propyne, propene, and propane.

Table 3.4: Average absolute deviation (%AAD) from experiment or correlations[34, 36, 39] for vapor pressure and saturated liquid density predicted by optimized Mie and SPEAD models[7].

Compound	P		ρ_L	
	Mie	SPEAD	Mie	SPEAD
ethyne	1.20	---	0.77	---
propyne	2.60	---	0.31	---
1-butyne	2.12	4.74	2.05	1.71
2-butyne	1.52	5.23	0.36	13.45
1-pentyne	7.57	9.57	1.25	3.98
2-pentyne	21.48	5.02	2.15	9.99
1-hexyne	3.39	6.62	3.01	0.52
2-hexyne	23.78	1.87	2.64	7.07
1-heptyne	3.08	---	1.75	---
1-octyne	2.04	---	1.59	---
1-nonyne	1.34	5.00	0.96	0.87
propadiene	6.17	---	0.40	---

Table 3.5: Critical parameters and normal boiling points predicted by the optimized Mie potentials for propadiene and alkynes. Numbers in parenthesis correspond to the uncertainty in the last digit.

Compound	T_c (K)		ρ_c (g/cm^3)		P_c (bar)		T_{NBP} (K)	
	Sim.	Lit.[154, 155]	Sim.	Lit[154, 155]	Sim.	Lit[154, 155]	Sim.	Lit[158]
ethyne	312.1(1)	308.3(1)	0.2301(3)	232(4)	66.04(7)	61.4(1)	187.78(3)	189.0(6)
propyne	402.74(8)	402.4(2)	0.2491(2)	245(5)	57.28(6)	56.3(2)	251.22(4)	250.0(5)
1-butyne	436.6(1)	440(2)	0.2460(2)	260(30)	46.35(9)	46(2)	280.43(8)	283(7)
2-butyne	471.5(2)	473.2	0.2530(4)	245	50.16(5)	48.7	300.26(5)	300(1)
1-pentyne	474.6(2)	481.2	0.2468(3)	246	40.13(7)	41.7	311.52(6)	313.0(7)
2-pentyne	496.1(1)	519	0.2508(3)	247	41.06(7)	40.3	327.14(7)	329.0(9)
1-hexyne	513.1(1)	516.2	0.2461(2)	255	35.9(1)	36.2	343.9(1)	344.3(9)
2-hexyne	525.5(1)	549	0.2483(2)	248	36.35(9)	35.3	353.1(1)	357.1(8)
1-heptyne	547.60(8)	547	0.2432(1)	249	33.22(9)	32.1	372.7(1)	372(1)
1-octyne	575.8(1)	574	0.2412(2)	249	30.1(1)	28.8	399.3(1)	399(2)
1-nonyne	601.1(1)	598.05	0.2395(2)	250	27.52(9)	26.1	424.3(1)	423(1)
propadiene	387.59(9)	394	0.2478(2)	0.2428	53.95(4)	52.5	238.29(3)	240(2)
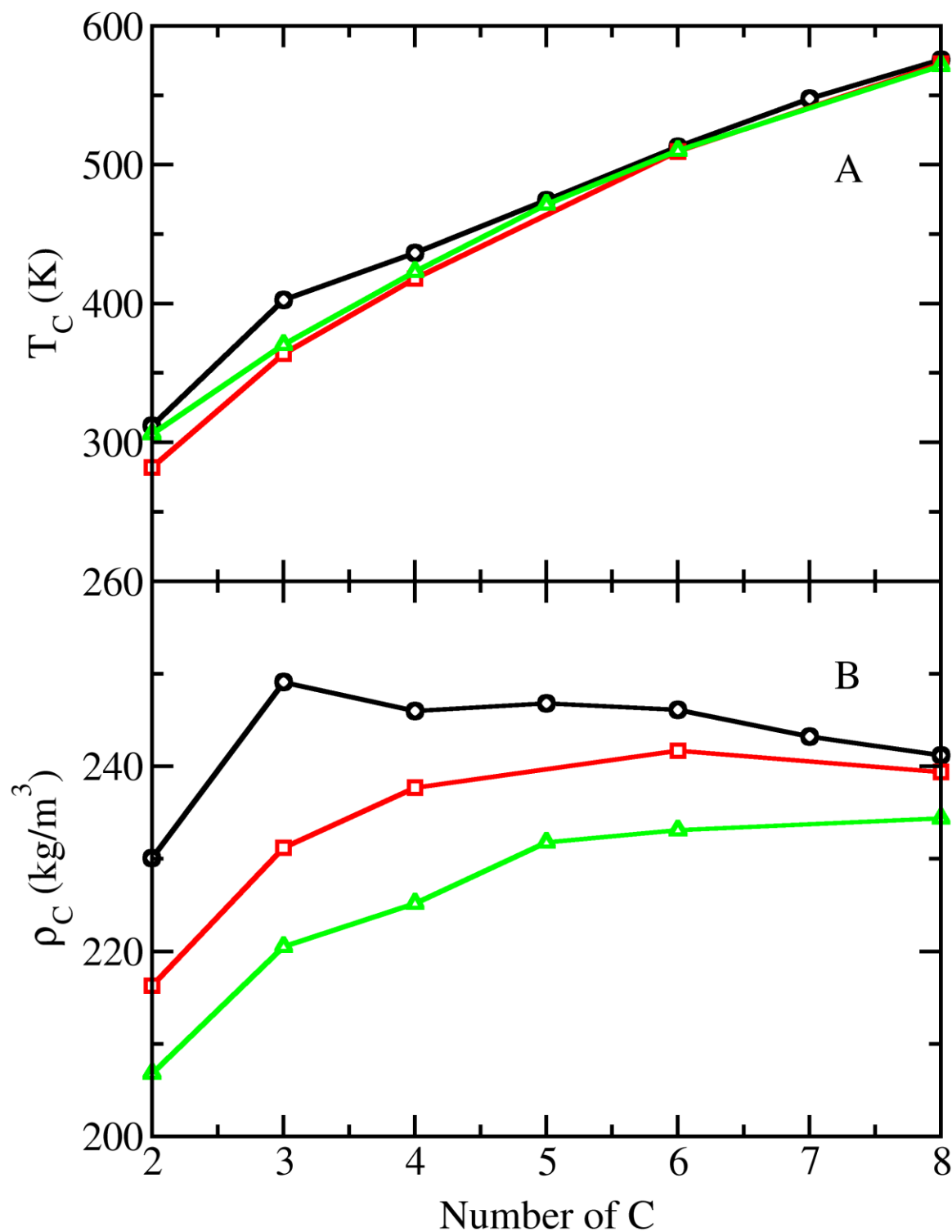
Figure 3.4: Critical properties predicted by Mie potentials versus molecule length for n-alkanes, 1-alkenes, and 1-alkynes. Figure 3.4A: critical temperature; 1-alkynes (black circles); 1-alkenes (red squares); n-alkanes (green triangles); Figure 3.4B: critical density; 1-alkynes (black circles); 1-alkenes (red squares); n-alkanes (green triangles).

3.4.3   Liquid Phase Structure

Radial distribution functions (RDF) were calculated from NPT simulations and used to provide insight into how liquid structure impacts the critical temperatures of hydrocarbons. RDFs for CH_3-CH_3 interactions in propyne are compared to propane, propene, 1-hexyne, and hexane in Figure 3.5A. The first peak of propyne is located at 3.95 Å compared to 4.15 Å for propene, propane, 1-hexyne, and n-hexane. The first peak height for propyne is approximately 25% higher than propene and propane, which have similar peak heights. These results illustrate how propyne molecules are able to pack more efficiently, leading to stronger intermolecular forces and ultimately increased critical temperatures and densities compared to n-alkenes and n-alkanes. The similarity of the RDF for 1-hexyne and n-hexane is consistent with the critical temperatures presented in Figure 3.4A, which show convergence of the critical temperature curves for alkynes, alkenes, and alkanes.

Additional radial distribution functions for second atom interactions in propane (CH_2-CH_2), propene (CH-CH), propyne (C-C), 1-hexyne (C-C), and n-hexane (CH_2-CH_2) are presented in Figure 3.5B. Propane and propene have similar peak heights and locations (5.35 Å), while the first peak for C-C interactions in propyne is shifted by 1.4 Å to 3.95 Å. The first peak of the C-C RDF for 1-hexyne is at 4 Å, compared to 5.35 Å for the CH_2-CH_2 RDF in *n*-hexane. The closer approach of the C(sp) groups in alkynes is consistent with the increased critical density predicted by simulation compared to *n*-alkanes and alkenes.
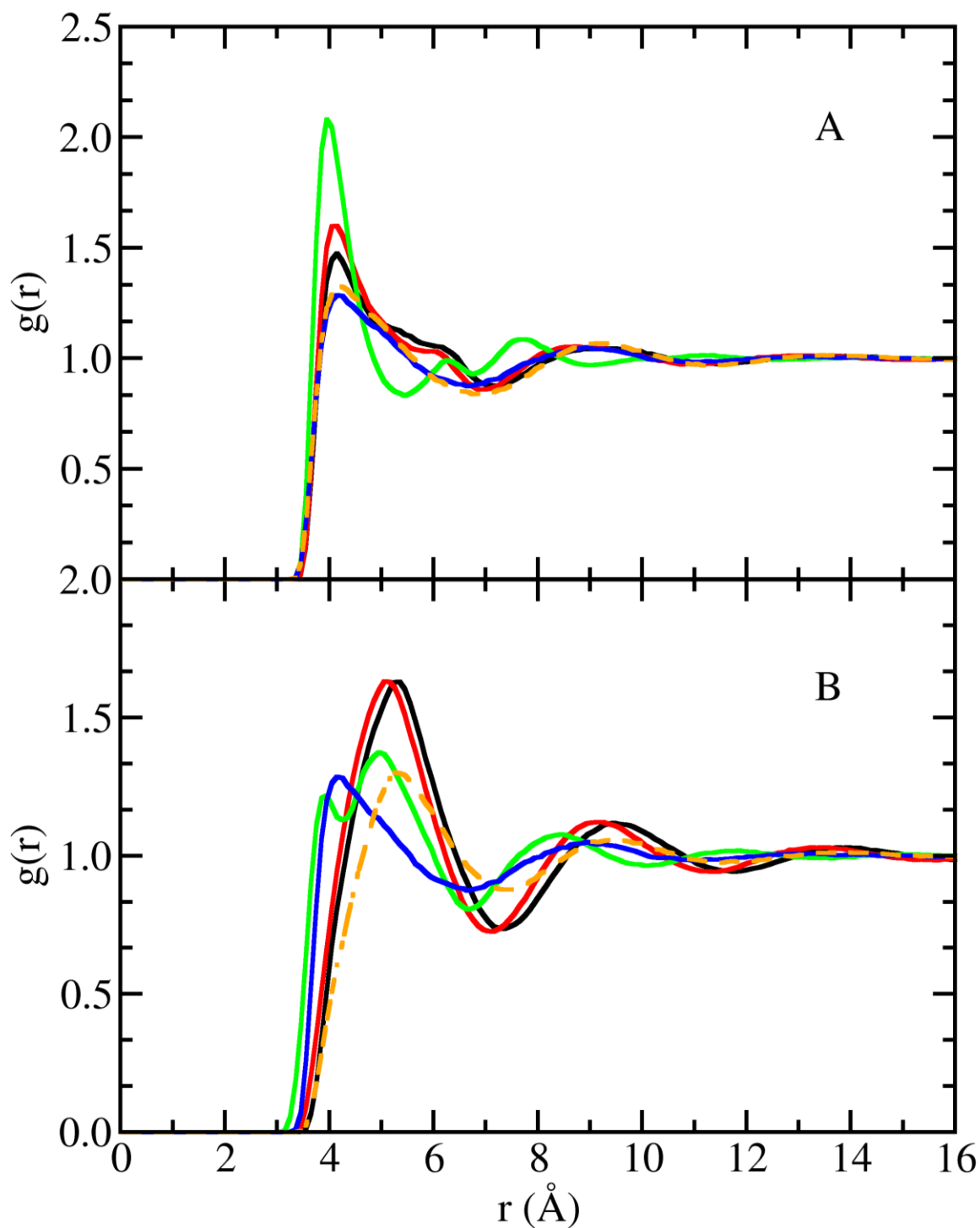
Figure 3.5: Radial distribution functions for propane, propene, propyne, 1-hexyne, and hexane. Figure 3.5A: radial distribution functions for CH3-CH3 interactions; propane (black line); propene (red line); propyne (green line); 1-hexyne (blue line); n-hexane (dashed orange); Figure 3.5B: radial distribution functions for second pseudo-atom in alkane, alkene, and alkyne; CH2-CH2 in propane (black line); CH-CH in propene (red line); C-C in propyne (green line); C-C in 1-hexyne (blue line); CH2-CH2 in n-hexane (dashed orange).

3.4.4 Binary Mixture Vapor-Liquid Equilibria

As mentioned earlier, pure alkynes are very reactive and are usually mixed with other compounds to form a stable solution. Propyne is usually mixed with propadiene to form MAPP gas. Therefore, it is important for the Mie potentials to yield reliable predictions for the pressure composition behavior of propadiene+propyne. The pressure composition diagram predicted by simulation for propadiene+propyne over the temperature range 303.15 to 353.15 K is shown in Figure 3.6. The optimized Mie potentials predict vapor pressures and liquid and vapor mole fractions that are in close agreement with experimental data[159]. Small deviations are observed for calculations performed at higher temperatures and for large concentrations of propadiene due to limitations in the propadiene force field, which over-predicts the vapor pressure by 9% at 358 K.
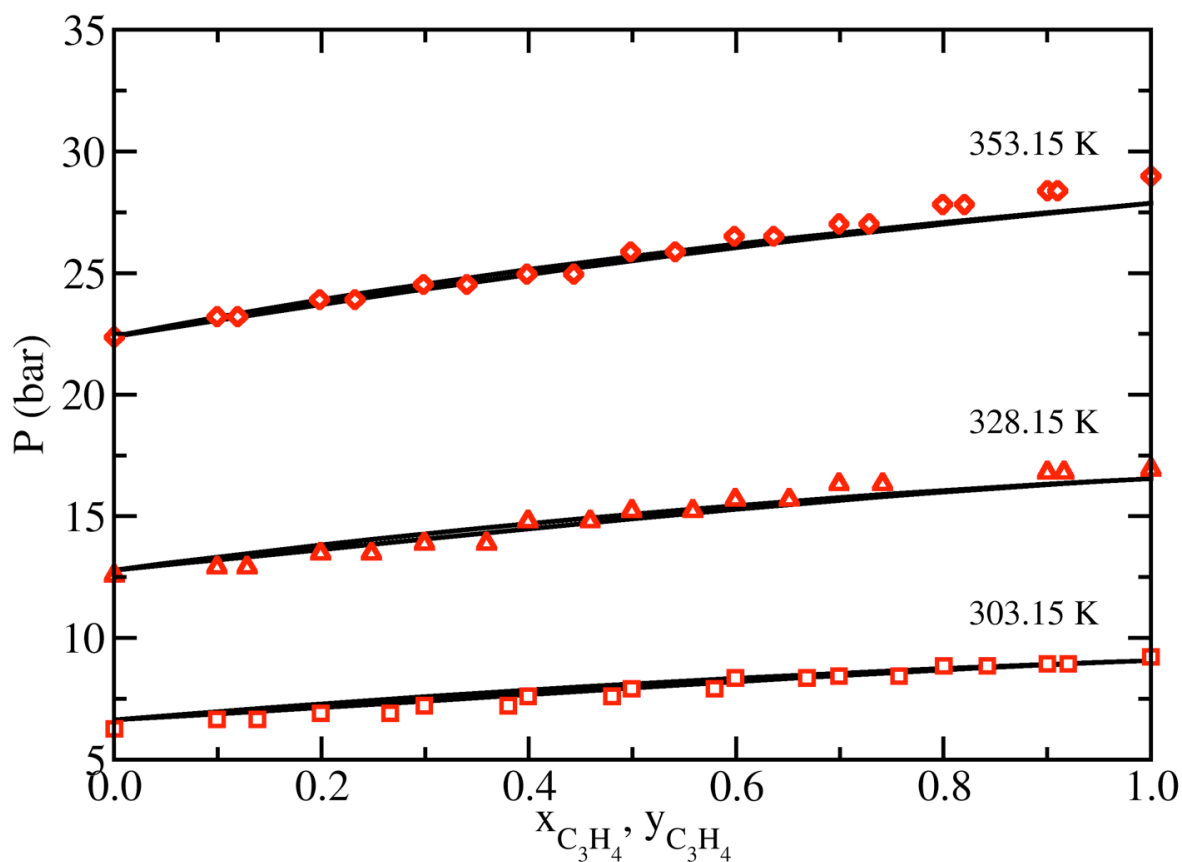
Figure 3.6: Pressure-composition diagram for propadiene+propyne over the temperature range $303.25 \leq T \leq 353.15$ K. Data are represented by: experiment (black lines)[159], optimized Mie potentials (red symbols).

As an additional test of transferability the Mie potentials were used to predict the pressure composition behavior of propane+propyne and propene+propyne over a wide temperature range (278.15 to 353.15 K). The pressure-composition diagram for propane+propyne is shown in Figure 3.7. The Mie potentials were unable to reproduce the azeotropic behavior observed experimentally. These results show that interactions between propane and propyne are over-predicted by approximately 4%. Additional calculations were performed with a propyne force field that included explicit representation of the dipole moment via partial charges. The results of these calculations are presented in the Appendix A, Figure A7, and show that the inclusion of electrostatic interactions does not improve agreement with experiment.
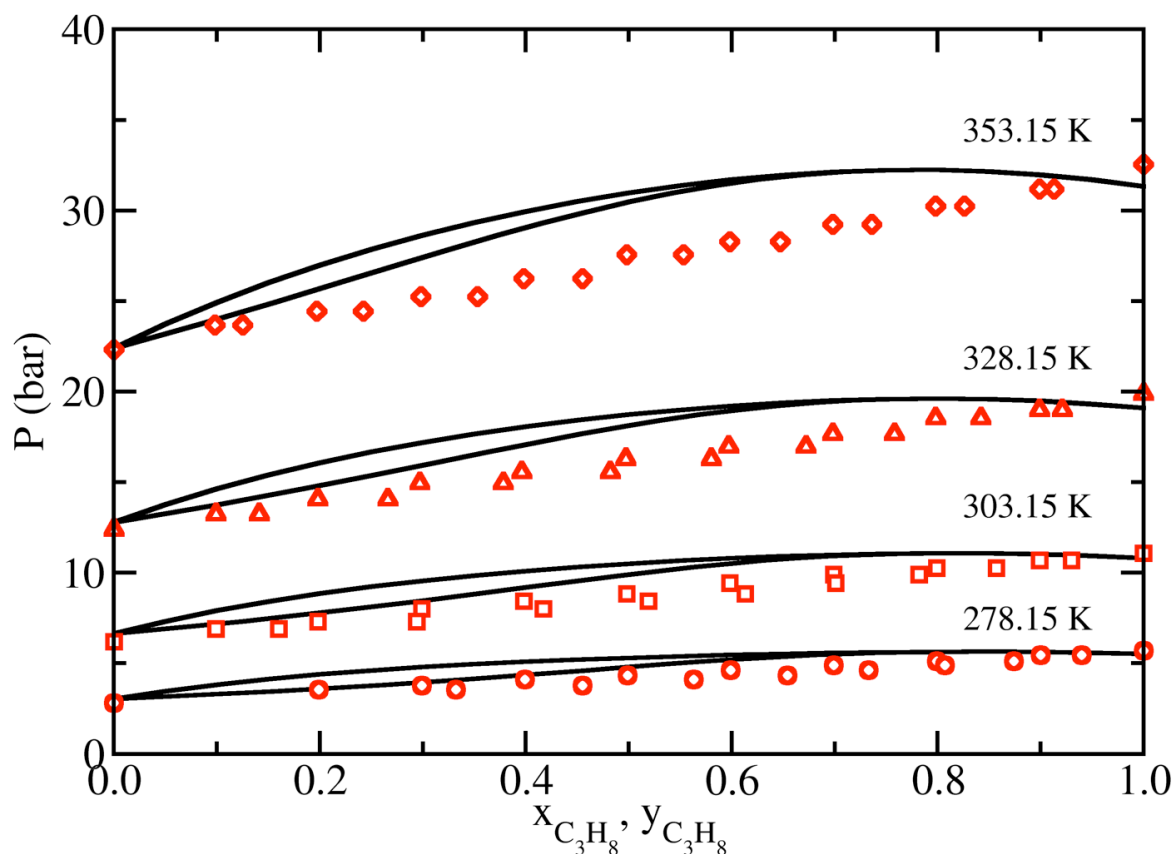
Figure 3.7: Pressure-composition diagram for propane+propyne over the temperature range $278.25 \leq T \leq 353.15$ K. Data are represented by: experiment (black lines)[159], optimized Mie potentials (red symbols).

The pressure composition diagram for propene+propyne is shown in Figure 3.8, and the predictions of simulation are in close agreement with experiment[159]. Small deviations were observed at higher temperatures, which indicate that unlike molecule interactions are slightly too strong.
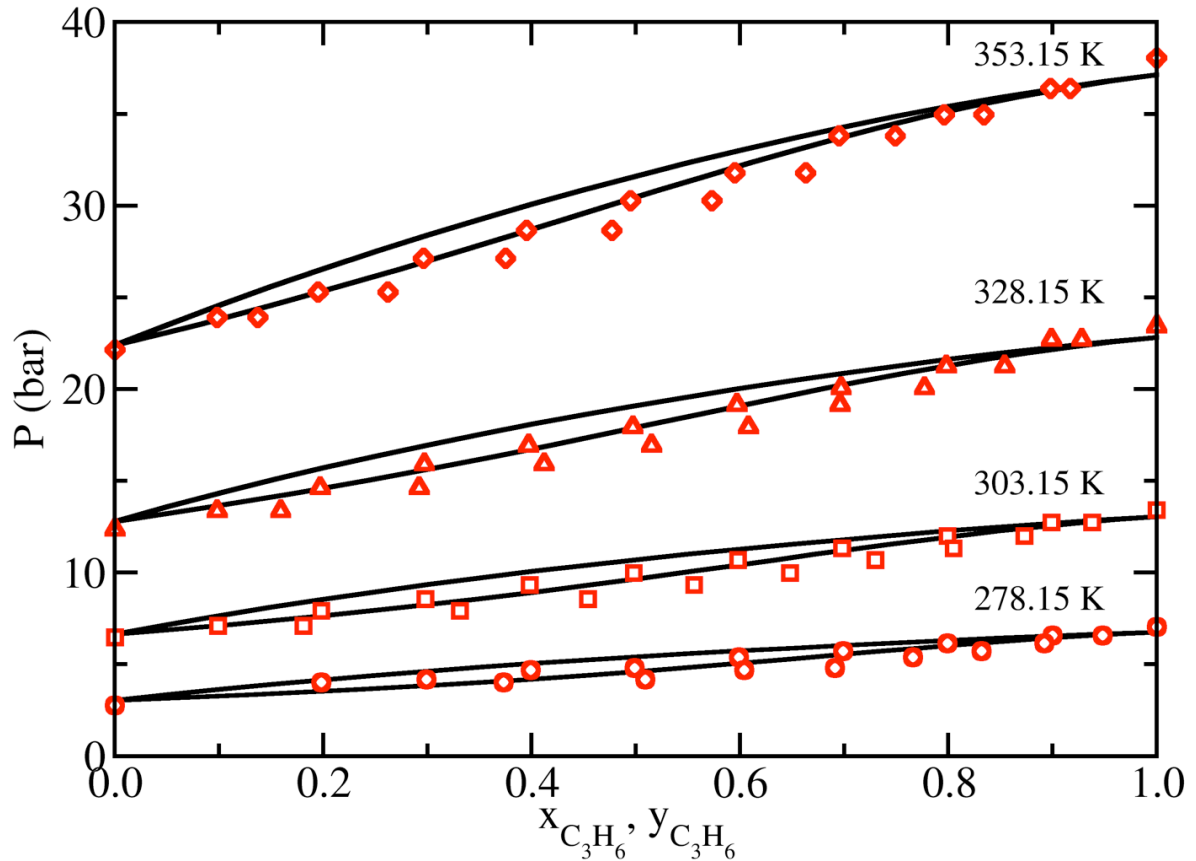
Figure 3.8: Pressure-composition diagram for propene+propyne binary system over the temperature range $278.25 \leq T \leq 353.15$ K. Data are represented by: experiment (black lines)[159], optimized Mie potentials (red symbols).

CHAPTER 4 MOLECULAR EXCHANGE MONTE CARLO: A GENERALIZED METHOD FOR IDENTITY EXCHANGES IN GRAND CANONICAL MONTE CARLO SIMULATION

4.1 Introduction

In Monte Carlo simulations in the grand canonical ensemble (GCMC), the chemical potential, volume and temperature are fixed (μVT = constant). Sampling of phase space is achieved through a variety of trial moves, such as displacement, and molecule insertion and deletion. For complex molecular typologies, additional trial moves, such as rigid body rotation and configurational-bias regrowth[160, 161], may be included to improve the sampling of conformational degrees of freedom. During the course of the simulation, the conjugate variables N (number of molecules) and E (potential energy) fluctuate. Because GCMC allows for the simulation of an open system, it has been used extensively to study the adsorption of gases in porous materials[15-18]. When combined with histogram-reweighting methods[130, 131], GCMC simulations provide precise predictions of vapor-liquid equilibria for pure fluids and mixtures[124, 162], and have been used to determine critical micelle concentrations for model surfactants[163].

Perhaps the greatest challenge with GCMC simulations, however, is achieving a sufficient number of accepted molecular insertion/deletion moves to ensure adequate sampling of phase space. Therefore, significant effort has been expended to develop algorithms that improve the acceptance rate for molecule insertions and deletions. Biasing methods, such as rotational, energy and cavity-bias, were used to improve the efficiency of simulations for the adsorption of benzene and p-xylene in silicalite[65]. The introduction of configurational-bias Monte Carlo enabled the successful simulation of chain molecule adsorption in zeolites[66], which was followed by the coupled-decoupled[67] and reservoir methods[164, 165], which

extended the complexity of systems that could be simulated to include molecules with branch points and rings.

These aforementioned biasing methods have greatly extended the complexity of systems that may be simulated with GCMC simulations, however, at high densities and low temperatures, the acceptance rate for molecule transfers is still unacceptably low due to the difficulty in finding a favorably sized cavity to insert a molecule. For example, in simulations of branched alkanes acceptance rates for molecule transfers at $0.7T_c$ were approximately 0.3%[124]. Others have sought to address these issues through the use of cavity-bias[68-70], to identify favorable locations to attempt molecule insertions, or continuous fractional component Monte Carlo[71, 72], and expanded ensembles[73, 74], where molecules are gradually inserted while the system is allowed to relax locally to minimize steric and energetic penalties due to molecule insertion.

For mixtures, a straightforward approach is to introduce a trial move where the identity of one molecule is changed to that of another[75]. The benefit of such a move is that steric overlaps are reduced significantly, leading to enhanced acceptance for the particle exchange. The identity exchange move has been used in many simulations of single particles in various ensembles, such as semi-grand[80, 166], Gibbs[75, 167, 168] and grand canonical[76, 81, 82]. The methodology has been extended to allow for the exchange of multiple solvent molecules with a polymer chain composed of solvent monomers without changing the coordinates of either polymer or solvent[76]. For the simulation of mixtures of colloids and solvent, it is necessary to swap a large colloid particle for multiple smaller solvent particles. By swapping multiple solvent particles, it is possible to create large enough voids such that a reasonable acceptance rate may be obtained for the insertion of colloid particles[81, 82]. For the exchange of a large particle with multiple small ones, Vink *et al.* used simple random insertions to determine the coordinates for the solvent particles. When inserting a large number of solvent

particles, the potential for overlap increases, reducing the efficiency of the method. To address this issue, Kindt introduced the idea of "solvent repacking" for two-dimensional hard-disk and size asymmetric three-dimensional Lennard-Jones systems, where configurational-bias was used to determine the positions of solvent particles in the large-small particle identity exchange[82, 169]. While a number of publications state that an identity exchange move was used for molecular systems[77-79], a detailed description of the algorithm and the acceptance criteria have not been published to date.

The previously described methods for identity exchange were generally applicable to only the special cases for which they were developed, e.g. single particle exchanges[80], a polymer composed of solvent monomers[76], or large hard particles or disks in a solvent of smaller hard particles[81, 82]. These methods are difficult to generalize to molecular systems of arbitrary molecular topology, and their computational performance is expected to be highly correlated with the type of system for which the move was originally developed. To address these issues, a generalized identity exchange move for simulations in the grand canonical ensemble, referred to as Molecular Exchange Monte Carlo (MEMC), is presented that works for systems of any molecular topology. Three different approaches for the insertion of the large molecule are presented. A derivation of acceptance criteria and the algorithms for performing the MEMC move is provided in the next section for each of the three approaches, while the detailed computational procedure and mathematical calculations are included in the Appendix B. The utility of the three methods and their computational efficiency is illustrated for selected binary mixtures in the Results and Discussion. The key findings of the work are summarized in the Chapter 7. The result of this work has been published in Journal of Chemical Physics[83].

4.2   Methods

To describe the MEMC move in the grand canonical ensemble, it is helpful to consider the case of a large molecule that is exchanged with multiple smaller molecules. However, the methods may be applied without modification to the exchange of molecules of similar size. The original state is called the *old* state, while the state created by the attempted exchange move is called the *new* state. For a given configuration, with N_L large and N_S small molecules, an '*insertion move*" is an attempt to exchange one large molecule with N_{EX} small molecules inside a predefined exchange sub-volume V_{EX}, and a "*deletion move*" is an attempt to exchange N_{EX} small molecules for a large one. The exchange sub-volume is defined as an orthogonal box, where the length of the box in the x and y dimensions are set to the same values for simplicity and the z dimension is set independently. If desired, all three sub-volume box dimensions could be set independently. An orthogonal sub-volume is used instead of a cube or sphere to accommodate large molecules with different aspect ratios. Depending on the method used, the orientation of the exchange sub-volume z-axis may also be varied. Although not used in this work, it is also possible to optimize N_{EX} and V_{EX} "on the fly" during a simulation to maximize the acceptance rate.

The acceptance criterion for a molecular exchange move that satisfies the detailed balance equation is written as

$$K(old \rightarrow new) = K(new \rightarrow old) \qquad (4.1)$$

where $K(i \rightarrow j)$ is the flux of probability from state i to state j. The probability flux is equal to the product of the probability of finding the system in state i, the probability of generating a move that takes state i to state j, and the probability of accepting the move:

$$K(old \rightarrow new) = \mathcal{N}(old) \times \alpha(old \rightarrow new) \times acc(old \rightarrow new) \qquad (4.2)$$

Based on the detailed balance Eq. 4.1, the ratio of the probability of accepting the move from $old \rightarrow new$ to that of its reverse move $new \rightarrow old$ is:

$$\frac{acc(old \rightarrow new)}{acc(new \rightarrow old)} = \frac{\mathcal{N}(new)}{\mathcal{N}(old)} \times \frac{\alpha(new \rightarrow old)}{\alpha(old \rightarrow new)} \tag{4.3}$$

In the deletion move, where one large molecule is exchanged for N_{EX} small molecules, the ratio of the probability of being in the new configuration to the probability of being in the old configuration is

$$\frac{\mathcal{N}(new)}{\mathcal{N}(old)} = \frac{e^{-\beta U(new)}e^{\beta[\mu_L(N_L-1)+\mu_S(N_S+N_{EX})]}}{e^{-\beta U(old)}e^{\beta[N_L\mu_L+N_S\mu_S]}} = \frac{e^{\beta[N_{EX}\mu_S-\mu_L]}}{e^{\beta[U(new)-U(old)]}} \tag{4.4}$$

where $\beta = 1/k_B T$, μ_L and μ_S are the imposed chemical potentials of large and small molecules, respectively. $U(old)$ and $U(new)$ are the potential energies of the system in configuration old and configuration new, respectively.

For the insertion move, where N_{EX} small molecules are exchanged for one large molecule, the ratio of the probability of being in the new configuration to the probability of being in the old configuration is

$$\frac{\mathcal{N}(new)}{\mathcal{N}(old)} = \frac{e^{-\beta U(new)}e^{\beta[\mu_L(N_L+1)+\mu_S(N_S-N_{EX})]}}{e^{-\beta U(old)}e^{\beta[N_L\mu_L+N_S\mu_S]}} = \frac{e^{\beta[\mu_L-N_{EX}\mu_S]}}{e^{\beta[U(new)-U(old)]}} \tag{4.5}$$

The probability of generating the new state, for both insertion and deletion of the large molecule, is given by the product of the probability of locating the center of the exchange sub-volume at a particular point within the simulation box, the probability of choosing N_{EX} particular small molecules, the probability of choosing a particular large molecule, the probability of generating trial configurations for N_{EX} small molecules, and the probability of generating trial configurations for the large molecule,

$$\alpha(old \rightarrow new) = P_{sub-v}(old \rightarrow new) \times P_{pick-S}(old \rightarrow new) \times$$

$$P_{pick-L}(old \rightarrow new) \times P_{pos-S}(old \rightarrow new) \times P_{pos-L}(old \rightarrow new) \tag{4.6}$$

Depending on how the center of the exchange sub-volume is located, the molecules to be exchanged are chosen, and how trial positions are generated, different algorithms to perform the MEMC move may be devised.

4.2.1 ME-1

For the large molecule insertion move, the exchange sub-volume V_{EX} with a random geometric center and a random orientation is defined within the simulation box. For a large molecule deletion move, the geometric center of V_{EX} is located at the centroid of the selected large molecule and its z-axis is aligned with the backbone of the large molecule. See Figure 4.1 for more details.
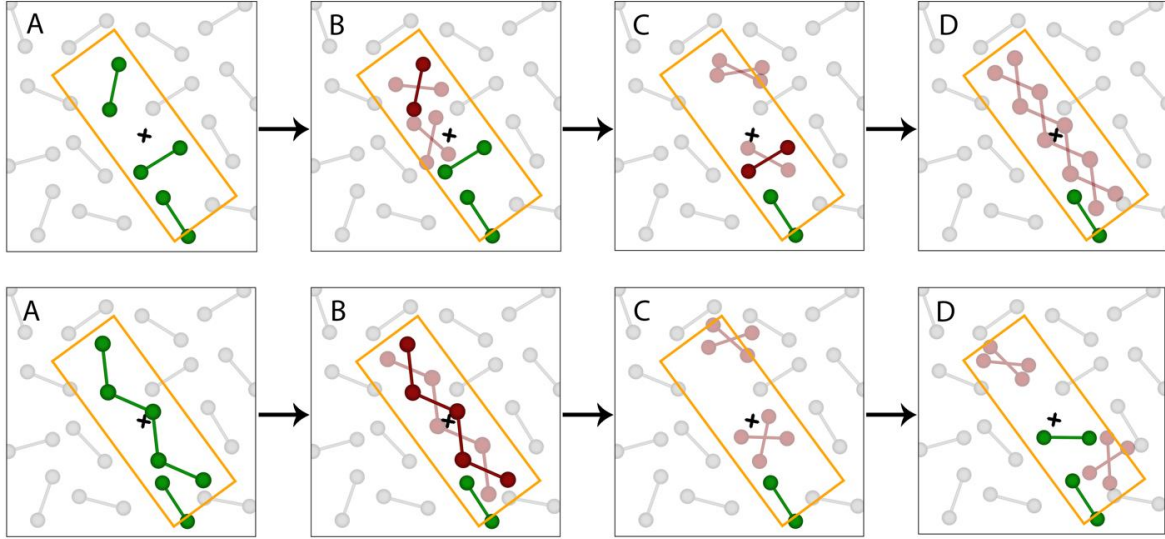
Figure 4.1: Schematic of the ME-1 algorithm. Selected or inserted molecule (green), trial position (light red), and actual position of the molecule (solid red). **Top row**, represents the exchange of two small molecules with one large molecule (insertion). The exchange sub-volume is defined as the orange box. (A) Identifying small molecules within the sub-volume with a random geometric center and orientation. (B) Generating CBMC trials (rotation and centroid location) for one of the small molecules and then removing it. (C) Generating CBMC trials (rotation and centroid location) for the second small molecule and then removing it. (D) Aligning the backbone of the large molecule with the sub-volume and performing CBMC rotational trials around the z-axis of the sub-volume. **Bottom row**, represents the exchange of a large molecule (deletion) with two small molecules. (A) Aligning the sub-volume with large molecule's backbone with geometric center placed at centroid of the large molecule, and identifying the small molecules within the sub-volume. (B) Generating CBMC rotational trials around the z-axis of the sub-volume and then removing it. (C) Generating CBMC trials (rotation and centroid location) for the first small molecule and then inserting it into the sub-volume. (D) Generating CBMC trials (rotation and centroid location) for the second small molecule and then inserting it into the sub-volume.

The algorithm for the insertion of a large molecule after deletion of small molecule(s) is as follows:

1. Define an orthogonal exchange sub-volume V_{EX}, with its geometric center located randomly within the simulation box of volume V (with the probability proportional to V^{-1}) and a random orientation. Determine the total number of small molecules within the exchange sub-volume ($N_{S,VEX}$) based on their geometric center.

2. Reject move if $N_{S,VEX} < N_{EX}$, otherwise continue.

3. Select N_{EX} small molecules out of $N_{S,VEX}$ found in the exchange sub-volume with the probability of $N_{EX}!\left(N_{S,VEX} - N_{EX}\right)!/N_{S,VEX}!$.

4. Repeat steps a and b for N_{EX} cycles ($i = 1, 2, \ldots, N_{EX}$) to delete the selected small molecules.

    a. Generate $j - 1$ random trial positions for the centroid of the ith small molecule within the exchange sub-volume V_{EX}. The original position of the centroid of the ith small molecule will be included as the jth term.

    b. For each trial position p, generate k random trial orientations around the molecule's centroid (except the jth centroid, where $k - 1$ random trial orientations are generated and the original orientation of the molecule will be included as the kth term) and calculate the Rosenbluth weight $W_{i,old} = \sum_{p=1}^{j} \sum_{r=1}^{k} exp\left(-\beta U_{i,p,r}\right)$, were $U_{i,p,r}$ is the interaction energy of the ith molecule to be removed in position p and orientation r with all other molecules, excluding those removed in the earlier cycles of the move. Finally, remove the molecule from the simulation box. Calculate $P_{i,old} = \frac{exp\left(-\beta U_{i,j,k}\right)}{W_{i,old}}$, were $U_{i,j,k}$ is the interaction energy of the ith small molecule at its original centroid position and orientation with all other molecules remaining in the simulation box. $P_{i,old}$ is the probability of inserting the ith small molecule back in its original configuration in the reverse move ($new \rightarrow old$).

5. Insert the centroid of the large molecule at the geometric center of the exchange sub-volume V_{EX} and align the backbone of the large molecule with the z-axis of the exchange sub-volume. Generate k random trial orientations for the large molecule around the z-axis of the sub-volume (two-dimensional rotation). Calculate the Rosenbluth weight $W_{new} =$

$\sum_{r=1}^{k} exp(-\beta U_r)$, where U_r is the interaction energy of the inserted large molecule at orientation r with all other molecules in the simulation box.

6. Select one of the generated trial configurations with the probability $P_{new} = \frac{exp(-\beta U_r)}{W_{new}}$ and insert the large molecule.

The algorithm for the deletion of a large molecule and subsequent insertion of small molecule(s) is as follows:

1. Select a large molecule out of N_L large molecules within the simulation box with probability of $1/N_L$.

2. Define an orthogonal exchange sub-volume with its geometric center placed at the centroid of the selected large molecule, and its z-axis aligned with the backbone of the large molecule. Determine the number of small molecules $N_{S,VEX}$ within the exchange sub-volume.

3. Generate $k - 1$ random trial orientations around the z-axis of the sub-volume The original orientation will be included as the kth term in the Rosenbluth weight. The Rosenbluth weight is calculated as $W_{old} = \sum_{r=1}^{k} exp(-\beta U_r)$, where U_r is the interaction energy of the large molecule in orientation r with all other molecules in the simulation box. Calculate the probability $P_{old} = \frac{exp(-\beta U_k)}{W_{old}}$, where U_k is the interaction energy of the large molecule at the original orientation with all other molecules in the simulation box. P_{old} is the probability of inserting the large molecule at its original configuration in the reverse move ($new \rightarrow old$). Then remove the large molecule from the simulation box.

4. Repeat the steps a→c for N_{EX} cycles ($i = 1, 2, ..., N_{EX}$) to insert the small molecules with the probability of $N_{EX}!/V_{EX}^{N_{EX}}$.

a. Generate j random trial positions for the centroid of the ith small molecule within V_{EX}.

b. For each trial position p, generate k random trial orientations around the molecule's centroid (three-dimensional rotation) and calculate the Rosenbluth weight $W_{i,new} = \sum_{p=1}^{j} \sum_{r=1}^{k} exp(-\beta U_{i,p,r})$, where $U_{i,p,r}$ is the interaction energy of the ith inserted small molecule at position p and orientation r with all the other molecules, including those added in the earlier cycles of the move.

c. Pick one of the generated trial configurations with probability $P_{i,new} = \frac{exp(-\beta U_{i,p,r})}{W_{i,new}}$ and insert the small molecule.

Based on the two algorithms described above, for the large molecule insertion, the ratio of the probability of generating the move new $(N_L + 1, N_S - N_{EX}) \rightarrow old$ (N_L, N_S) to that of the reverse move is:

$$\frac{\alpha(new \rightarrow old)}{\alpha(old \rightarrow new)} = \frac{\frac{1}{N_L + 1}}{\frac{1}{V}} \times \frac{\frac{N_{EX}!}{V_{EX}^{N_{EX}}}}{\frac{N_{EX}!\left(N_{S,VEX} - N_{EX}\right)!}{N_{S,VEX}!}} \times \frac{\prod_{i=1}^{N_{EX}} P_{i,old}}{P_{new}} \qquad (4.7)$$

Simplifying Eq. 4.7 and substituting into Eq. 4.3, produces the acceptance criteria for the large molecule insertion.

$$acc(old \rightarrow new) = min\left\{ 1, \frac{V}{N_L + 1} \times \frac{N_{S,VEX}!}{V_{EX}^{N_{EX}}\left(N_{S,VEX} - N_{EX}\right)!} \times \frac{W_{new}}{\prod_{i=1}^{N_{EX}} W_{i,old}} \times e^{\beta[\mu_L - N_{EX}\mu_S]} \right\} \qquad (4.8)$$

For the large molecule deletion move, the ratio of the probability of generating the move new $(N_L - 1, N_S + N_{EX}) \rightarrow old$ (N_L, N_S) to that of the reverse move is:

$$\frac{\alpha(new \rightarrow old)}{\alpha(old \rightarrow new)} = \frac{\frac{1}{V}}{\frac{1}{N_L}} \times \frac{\frac{N_{EX}! \, N_{S,VEX}!}{(N_{S,VEX} + N_{EX})!}}{\frac{N_{EX}!}{V_{EX}^{N_{EX}}}} \times \frac{\prod_{i=1}^{N_{EX}} P_{i,new}}{P_{old}} \tag{4.9}$$

Simplifying Eq. 4.9 and substituting into Eq. 4.3, produces the acceptance criteria for the large molecule deletion move.

$$acc(old \rightarrow new) = min\left\{ 1, \frac{N_L}{V} \times \frac{V_{EX}^{N_{EX}} \times N_{S,VEX}!}{(N_{S,VEX} + N_{EX})!} \times \frac{\prod_{i=1}^{N_{EX}} W_{i,new}}{W_{old}} \times e^{\beta[N_{EX}\mu_S - \mu_L]} \right\} \tag{4.10}$$

The energy difference between configuration new and old, $U(new) - U(old)$, does not appear directly in the acceptance criteria because their Boltzmann weight is already included in the probabilities used for selecting the position of the molecules.

The acceptance criterion derived for ME-1 is identical to the one introduced by Vink and Horbach[81]. This move performs well for binary mixtures with low concentrations of large molecules. However, the acceptance rate of the move decreases significantly as the concentration of large molecules increases, and the chance of finding N_{EX} small molecules in the exchange sub-volume becomes very low. To address this limitation, ME-2 was developed.

4.2.2 ME-2

In ME-1, for the insertion of a large molecule, the exchange sub-volume V_{EX} is defined with a random orientation and position. However, as the mole fraction of small molecules decreases, the required number of small molecules are frequently not available within the exchange sub-volume. Therefore, a large fraction of the attempted ME-1 moves will be rejected. In the ME-2 approach, the geometric center of V_{EX} is placed on the centroid of a randomly selected small molecule. If the small molecule is monoatomic, the orientation of V_{EX} is assigned randomly, otherwise its z-axis is aligned with the backbone of the small molecule. The large molecule deletion is identical to ME-1. An illustration of the ME-2 algorithm is provided in Figure 4.2.
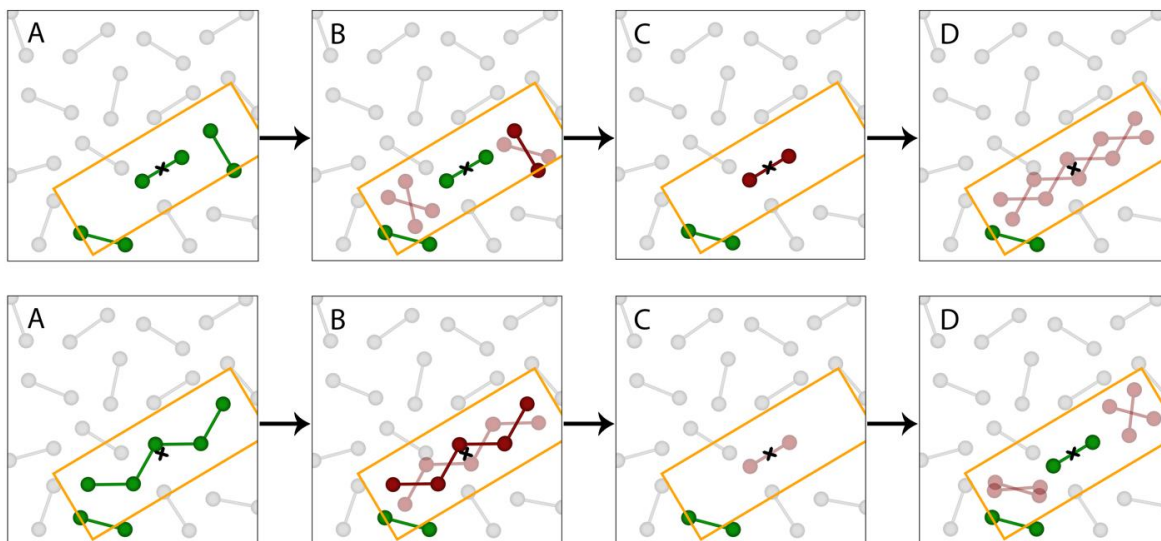
Figure 4.2: Schematic of the ME-2 algorithm. Selected or inserted molecule (green), trial position (light red), and actual position of the molecule (solid red). **Top row**, represents the exchange of two small molecules with one large molecule (insertion). The sub-volume is defined as the orange box. (A) Aligning the sub-volume with a randomly selected small molecule's backbone with geometric center placed at centroid of the selected small molecule, and identifying the small molecules within the sub-volume. (B) Generating CBMC trials (rotation and centroid location) for one of the small molecules and then removing it. (C) Generating CBMC rotational trials around the z-axis of the sub-volume and then removing it. (D) Aligning the backbone of the large molecule with the sub-volume and performing CBMC rotational trials around the z-axis of the sub-volume. **Bottom row** represents the exchange of one large molecule with two small molecules (deletion). (A) Aligning the sub-volume with large molecule's backbone with geometric center placed at centroid of the large molecule, and identifying the small molecules within the sub-volume. (B) Generating CBMC rotational trials around the z-axis of the sub-volume and then removing it. (C) Placing the centroid of the first small molecule at the geometric center of the sub-volume and generate the CBMC rotational trials around the z-axis of the sub-volume and then inserting it into the sub-volume. (D) Generating CBMC trials (rotation and centroid location) for the second small molecule and then inserting it into the sub-volume.

The algorithm for the insertion of a large molecule after deletion of small molecule(s) is as follows:

1. Select one molecule out of N_S small molecules in the simulation box with the probability of $1/N_S$. This molecule will be the last molecule to be removed from the system.

2. Define V_{EX} with its geometric center placed at the centroid of the small molecule selected in step 1. The z-axis of the exchange sub-volume is aligned with the backbone of the small molecule. If the small molecule is monoatomic, the orientation of V_{EX} is assigned randomly.

Determine the number of small molecules $N_{S,VEX}$ within V_{EX} ($N_{S,VEX}$ includes the molecule selected in step 1).

3. Reject the move if $N_{S,VEX} < N_{EX}$, otherwise continue.

4. Select $N_{EX} - 1$ small molecules out of $N_{S,VEX} - 1$, with probability $(N_{EX} - 1)! \left(N_{S,VEX} - N_{EX}\right)! / \left(N_{S,VEX} - 1\right)!$.

5. Repeat steps a and b of the large molecule insertion move of ME-1 for $N_{EX} - 1$ cycles (i = $1, 2, ..., N_{EX} - 1$) to delete the selected small molecules.

6. For the last small molecule to be deleted, generate $k - 1$ random trial orientations around the z-axis of the sub-volume. If the small molecule is monoatomic, orientations are generated around its centroid. The original orientation will be included as the kth term in the Rosenbluth weight. The Rosenbluth weight is calculated from $W_{N_{EX},old} = \sum_{r=1}^{k} exp\left(-\beta U_{N_{EX},r}\right)$, where $U_{N_{EX},r}$ is the interaction energy of the last small molecule in orientation r with all other molecules in the simulation box. Finally, remove the last small molecule from the simulation box and calculate $P_{N_{EX},old} = \frac{exp\left(-\beta U_{N_{EX},k}\right)}{W_{N_{EX},old}}$, where $U_{N_{EX},k}$ is the interaction energy of the last small molecule at its original configuration with all other molecules remaining in the simulation box. $P_{i,old}$ is the probability of inserting the ith small molecule back at its original configuration in the reverse move ($new \rightarrow old$).

7. Insert the large molecule according to steps 5 and 6 of ME-1.

The algorithm for the deletion of a large molecule and subsequent insertion of small molecule(s) is as follows:

1. Follow steps 1-4 of the ME-1 large molecule deletion move.

2. Insert the centroid of the first small molecule at the geometric center of V_{EX} and align its backbone with the z-axis of the exchange sub-volume. Generate k random trial orientations around the z-axis of the sub-volume. If small molecules are monoatomic, the orientation is

assigned randomly around its centroid. Calculate the Rosenbluth weight $W_{1,new} = \sum_{r=1}^{k} exp(-\beta U_{1,r})$, where $U_{1,r}$ is the interaction energy of the first small molecule inserted at orientation r with all other molecules in the simulation box.

3. Select one of the trial orientations with the probability $P_{1,new} = \frac{exp(-\beta U_{1,r})}{W_{1,new}}$.

4. Repeat steps a→c of the large molecule deletion move of ME-1 for $N_{EX} - 1$ cycles ($i = 2, ..., N_{EX}$) to insert the small molecules with probability $(N_{EX} - 1)!/V_{EX}^{(N_{EX}-1)}$.

Based on the two algorithms described above, for the large molecule insertion move, the ratio of the probability of generating move new $(N_L + 1, N_S - N_{EX}) \rightarrow old$ (N_L, N_S) to that of the reverse move is:

$$\frac{\alpha(new \rightarrow old)}{\alpha(old \rightarrow new)} = \frac{\frac{1}{N_L + 1}}{\frac{1}{N_S}} \times \frac{\frac{(N_{EX} - 1)!}{V_{EX}^{(N_{EX}-1)}}}{\frac{(N_{EX} - 1)!\left(N_{S,VEX} - N_{EX}\right)!}{\left(N_{S,VEX} - 1\right)!}} \times \frac{\prod_{i=1}^{N_{EX}} P_{i,old}}{P_{new}} \qquad (4.11)$$

Simplifying Eq. 4.11 and substituting into Eq. 4.3 results in the acceptance criterion for the large molecule insertion move:

$$acc(old \rightarrow new) = min\left\{1, \frac{N_S}{N_L + 1} \times \frac{\left(N_{S,VEX} - 1\right)!}{V_{EX}^{(N_{EX}-1)}\left(N_{S,VEX} - N_{EX}\right)!} \times \frac{W_{new}}{\prod_{i=1}^{N_{EX}} W_{i,old}} \times e^{\beta[\mu_L - N_{EX}\mu_S]}\right\} \qquad (4.12)$$

For the large molecule deletion move, the ratio of the probability of generating configuration new $(N_L - 1, N_S + N_{EX}) \rightarrow old$ (N_L, N_S) to that of the reverse move is:

$$\frac{\alpha(new \rightarrow old)}{\alpha(old \rightarrow new)} = \frac{\frac{1}{(N_S + N_{EX})}}{\frac{1}{N_L}} \times \frac{\frac{(N_{EX} - 1)! N_{S,VEX}!}{\left(N_{S,VEX} + N_{EX} - 1\right)!}}{\frac{(N_{EX} - 1)!}{V_{EX}^{(N_{EX}-1)}}} \times \frac{\prod_{i=1}^{N_{EX}} P_{i,new}}{P_{old}} \qquad (4.13)$$

Simplifying Eq. 4.13 and substituting into Eq. 4.3 results in the acceptance criterion for the large molecule deletion move.

$$acc(old \rightarrow new) = min\left\{1, \frac{N_L}{(N_S + N_{EX})} \times \frac{V_{EX}^{(N_{EX}-1)} \times N_{S,VEX}!}{(N_{S,VEX} + N_{EX} - 1)!} \times \frac{\prod_{i=1}^{N_{EX}} W_{i,new}}{W_{old}} \times e^{\beta[N_{EX}\mu_S - \mu_L]}\right\} \qquad (4.14)$$

If $N_{EX} = 1$, the acceptance criteria given in Eqs. 4.13 and 4.14 simplifies to that of the standard identity-exchange acceptance move[80].

$$acc(N_L \rightarrow N_L + 1) = min\left\{1, \frac{N_S}{(N_L + 1)} \times \frac{W_{new}}{W_{old}} \times e^{\beta[\mu_L - \mu_S]}\right\} \qquad (4.15)$$

$$acc(N_L \rightarrow N_L - 1) = min\left\{1, \frac{N_L}{(N_S + 1)} \times \frac{W_{new}}{W_{old}} \times e^{\beta[\mu_S - \mu_L]}\right\} \qquad (4.16)$$

4.2.3 ME-3

For the large molecule insertion move in ME-2, the large molecule is inserted as a rigid body and its backbone is aligned with the z-axis of the V_{EX}. This move performs well for large molecules with a straight backbone. However, the acceptance rate decreases for a large molecule with nonlinear geometry as it becomes significantly more difficult to fit a complex rigid body into the void space created after deleting the small molecule(s). Therefore, a modification to ME-2 was developed to address this limitation.

In the ME-3 algorithm, a predefined atom of the large molecule is first placed at the geometric center of V_{EX} and the molecule is built segment by segment using the coupled-decoupled configurational-bias Monte Carlo (CBMC) algorithm[67]. For the large molecule deletion move, the exchange sub-volume is defined with a random orientation, with its geometric center placed at the same predefined atom of the large molecule to be deleted. Next, the Rosenbluth weight W_{old} of the large molecule is calculated. Insertion and deletion of N_{EX} small molecules are identical to the ME-2 method. Figure 4.3 illustrates the ME-3 algorithm.
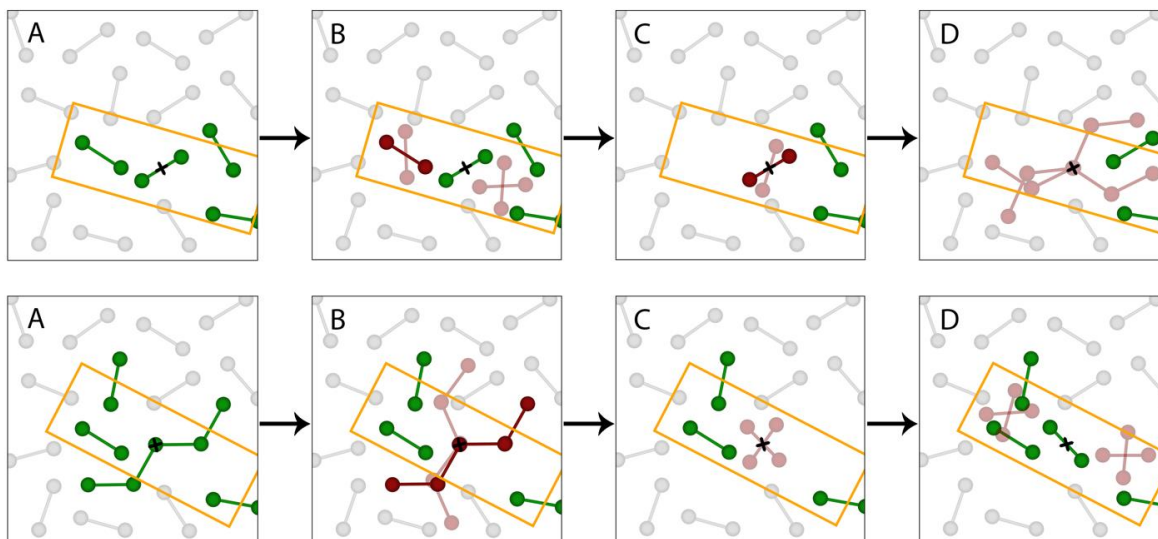
Figure 4.3: Schematic of the ME-3. Selected or inserted molecule (green), trial position (light red), and actual position of the molecule (solid red). **Top row**, represents the exchange of two small molecules with one large molecule (insertion). The sub-volume is defined as the orange box. (A) Defining the sub-volume with a random orientation, where its geometric center is placed at a randomly selected small molecule's centroid, and identifying the small molecules within the sub-volume. (B) Generating CBMC trials (rotation and centroid location) for one of the small molecules and then removing it. (C) Generating CBMC rotational trials around its centroid of the selected small molecule and then removing it. (D) Placing the predefined atom of the large molecule at the geometric center of the sub-volume and growing the large molecule using coupled-decoupled CBMC technique. **Bottom row**, represents the exchange of a large molecule with two small molecules (deletion). (A) Defining the sub-volume with a random orientation with geometric center placed at the predefined atom of the large molecule, and identifying the small molecules within the sub-volume. (B) Generating coupled-decoupled CBMC trials and then removing it. (C) Placing the centroid of the first small molecule at the geometric center of the sub-volume, generating CBMC rotational trials around its centroid and then inserting it into the sub-volume. (D) Generating CBMC trials (rotation and centroid location) for the second small molecule and then inserting it into the sub-volume.

The algorithm for the insertion of a large molecule after deletion of small molecule(s) is as follows:

1. Select one molecule out of N_S small molecules in the simulation box with probability $1/N_S$.

   This molecule will be the last molecule to be removed from the system.

2. Define an orthogonal exchange sub-volume V_{EX} with a random orientation and its

   geometric center placed at the centroid of the small molecule selected above. Then

determine the number of small molecules $N_{S,VEX}$ within V_{EX} ($N_{S,VEX}$ includes the molecule selected in step 1).

3. Repeat steps 3-6 of the ME-2 method to delete N_{EX} small molecules from simulation box.

4. Insert the predefined atom of the large molecule at the center of V_{EX} and perform coupled-decoupled configurational-bias Monte Carlo to grow the large molecule segment by segment. Calculate the Rosenbluth weight W_{new}.

5. Insert the large molecule by selecting one of the generated trial configurations with the probability P_{new}.

The algorithm for the deletion of a large molecule and subsequent insertion of small molecule(s) is as follows:

1. Within the simulation box of volume V, pick one large molecule out of N_L with probability of $1/N_L$.

2. Define an orthogonal exchange sub-volume V_{EX} with a random orientation and place its geometric center at the predefined atom of the selected large molecule. Determine the number small molecules $N_{S,VEX}$ within the exchange sub-volume.

3. Perform coupled-decoupled CBMC for the large molecule and calculate the Rosenbluth weight W_{old} and P_{old}.

4. Repeat steps 2-4 of ME-2 to insert N_{EX} small molecules within V_{EX}.

The forward to reverse probability ratios for generating the large molecule insertion and the large molecule deletion moves are identical to those given in Eq. 4.11 and 4.13, respectively. The acceptance criteria for the ME-3 algorithm is identical to that of ME-2 and are given by Eq. 4.12 and 4.14.

4.3   Simulation Methodology

The three molecular exchange algorithms described in this chapter were implemented in the development version of GPU Optimized Monte Carlo[57] (GOMC), which is available to the public on GitHub[170]. GOMC is an object-oriented Monte Carlo simulation engine, capable of performing simulations in canonical, isobaric-isothermal, and grand canonical ensembles, as well as Gibbs ensemble Monte Carlo. GOMC is designed for the simulation of complex molecular topologies and supports a variety of potential functions, such as Lennard-Jones and Mie potentials. Coulomb interactions are also supported via the Ewald summation method[122]. GOMC is capable of parallel computation, either on multicore CPUs or GPUs.

Phase diagrams were determined from histogram-reweighting Monte Carlo simulations in the grand-canonical ensemble[42]. A cubic box size of 25 Å × 25 Å × 25 Å was used for methane+ethane, methane+propane, methane+n-butane, and water+impurity. For perfluorobutane+n-butane and methane+n-pentane, a box size of 30 Å × 30 Å × 30 Å was used, while for 2,2,4-trimethylpentane+neopentane a box size of 40 Å × 40 Å × 40 Å was used. Initial configurations were generated with Packmol[142]. Psfgen was used to generate coordinate (*.pdb) and connectivity (*.psf) files[143]. Potentials were truncated at 10 Å and analytical tail corrections were applied[144]. To enhance the acceptance rate for molecule insertions, the coupled-decoupled configurational-bias Monte Carlo (CBMC) algorithm was used[67]. For all liquid phase simulations, unless otherwise noted in the Results and Discussion, configurational-bias parameters were: 100 angle trials, 100 dihedral trials, 10 trial locations for the first site, and 8 trial locations for secondary sites. For standard GCMC simulations, a move ratio of 20% displacements, 10% rotations, 10% regrowth, and 60% molecule transfers was used. For simulations that included the molecular exchange move, 30% molecular exchanges were performed with a corresponding reduction in the percentage of attempted molecule transfers.

Uncertainties used in the calculation of the statistical efficiency of the methods were calculated as the standard deviation determined from five unique simulation trajectories, each started from a unique initial configuration and random number seed. All simulations, except those used to generate phase diagrams, were run for 2×10^7 Monte Carlo steps (MCS), without equilibration period. Simulations used to generate phase diagrams were run for 5×10^7 MCS with a 5×10^6 MCS equilibration period. Every 200-500 MCS, the instantaneous state of the system (N_1, N_2, E) was saved as a histogram. Every one million MCS, the natural log of distribution of large particle $\ln(P_N)$ for each simulation was determined, and the standard deviation and efficiency were calculated for each binary system for a variety of compositions along the vapor-liquid coexistence curve. Calculations were performed on one core of an Intel Xeon E5-4627v4 2.6 GHz CPU.

The efficiency was computed using the calculated standard deviation and the CPU time.

$$\eta = (\sigma^2 s)^{-1} \tag{4.17}$$

where σ is average uncertainty in natural log of large particle distribution and s is the CPU time in seconds.

4.4 Results and Discussion

In this chapter, a number of examples are provided to illustrate the effect of molecular exchange moves on the statistical sampling in grand canonical histogram reweighting Monte Carlo simulations. Mixtures simulated include perfluorobutane+n-butane, and methane +ethane, +propane, +n-butane, and +n-pentane. Additional calculations were performed to generate pure fluid phase diagrams for water and 2,2,4-trimethylpentane to demonstrate the utility of the method and to provide comparisons to prior work[171-173]. For binary mixture phase diagrams, all calculations were performed at temperatures below $0.7T_c$. For pure fluid phase diagrams, calculations were performed from the critical temperature to $0.44T_c - 0.51T_c$.

Performing grand canonical Monte Carlo simulations, using standard configurational-bias methods[67], below $0.7T_c$ is a challenging task, and therefore a good test to evaluate the improvement in sampling of phase space provided by the proposed algorithms.

4.4.1   Methane+n-alkane

Methane+n-alkane systems are well studied and extensive experimental data may be found in the literature[174-181]. In general, the determination of vapor-liquid coexistence for these systems at temperatures above $0.7T_c$ can be done using standard configurational-bias methods in grand canonical or Gibbs ensemble Monte Carlo simulations[24, 124, 141, 182]. However, below $0.7T_c$, acceptance rates for the insertion of n-alkanes into a liquid phase drops to approximately 0.1%, which necessitates long simulations to obtain convergence of the simulations. In this section, the effect of the three ME algorithms on the convergence of grand canonical Monte Carlo simulations is assessed for mixtures of methane +ethane, +propane, +n-butane, and +n-pentane, and the effectiveness of performing a two for one exchange is evaluated.

The methane+n-butane mixture is presented first as an example of the validation process used in the development of the molecular exchange methods. Grand canonical Monte Carlo (GCMC) simulations were performed for a variety of temperatures, chemical potentials, and move ratios using both standard configurational-bias insertions/deletions and the ME-1, ME-2, and ME-3 methods. Probability distributions of states sampled during the simulation were collected and compared to reference distributions determined using standard configuration-bias insertions. An example of this is shown in Figure 4.4, for gas $(\mu_{butane} = -2960, \mu_{methane} = -2000)$ and liquid $(\mu_{butane} = -2840, \mu_{methane} = -2000)$ phase simulations at 277 K. As expected, the probability distributions produced by the ME-3 algorithm are an exact match to the reference distributions. Additional data for the ME-1 and ME-2 algorithms are presented in the Appendix B, Figures B1 and B2.
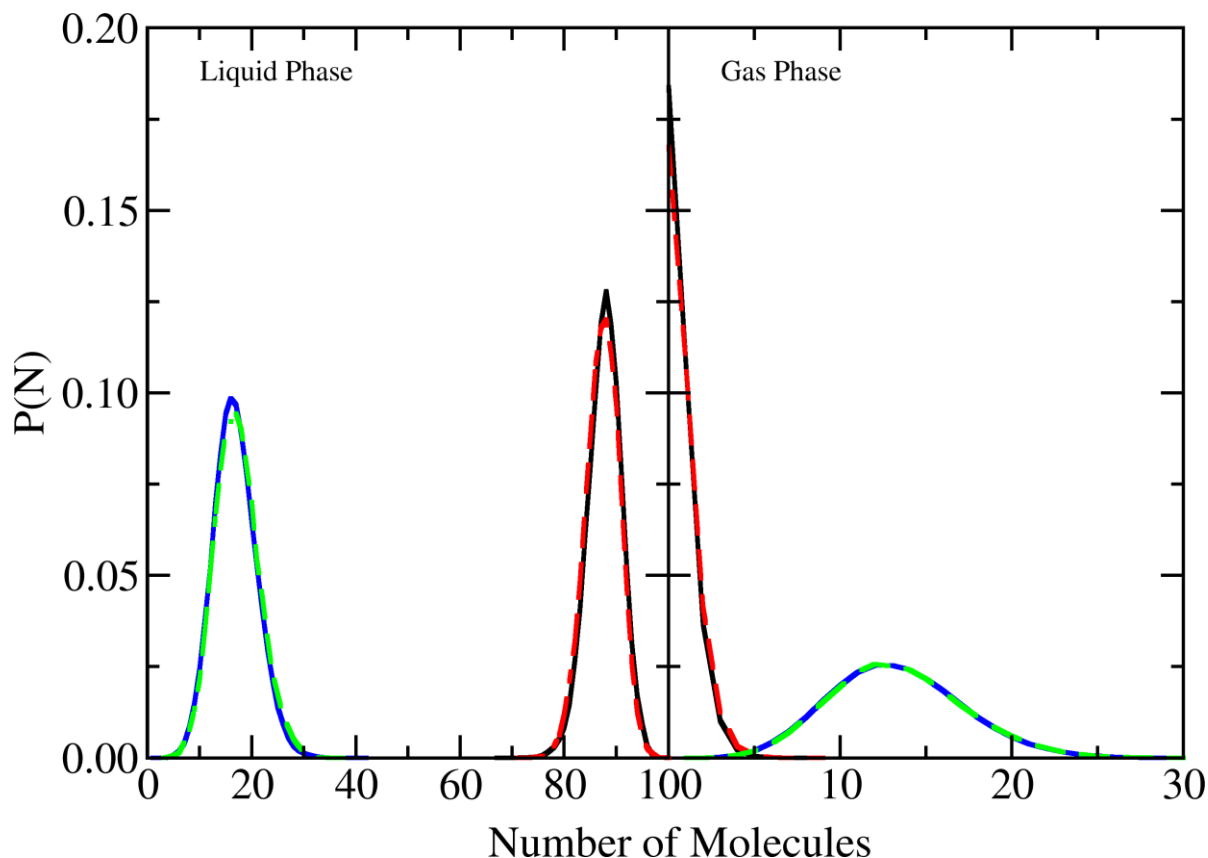
Figure 4.4: Probability distributions predicted from gas $(\mu_{butane} = -2960, \mu_{methane} = -2000)$ and liquid $(\mu_{butane} = -2840, \mu_{methane} = -2000)$ phase GCMC simulations of methane+n-butane at 277 K. Solid lines denote the probability distributions for n-butane (black) and methane (blue) using standard configurational-bias insertions and deletions. Dashed lines denote the probability distributions for n-butane (red) and methane (green) using the ME-3 algorithm.

In Figure 4.5, the pressure vs. composition diagram for methane+n-butane at 277 K, predicted using both the coupled-decoupled configurational-bias method[67] and the ME-3 algorithm, is shown. Interactions between molecules were described with Optimized Mie Potentials for Phase Equilibria[24]. In addition to showing excellent agreement with experimental data[181], the ME-3 algorithm produced results that are statistically indistinguishable from standard configurational-bias insertions, providing further validation of the method.
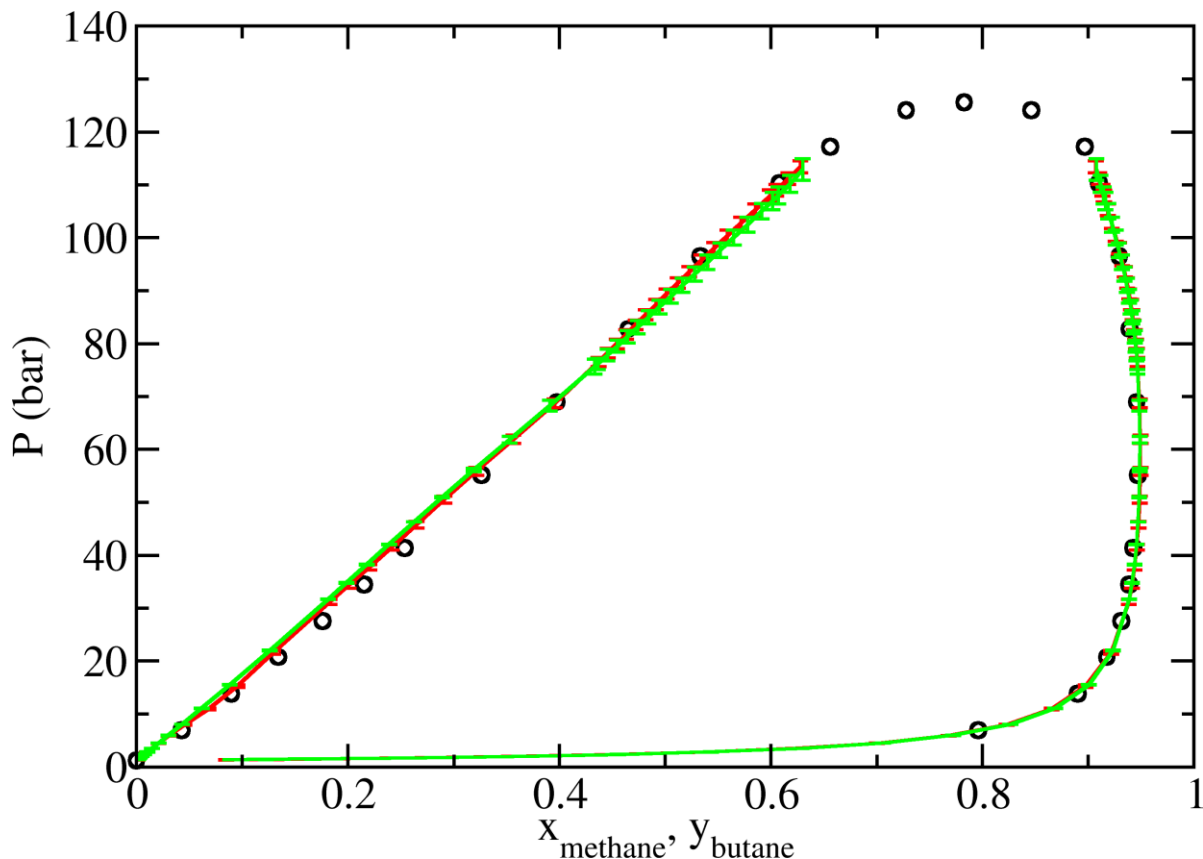
Figure 4.5: Pressure composition diagram for methane+n-butane at 277 K predicted from GCMC+histogram reweighting simulations using Mie potentials*[24]*. Experimental data (circles)[181], standard configurational-bias insertions (red lines), ME-3 algorithm (green lines).

In Table 4.1, the acceptance rate for molecule transfers as a function of composition is presented for each methane+n-alkane binary mixture. Calculations were performed for liquid phase simulations along the coexistence curve at 186 K (methane+ethane), 213 K (methane+propane), 225 K (methane+n-butane), and 273 K (methane+n-pentane). The systems exhibit similar general trends, with acceptance rates climbing as the critical point of the mixture is reached. For $x_{methane} < 0.5$, acceptances rates for the insertion of the larger n-alkane using configurational-bias were less than 1%. When performing a one to one exchange, ME-3 was found to produce the largest improvement in acceptance rates for the large molecule, producing improvements of 2X for methane+n-pentane at $x_{methane} = 0.7$ to 70X for methane+ethane at $x_{methane} = 0.1$. The ME-2 algorithm also produced significant enhancement in the

acceptances rates for the insertion of the longer n-alkane, while the ME-1 algorithm was found to yield significantly lower acceptance rates than traditional configurational-bias insertions. Because the ME-2 algorithm uses a rigid swap and the centroid of the large molecule is placed at the geometric center of the exchange sub-volume, only a fraction of the sub-volume is guaranteed to be empty. In most of the ME-2 exchanges, it is likely that some atoms from the large molecule will overlap with existing molecules, lowering acceptance rates compared to ME-3. The ME-3 algorithm uses the same initial placement for the central atom as ME-2, but grows the rest of the large molecule, allowing it to find more energetically favorable configurations than are possible through a rigid molecule insertion, leading to greater acceptance rates for the exchange move. As expected, the more similar the large and small molecule were in terms of excluded volume, the greater the success of the molecular exchange. It is also interesting to note that even for the highly asymmetric mixture of methane+n-pentane, acceptance rates for molecule transfers were improved substantially through the inclusion of the molecular exchange move.

The molecular exchange algorithm allows for trial moves where any number of small molecules may be exchanged for one large molecule. An example of this is shown in Table 1, where acceptance rates are presented for exchange of two methanes with one n-butane or n-pentane ($N_{EX} = 2$). For the ME-3 algorithm, acceptance rates are always lower than the one for one exchange, although, this difference decreases as the chain length of the large molecule increases. Part of the decrease in the acceptance rate stems from the reduced probability of finding two methane molecules in the sub-volume to exchange at low methane concentrations. For ME-2, acceptance rates are slightly lower for the exchange of two methanes with one n-butane, compared to the one for one exchange. However, for the exchange of two methanes with one n-pentane, slight improvements in the acceptance rates were observed. The ME-1 algorithm shows a slight improvement in acceptance rates for the exchange of two methanes

with one n-butane or n-pentane, although in all cases, acceptance rates for the ME-1 algorithm are lower than configurational-bias insertions.

While size of the sub-volume does not have an effect on the acceptance rates for the ME-2 and ME-3 algorithms for a one to one exchange, it was found to have an effect on the acceptance rates for the two to one exchange, as shown in Table 1. Increasing the size of the sub-volume increases the probability that a second small molecule will be found within the sub-volume, leading to an increased overall acceptance rate for the MEMC move. Therefore, it is possible to optimize acceptance rates for the two to one exchange ratio by performing a series of short simulations for a range of sub-volume box lengths, and by using a heuristic that the sub-volume should be large enough to contain the entire large molecule. For methane+n-butane, the optimum exchange sub-volume size for a two for one exchange was found to be 8.8 Å $\times$ 8.8 Å $\times$ 11.8 Å for ME-3 and ME-2, while for ME-1 it was 5 Å $\times$ 5 Å $\times$ 8 Å.

Table 4.1: n-alkane insertion/removal acceptance percentages in GCMC liquid phase simulations of methane+n-alkane mixtures for CBMC, ME-1, ME-2, and ME-3 methods.

Binary system	Sub-volume size (Å)	N_{EX}	x_{CH_4}	CBMC	ME-1	ME-2	ME-3
methane+ethane	$5 \times 5 \times 6$	1	0.1	0.33	0.11	11.68	23.62
			0.5	1.47	0.96	16.20	33.33
			0.9	8.3	4.18	24.09	47.84
methane+propane	$5 \times 5 \times 7$	1	0.1	0.08	0.05	3.42	4.13
			0.4	0.38	0.40	5.67	7.21
			0.8	5.18	3.36	13.56	18.36
methane+n-butane	$5 \times 5 \times 8$	1	0.1	0.14	0.025	0.835	2.373
			0.3	0.33	0.099	1.207	3.421
			0.6	2.52	0.948	3.378	8.128
	$5 \times 5 \times 8$	2	0.1	0.14	0.019	0.196	0.362
			0.3	0.33	0.144	0.557	0.928
			0.6	2.52	1.262	2.288	3.160
	$8.8 \times 8.8 \times 11.8$	2	0.1	0.14	0.022	0.398	0.984
			0.3	0.33	0.086	0.821	1.860
			0.6	2.52	0.621	2.682	5.252
methane+n-pentane	$5 \times 5 \times 9$	1	0.1	0.064	0.007	0.209	0.824
			0.5	0.397	0.116	0.638	2.163
			0.7	2.461	0.666	1.72	4.814
	$5 \times 5 \times 9$	2	0.1	0.639	0.006	0.086	0.189
			0.5	0.397	0.270	0.736	1.160
			0.7	2.461	1.332	2.389	3.170
	$8.8 \times 8.8 \times 13$	2	0.1	0.639	0.008	0.145	0.455
			0.5	0.397	0.102	0.675	1.806
			0.7	2.461	0.473	2.054	4.133

A more detailed analysis of the statistical uncertainty and efficiency for an exchange ratio of one to one is provided in Figure 4.6 for the methane+n-butane mixture. A direct comparison between the efficiencies obtained for the one to one and one to two exchange ratios are presented in the Appendix B, Figure B3. Uncertainties were determined from probability distributions collected from liquid phase grand canonical Monte Carlo simulations performed along the vapor-liquid coexistence curve. For all mole fractions investigated, the ME-3 algorithm shows the fastest convergence of the n-particle probability distribution, converging

in approximately half the number of Monte Carlo steps of ME-2. Both the ME-3 and ME-2 algorithms produce similarly converged probability distributions after 2×10^7 MCS, with average uncertainties of approximately 0.05. The ME-1 algorithm and configurational-bias insertions show similar convergence properties. However, with 2×10^7 MCS each produced uncertainties that were approximately double those of the ME-3 and ME-2 methods.
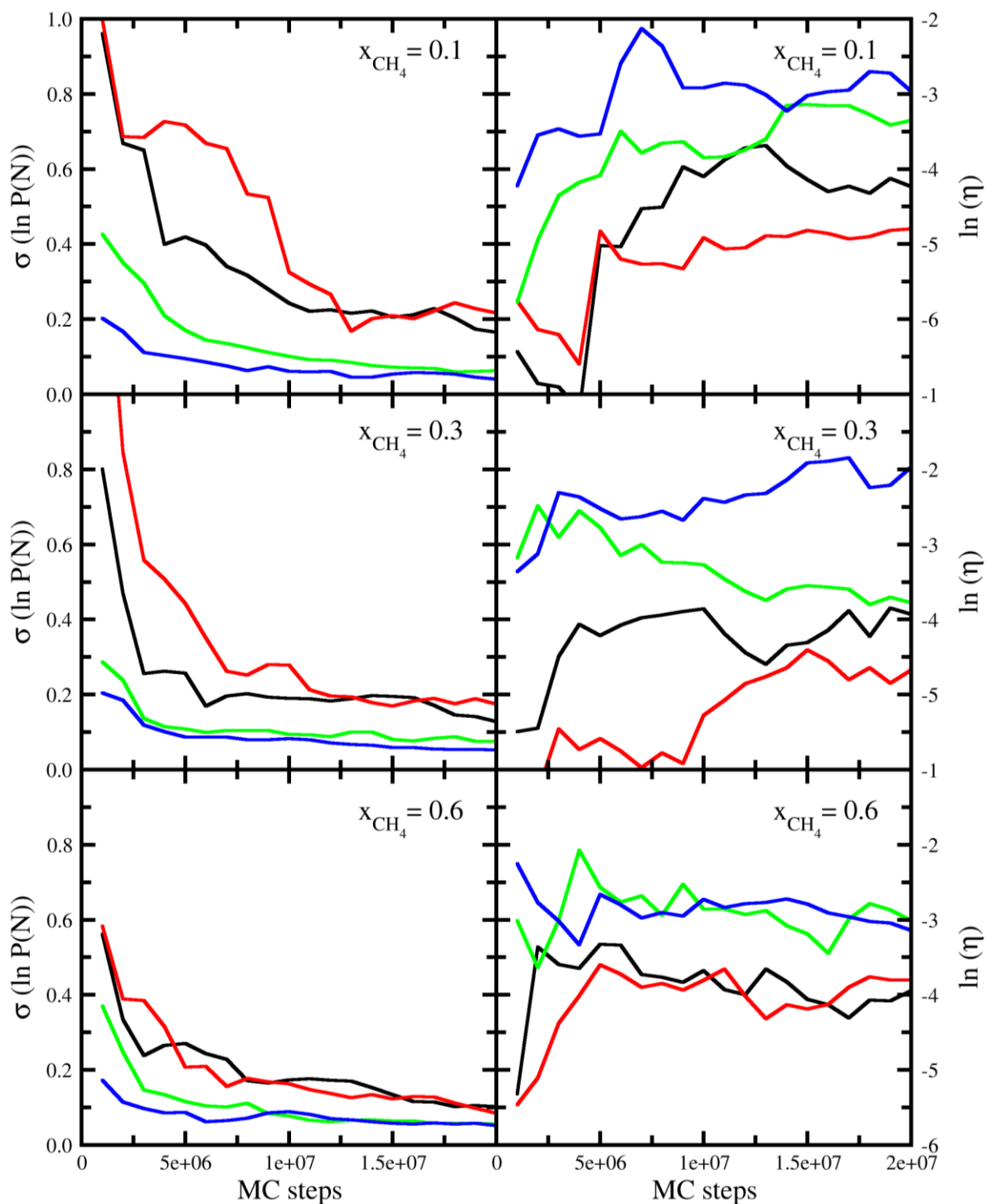
Figure 4.6: Efficiency and standard deviation in methane+n-butane at 255 K. Lines represent the efficiency and average uncertainty in probability distributions generated from GCMC simulations. Standard configurational-bias insertions (black), ME-1 (red), ME-2 (green), and ME-3 (blue). The MEMC move was performed with the exchange ratio of one butane with one methane.

In Figure 4.7, the probability distributions resulting from GCMC simulations with the various ME methods using an exchange ratio of one to one are presented for $x_{methane} = 0.3$, while data for other mole fractions are given in Appendix B, Figures B4 and B5. The probability distributions resulting from GCMC simulations with the various ME methods using an exchange ratio of one to two are presented in Appendix B, Figure B6-8 for a range of mole fractions. All MEMC methods converge to the same distribution. ME-3 shows rapid convergence, and within only $5x10^6$ MCS the correct distribution is obtained. The ME-2 algorithm shows slightly slower convergence compared to ME-3, but is still more efficient that ME-1 or configurational-bias trial insertions.
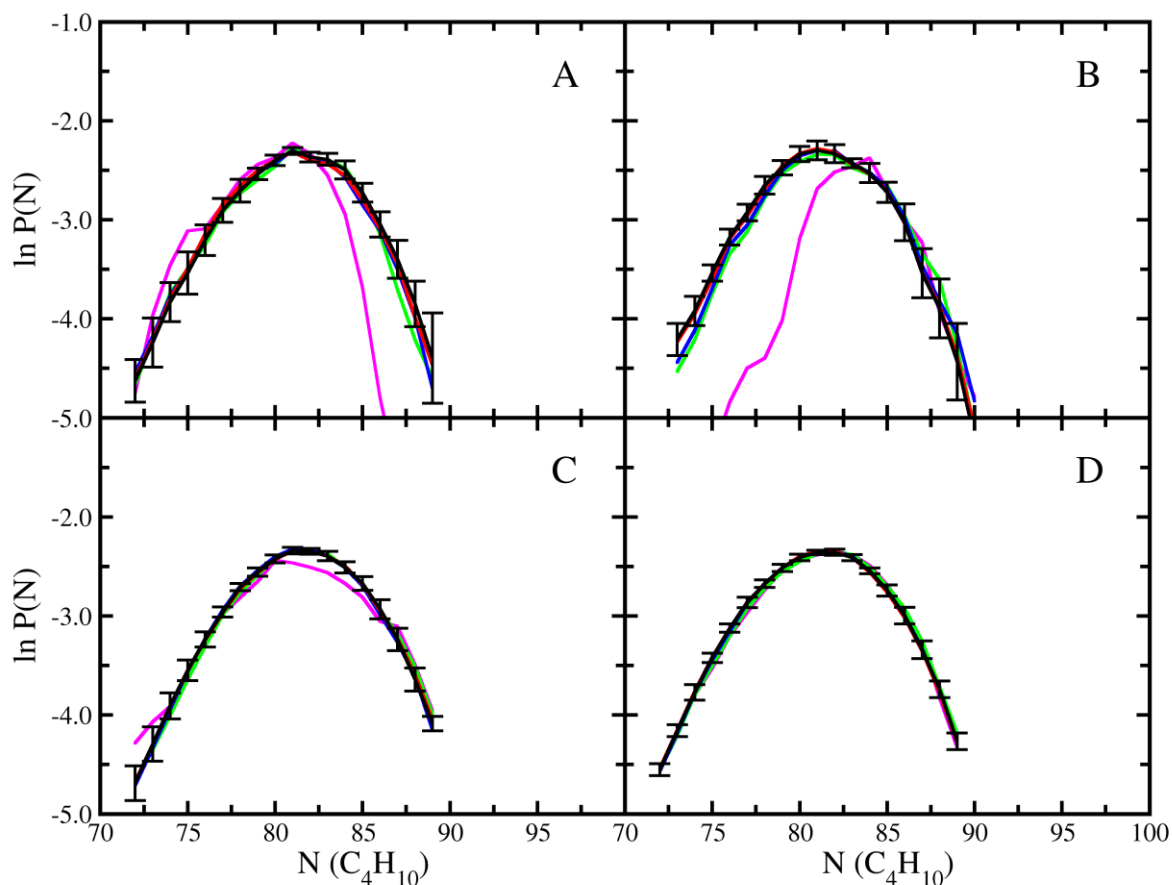
Figure 4.7: Probability distributions for methane+n-butane at 255 K and $x_{methane} = 0.3$. After simulations of: $1x10^6$ MCS (magenta), $5x10^6$ MCS (green), $1x10^7$ MCS (blue), $1.5x10^7$ MCS (red), and $2x10^7$ MCS (black) (A) Standard configurational-bias insertions, (B) ME-1 (C) ME-2 and (D) ME-3.

4.4.2 Perfluorobutane+n-butane

The perfluorobutane+n-butane system is an interesting case study because of its large deviations from Raoult's law, despite the fact that perfluorobutane and n-butane have very similar normal boiling points (270.96 K for C_4F_{10} and 272.61 K for C_4H_{10}) and both are non-polar with similar molecular geometries. This system has been modeled in the past with SAFT-VR[183], PC-SAFT[184] and GC-SAFT-VR[185], which showed close agreement with experimental data[186]. Gibbs ensemble Monte Carlo simulations using an identity exchange move have been used to study liquid-liquid equilibria for n-heptane+perfluoheptane[187], otherwise, grand canonical and Gibbs ensemble methods have rarely been applied to these kinds of mixtures. This is due, in part, to the difficulty in achieving an adequate number of accepted molecule transfers. For example, at 260 K, acceptance rates for the insertion of perfluorobutane in the neat liquid phase was approximately 0.075%.

In Figure 4.8, the pressure vs. composition diagram for perfluorobutane+n-butane at 260 K, predicted using the ME-3 algorithm and the Mie potentials developed by our group[24], is shown. The force field for perfluorobutane was modified slightly from the original work to use a more accurate seven term cosine series, which is described in detail in the Appendix B. Using standard Lorentz-Berthelot combining rules[120, 121] and no adjustable parameters for the cross interaction, very good agreement was achieved with experiment. The largest deviation results from the limitation in the united-atom force field for perfluorobutane, which over-predicts the vapor pressure at 260 K by approximately 0.1 bar.
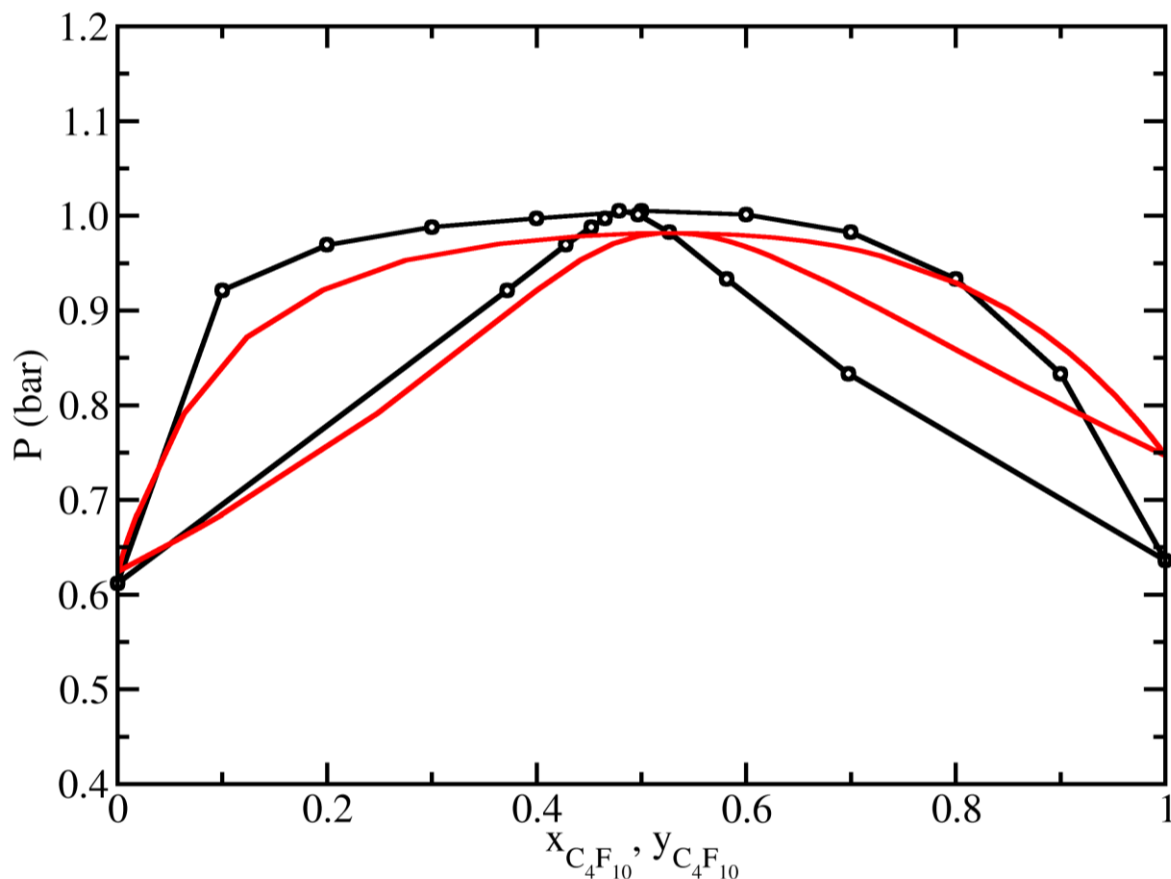
Figure 4.8: Pressure-composition diagram for perfluorobutane+n-butane at 259.95 K. The predictions from GCMC+histogram reweighting simulations using the ME-2 algorithm are given by (red line) while experiment data[186] are represented by (black circles). The line connecting the experimental data points is provided as a guide to the eye.

To evaluate the effectiveness of the molecular exchange move with a one to one exchange ratio and an exchange sub-volume of 6 Å × 6 Å × 9 Å, acceptance rates, uncertainties in the probability distributions, and efficiencies produced from the grand canonical Monte Carlo simulations were determined for liquid phase simulations at selected points along the coexistence curve. The effect of various simulation parameters on the performance of the CBMC and MEMC acceptance rates and efficiencies were also evaluated for liquid phase simulations containing 50 mol% n-butane, and are shown in Appendix B, Figure B9. Using the coupled-decoupled configurational-bias method[67], the probability of successfully inserting one perfluorobutane into a simulation box containing 10 mol%, 50 mol%, and 90 mol% of n-

butane was 0.073%, 0.026%, and 0.011%, respectively. The ME-1 algorithm increased acceptance rates approximately 4 times that of standard trial insertions for $x_{butane} > 0.50$, however, for lower concentrations of n-butane, no improvement was observed. For the ME-2 algorithm, acceptance rates of 4.92%, 4.17%, and 3.15% were obtained, while for ME-3, acceptance rates were 3.52%, 2.73 %, and 1.69%, respectively. For this system, the ME-2 algorithm produces the best acceptance rates because it works by aligning the backbone of perfluorobutane with the cavity left by the leaving n-butane. Acceptance rates were slightly lower for ME-3 since it grows the molecule using coupled-decoupled configurational-bias without requiring the backbone of the molecule to be aligned with the cavity created by the molecule that was removed.

The efficiency of the various molecular exchange algorithms is shown in Figure 4.9 as a function of Monte Carlo step for $x_{butane} = 0.1, 0.5,$ and 0.9. Uncertainties shown are the average over uncertainties for each histogram bin in the probability distribution. Both the ME-2 and ME-3 algorithms show that convergence of the probability distributions was achieved within 10 million MCS, while for ME-1 and configurational-bias insertions, convergence was not achieved within 20 million MCS. Depending on composition, ME-3 provides efficiencies that are between 12 and 200 times greater than configurational-bias insertions for the insertion of perfluorobutane. Based on the trajectory of the uncertainties, it is unlikely that convergence of the probability distributions using standard Monte Carlo insertions would ever occur. Despite the fact that the ME-2 method provides slightly better acceptance rates for the molecular exchange move, at most compositions, ME-3 produces slightly faster convergence and better efficiencies. By growing the inserted molecule with coupled-decoupled configurational-bias[67], larger rearrangements take place in the system, even though more of the trial moves are rejected than in ME-2.
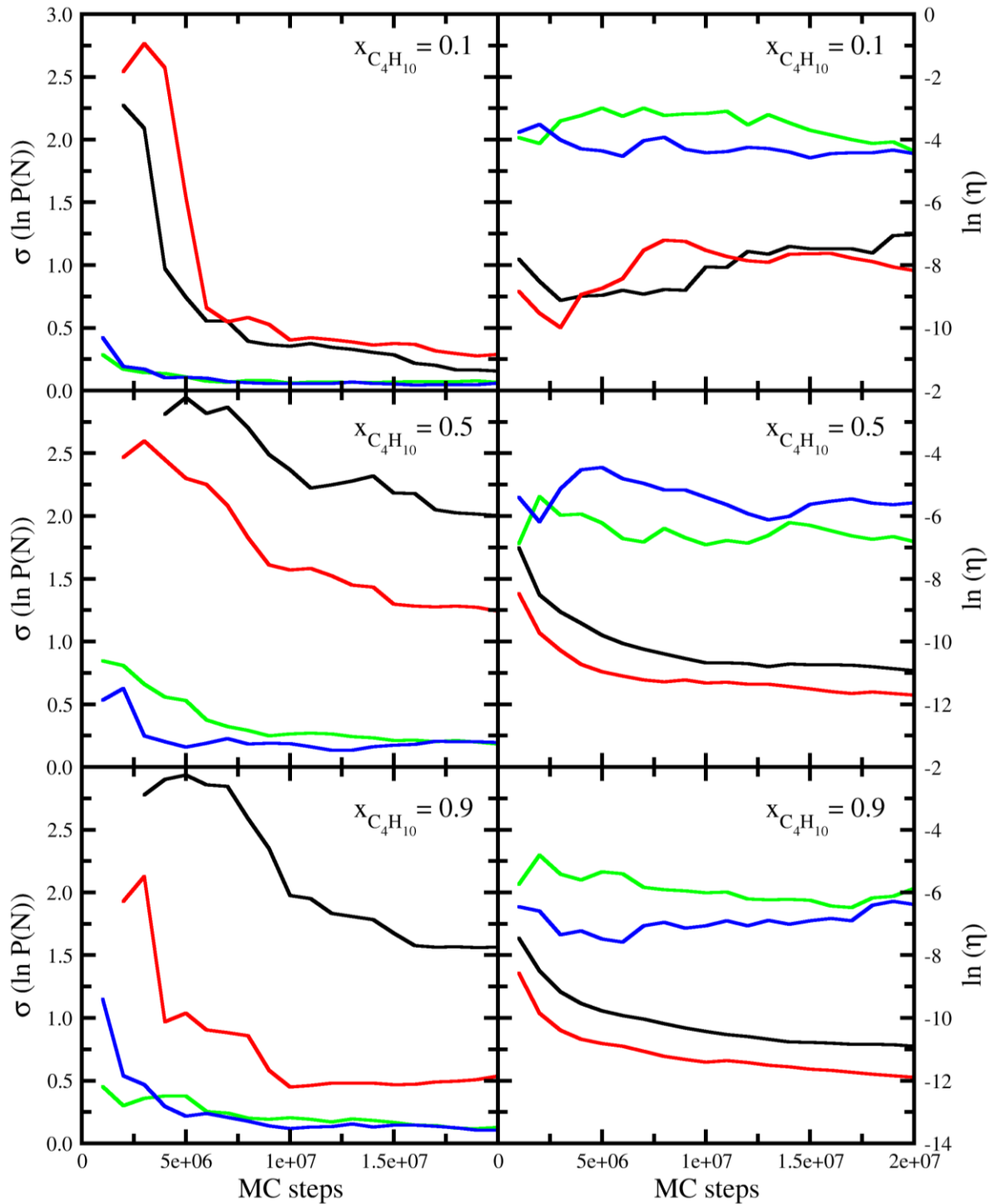
Figure 4.9: Efficiency and standard deviation in the perfluorobutane+n-butane binary mixture at 259.95 K. Lines represent the efficiency and average uncertainty in the perfluorobutane probability distribution; standard configurational-bias insertions(black), ME-1 (red), ME-2 (green), and ME-3 (blue). The MEMC moves were performed with an exchange ratio of one to one.

In Figure 4.10, the probability distributions resulting from GCMC simulations with the various ME methods are presented for $x_{butane} = 0.5$, while data for $x_{butane} = 0.1$ and 0.9 are given in Appendix B, Figures B10 and B11. The figure shows rapid convergence of the probability distributions for the ME-2 and ME-3 methods, while ME-1 and standard GCMC have not converged in 20 million MCS, although, the uncertainties calculated for ME-1 are approximately half those of standard GCMC. In Figure 4.11, heat maps are presented for the particle numbers and potential energies sampled during a liquid phase GCMC simulation. The heat maps illustrate how simulations with only configurational-bias insertions/deletions may become trapped in metastable states, resulting in poor sampling. Inclusion of the ME-3 algorithm produced a short equilibration period and a much broader sampling of the N_1, N_2, E phase space.
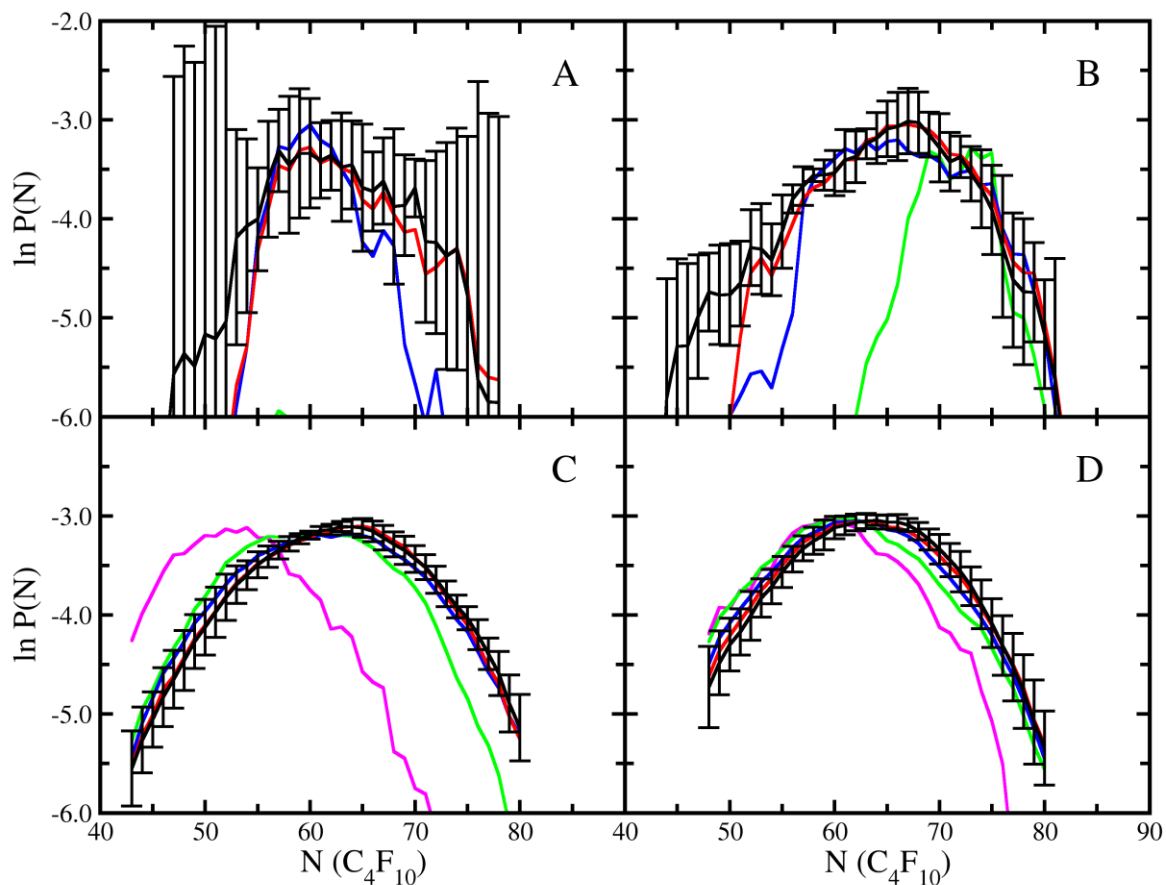
Figure 4.10: Molecule probability distribution for perfluorobutane+n-butane at $x_{butane} = 0.5$ and 259.95 K. After simulations of: 1×10^6 MCS (magenta), 5×10^6 MCS (green), 1×10^7 MCS (blue), 1.5×10^7 MCS (red), and 2×10^7 MCS (black) (A) Standard configurational-bias insertions, (B) ME-1 (C) ME-2 and (D) ME-3.
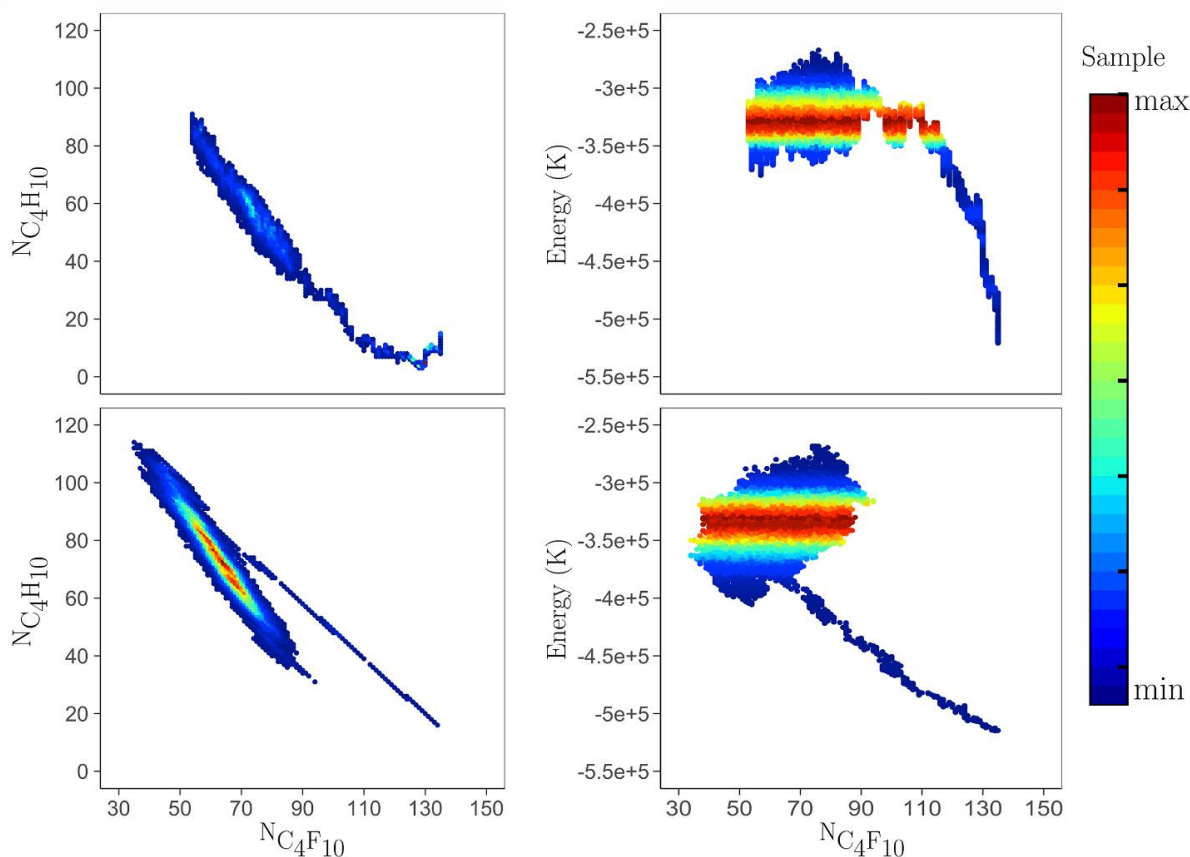
Figure 4.11: Heat maps of particle numbers (left panel) and potential energies (right panel) sampled during liquid phase grand canonical Monte Carlo simulations of perfluorobutane+n-butane at 259.95 K. Upper figures correspond to GCMC simulations with standard configurational-bias insertions/deletions, while the bottom figures correction to GCMC simulations with the ME-3 algorithm.

4.4.3 Water

In order to compare the performance of the MEMC move with other advanced sampling techniques, such as CBMC swap + identity switch[171](IS), continuous fractional component Monte Carlo (CFCMC)[172, 173], and configurational-bias continuous fractional component Monte Carlo (CB-CFCMC)[172], the vapor-liquid coexistence curve for SPC/E water[188] was predicted from the critical temperature to $0.44T_c$. To enhance the acceptance rate for insertions and deletions of water and to provide a uniform basis for comparison, the strategy of Bai and Siepmann was used[171]. For regular CBMC swaps, oxygen is inserted first, followed by the two hydrogen atoms. 16 trials were used for the first atom and 8 trials for all remaining atoms. Simulations were performed as a mixture that contained approximately 0-10

"impurity" molecules, where the impurity molecule had an identical geometry to the SPC/E water model, but with partial charges reduced by a factor of 2 and the oxygen atom Lennard-Jones epsilon reduced by a factor of 4 compared to SPC/E water. Swap moves were performed only for impurity molecules, while the MEMC move is performed to exchange the impurity with water and vice versa. Move frequencies were adjusted to yield approximately to the same number of accepted molecule transfers for the swap and MEMC moves. Due to the poor performance of the ME-1 method in prior calculations, only the performance of the ME-2 and ME-3 methods were evaluated. An exchange ratio of one to one was used for all calculations.

The phase diagram for SPC/E water predicted from GCMC simulations using the ME-2 or ME-3 algorithm is shown in Figure 4.12, with a comparison to prior simulations[189]. Additional information on vapor pressure is provided in Appendix B, Figure B12. Excellent agreement was observed, validating both the MEMC algorithms and the simulation code used to perform the calculations.
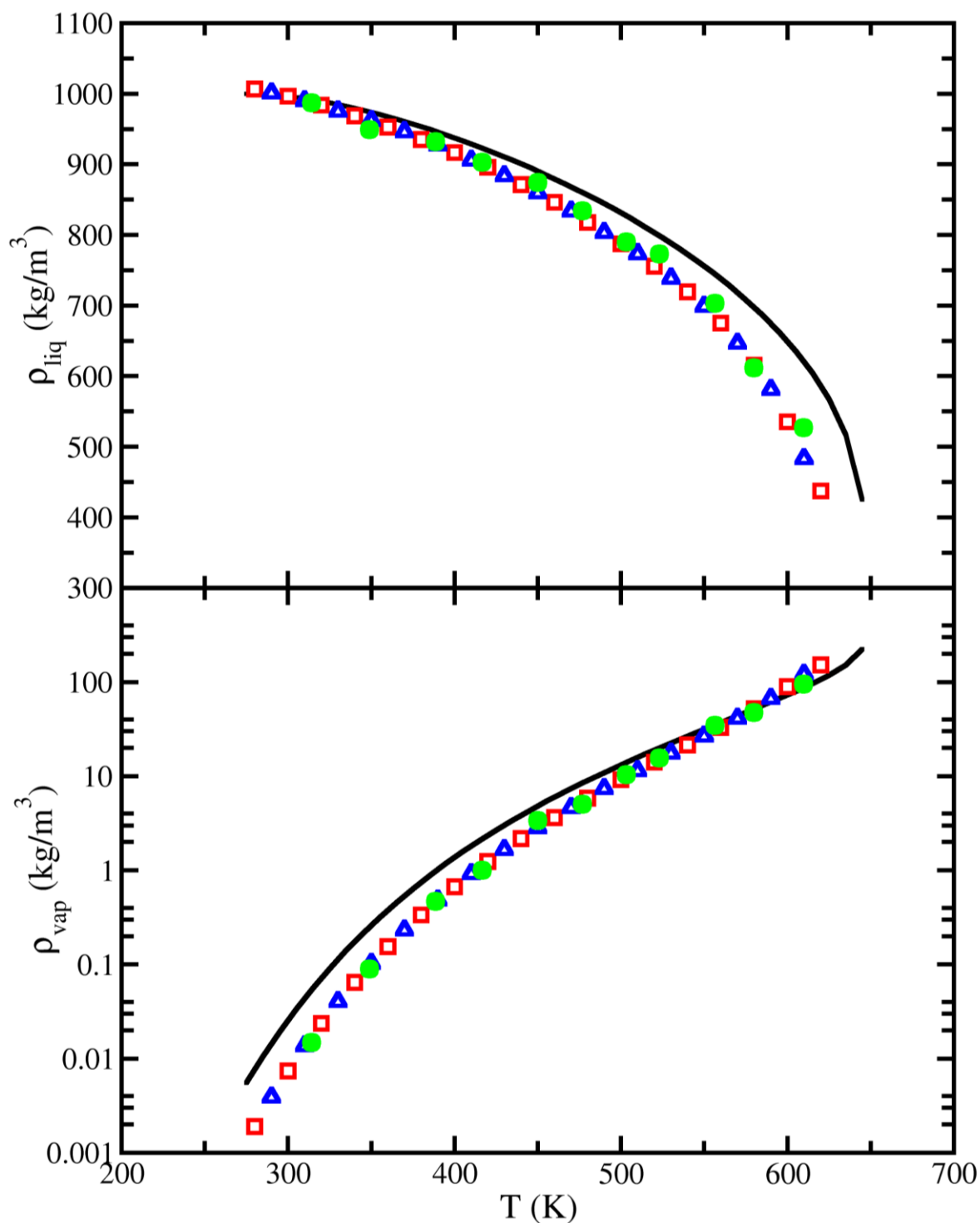
Figure 4.12: Vapor-liquid coexistence curve for SPC/E water predicted from GCMC+histogram reweighting simulations. NIST Chemistry WebBook[158] (solid lines), values obtained by Boulougouris et al.[189] (green circles), ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles).

To compare the performance of MEMC with other methods, the effective number of molecule transfers was calculated. The effective number of molecule transfers was defined as

the insertion of an impurity molecule by the swap move and its conversion to a regular water molecule by the MEMC move, or the conversion of regular water to impurity via MEMC and then deletion of impurity by the swap move. Exchanges of impurity to water and back to impurity were not counted. The effective acceptance rate was calculated from the effective number of molecule transfers divided by the sum of attempted swap and MEMC moves. The results of these calculations are summarized in Table 4.2, with comparisons to the work of Bai and Siepmann[171], and Torres-Knoop *et al*.[172]. At 283 K, the effective acceptance rates for the ME-2 and ME-3 algorithms are 7.6 and 1.4 times greater, respectively, than the IS algorithm[171]. While the S+IS method reuses atomic coordinates of the molecule to be removed, the MEMC methods perform multiple trial orientations to insert the water molecule. In ME-2, first the center of the sub-volume was placed at the geometric center of the impurity, second the z-axis of the sub-volume was aligned with the O-H bond of impurity, and then multiple rotational trials were performed around the z-axis of the sub-volume. Aligning the O-H bond of water and the sub-volume allows some of the original hydrogen bonding to be maintained, while finding an energetically favorable position for the oxygen atom through rotational trials around the z-axis of the sub-volume, leading to significant improvements in the effective acceptance. In the ME-3 method, the oxygen atom of water was placed at the geometric center of the impurity molecule, and multiple rotational trials were performed on a sphere to find the most energetically favorable position. In order to maintain the hydrogen bonding formed by the impurity molecule, a large number of rotational trials are required, leading to a significant decrease in the acceptance efficiency compared to ME-2 method.

Compared to the original CFCMC method of Shi and Maginn[173], at 280 K, the ME-2 method exhibits twice the effective acceptance rate, while the ME-3 method is approximately 40% lower. The continuous fractional component Monte Carlo (CFCMC) and configurational-bias continuous fractional component Monte Carlo (CB-CFCMC) methods of Torres-Knoop

et al.[172] produced the largest acceptance rates of all methods. At 280 K, CFCMC and CB-CFCMC had acceptance rates that were 2.25 and 3.6 times larger, respectively, than the ME-2 method.

Table 4.2: Comparison of Swap + MEMC move acceptance percentages with standard CBMC, S+IS[171], CFCMC[172, 173], and CB-CFCMC[172] for SPC/E water.

T (K)	%$P_{Imp-acc}$ (CBMC)		%$P_{Switch-acc}$			%$P_{Effective-acc}$			%$P_{water-acc}$ (CBMC)			%$P_{water-acc}$ (CFCMC)		%$P_{water-acc}$ (CB $-CFCMC$)
	This Work	Bai *et al.*	ME-2	ME-3	IS	ME-2	ME-3	S+IS	This work	Bai *et al.*	Torres-Knoop *et al.*	Shi. *et al.*	Torres-Knoop *et al.*	Torres-Knoop *et al.*
280	5.7	-	5.70	0.59	-	2.73	0.51	-	0.063	-	0.027	1.38	6.16	9.86
283	5.9	4.3	6.07	0.61	1.4	2.94	0.53	0.36	0.076	0.061	-	-	-	-
313	6.3	-	6.74	0.98	-	3.35	0.83	-	0.167	-	0.068	1.00	7.49	11.7
343	6.8	7.8	6.61	1.10	3.1	3.28	0.91	0.73	0.35	0.37	-	-	-	-
348	7.0	-	6.47	1.28	-	2.94	1.07	-	0.423	-	0.155	2.18	9.52	14.93
375	9.8	-	8.67	2.11	-	4.55	1.71	-	0.761	-	0.286	-	10.14	16.53
473	20.5	22	14.84	6.31	7.3	8.48	4.84	2.2	3.989	3.5	1.374	1.98	15.17	21.82
500	23	-	15.95	7.49	-	9.29	5.62	-	5.556	-	1.964	-	15.23	21.5

The acceptance efficiency was defined as the effective number of molecules transferred, divided by the total CPU time spent on swap and MEMC moves. In order to have a fair comparison between the acceptance efficiency of MEMC and S+IS, CFCMC, and CB-CFCMC methods, this quantity was normalized with respect to the acceptance efficiency of the standard CBMC method, minimizing the impact of CPU choice on the relative performance of the algorithms. The results of these calculations are listed in Table 4.3. At 280 K, the ME-2 method outperformed S+IS by 3.8 times, while the S+IS method is 23.9% better than ME-3. The performance of CFCMC and CB-CFCMC is 5-6 times greater than ME-2, although, it should be noted that the acceptance rates reported by Torres-Knoop *et al.* for standard swaps of water were approximately 2.4 times lower than those reported in this work, or Bai and Siepmann[171].

Table 4.3: Comparison of relative acceptance efficiency for the MEMC, S+IS[171], CFCMC[172] and CB-CFCMC[172] methods.

T (K)	ME-2	ME-3	S+IS	CFCMC[172]	CB-CFCMC[172]
280	38.8	7.61	-	243.47	195.28
283	34.1	6.49	10	-	-
313	19.33	4.91	-	97.07	85.27
343	11.04	3.32	3.45	-	-
348	7.97	3.02	-	52.18	42.69
375	6.39	2.47	-	33.16	27.59
473	2.08	1.25	1.23	7.74	6.85
500	1.65	1.04	-	6.52	5.18

4.4.4 2,2,4-Trimethylpentane

As mentioned earlier, achieving a statistically valid number of molecule insertions in low temperature $(T < 0.7T_c)$ simulations of branched alkanes can be challenging. Here, 2,2,4-trimethylpentane is used as an example to highlight how the MEMC move can significantly extend the range of temperatures where GCMC simulations may be used to predict vapor-liquid coexistence for a highly branched molecule. In this case, neopentane is used as the impurity molecule based on its similar structure to part of 2, 2, 4-trimethylpentane. This also illustrates the general nature of the MEMC algorithm, which does not require the molecules to be exchanged to be an integer numbers of each other. In Figure 4.13, the vapor-liquid coexistence curve for 2,2,4-trimethylpentane, using ME-2 algorithm and GCMC+histogram reweighting Monte Carlo simulations, is presented. Additional data for the ME-3 algorithms is presented in Appendix B, Figure B13. Using the ME-2 or ME-3 algorithms, it is possible to predict vapor-liquid coexistence to 280 K $(0.51T_c)$, while prior simulations using only coupled-decoupled configurational bias were limited to 390 K $(0.7T_c)$. In Table 4.4, a detailed comparison is presented for the acceptance rates for direct swaps of neopentane and 2,2,4-trimethylpentane, MEMC moves, effective acceptance rates and effective acceptance rates per CPU time. Effective acceptance rate and acceptance efficiency is calculated using a similar method explained in the water section. The results of additional calculations performed with different

CBMC parameters are given in Appendix B, Table B6. At all temperatures, the combination of impurity swap plus ME-2 or ME-3 method outperforms standard configurational-bias Monte Carlo. At 280 K, the relative acceptance efficiency (impurity swap+MEMC/standard CBMC) was 409 for ME-2 and 154 for ME-3. ME-2 is more effective than ME-3 for branched molecules because it inserts the entire molecule at the same time and aligns the backbone of the molecule to be inserted with the backbone of the molecule to be removed. ME-3 regrows the entire molecule using coupled-decoupled CBMC, however, many of these regrowths fail because they are unable to satisfy the internal molecular constraints due to the bond bending and torsional potentials[190]. In future work, it may be possible to improve the performance of the ME-3 algorithm for branched molecules by inclusion of the Jacobian-Gaussian scheme[191] for generating bending angle trials in the CBMC growth.

Table 4.4: Comparison of acceptance rates for swaps of the impurity molecule (neopentane), identity exchange via the MEMC algorithm, and swaps performed with standard configurational-bias Monte Carlo for 2,2,4-trimethylpentane.

T (K)	$\%P_{Imp-acc}$	$\%P_{Switch-acc}$		$\%P_{Effective-acc}$		$\%P_{acc}$	Effective acceptance per CPU time (s^{-1})			Relative acceptance efficiency	
	swap	ME-2	ME-3	ME-2	ME-3	CBMC	CBMC	ME-2	ME-3	ME-2	ME-3
280	0.013	0.89	0.03	0.013	0.008	0.00008	0.0003	0.109	0.041	409.2	153.7
330	0.10	2.21	0.15	0.096	0.057	0.0008	0.0026	0.917	0.288	356.9	112.0
390	0.85	5.69	0.55	0.653	0.274	0.022	0.0769	5.727	1.135	74.5	14.8
450	4.09	9.84	1.27	2.645	0.837	0.225	0.838	24.12	3.497	28.8	4.17
510	13.50	21.07	2.89	6.613	1.894	1.026	4.120	55.74	7.210	13.5	1.75
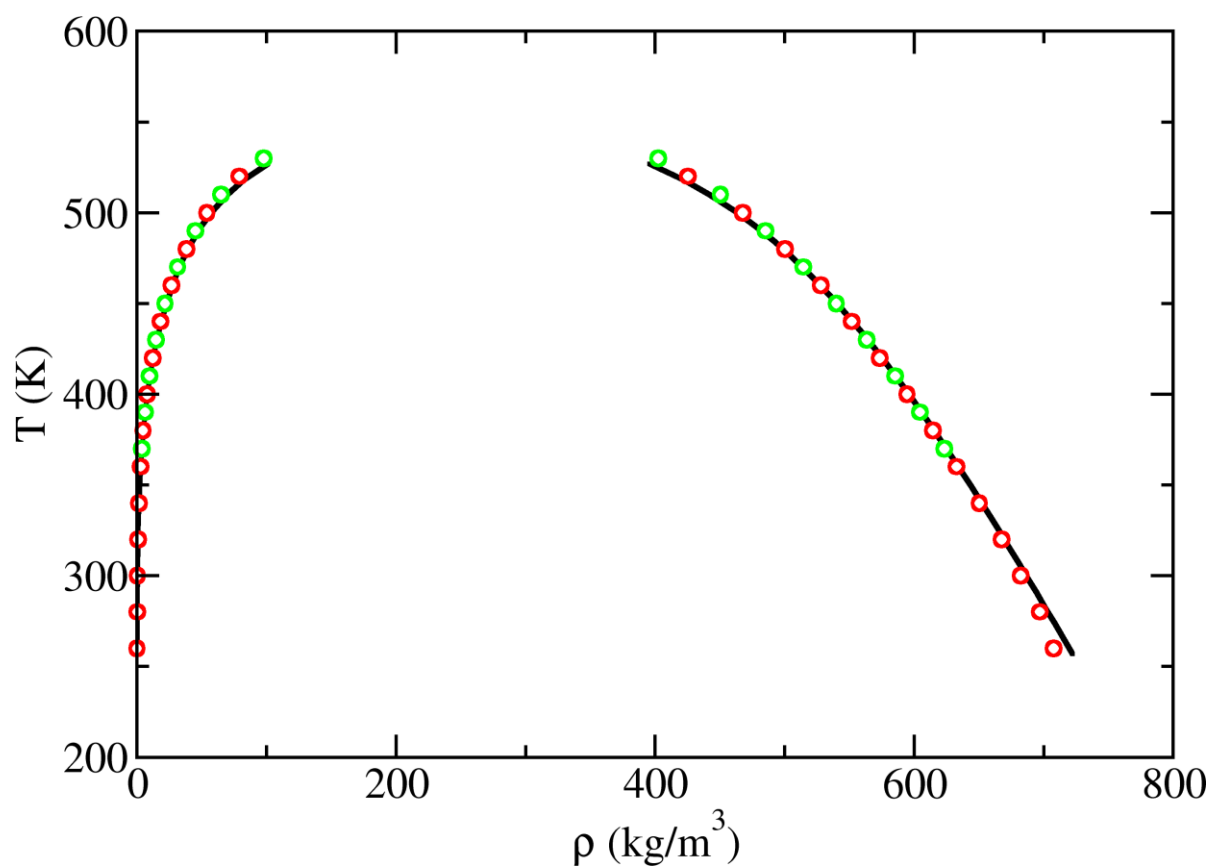
Figure 4.13: Vapor-liquid coexistence curve for 2,2,4-trimethylpentane predicted from GCMC+histogram reweighting simulations using Mie potentials[124]. Experimental data (solid lines)[151], ME-2 algorithm (red circles), and prior calculations using only configurational-bias Monte Carlo (green circles)[124].

CHAPTER 5 PREDICTION OF PHASE EQUILIBRIA AND GIBBS FREE ENERGIES OF TRANSFER USING MOLECULAR EXCHANGE MONTE CARLO IN THE GIBBS ENSEMBLE

5.1   Introduction

The Gibbs ensemble Monte Carlo (GEMC) method, developed by Panagiotopoulos[32], provides a robust, direct method for the calculation of phase equilibria from molecular simulation. The Gibbs ensemble method uses two simulation boxes that represent a sample taken deep from within each phase. These boxes are in thermodynamic contact, and the algorithm relies on three basic movements to achieve equilibrium. These moves are configurational or conformational movement within a cell (thermal equilibration), the transfer of molecules between boxes (chemical equilibrium) and the transfer of volume from one box to the other (mechanical equilibrium).

While originally proposed as a means to simulate the vapor-liquid equilibria of Lennard-Jones spheres[32, 167], numerous advances in Monte Carlo sampling methodology for the molecule swap move have been developed since its introduction, enabling GEMC to be used on increasingly complex and difficult to sample systems. The introduction of configurational-bias Monte Carlo (CBMC) sampling techniques to GEMC[192] enabled the simulation of VLE for n-alkanes up to C48[193]. By decoupling the various intra- and inter-molecular degrees of freedom, the coupled-decoupled configurational-bias method allowed for the efficient simulations of highly branched molecules[67]. Work by Martin[190], and Sepehri *et al*.[194, 195], focused on improving the success rate for molecule growths by using smarter methods for sampling intramolecular degrees of freedom. Methods such as, reservoir[164, 196, 197], rebridging configurational-bias[198], and self-adapting fixed end-point Monte Carlo[199] have been created that allow rings to be exchanged between simulation boxes. For systems with high densities and/or strong electrostatic interactions, which may preclude

successful insertions of molecules via bead-by-bead growth, expanded ensemble methods have been developed[71, 172, 173, 200-202], where a molecule is slowly deflated in one phase and inflated in another. Martin and Siepmann suggested that by swapping the identity of molecules between phases (a "swatch" move), significant improvements in sampling could be achieved in mixtures[77]. This technique has been used in a number of studies, such as: the simulation of water[171], liquid-liquid equilibria for hexane-perfluorohexane[187], and CO_2-polymer phase behavior[79]. In previous work, our group introduced a variation of the combined swap+identity exchange (swatch) move called molecular exchange Monte Carlo (MEMC) for grand canonical Monte Carlo simulations[83]. MEMC can be thought of as a generalized version of Siepmann's swatch move, where the molecules to be exchanged do not have to share any common atom types or coordinates.

In this Chapter, MEMC methods are presented for simulations in the Gibbs ensemble. A derivation of acceptance criteria and the algorithms for performing the MEMC move in GEMC are provided in the next section. The simulation details for determining the binary mixture phase diagrams and Gibbs free energies of transfer are provided in Simulation Methodology. In the Results and Discussion, the MEMC algorithm is validated with predictions of the methane+n-butane and n-butane+perfluorobutane pressure-composition diagrams, and free energies of transfer for n-alkanes in 1-octanol, hexadecane and 2,2,4-trimethylpentane. The key findings of the work are summarized in the Chapter 7. Additional results with their numerical values are provided in Appendix C. The result of this work has been published in Journal of Fluid Phase Equilibria[84].

5.2 Methods

In this Chapter, the molecular exchange Monte Carlo (MEMC) method, originally developed in the context of the grand canonical ensemble, is extended to Gibbs ensemble Monte Carlo. To describe the MEMC move in the Gibbs ensemble, box 1 is assumed to be the

higher density liquid phase, and box 2 is assumed to be the lower density gas phase. Attempts are made to exchange a large molecule with multiple smaller molecules in the dense phase (box 1).

For a given configuration, with N_1^L, N_2^L large and N_1^S, N_2^S small molecules in box 1 with volume V_1 and box 2 with volume of V_2, a "*deletion move*" is an attempt to remove one large molecule and insert N_{EX} small molecules inside a predefined exchange sub-volume V_{EX} in box 1. An "*insertion move*" is an attempt to remove N_{EX} small molecules, and insert one large molecule in box 1. The exchange sub-volume is defined as an orthogonal box, where the length of the box in the x-,y-, and z-dimensions can be set independently, however, in this work x=y, while z is set independently. An orthogonal sub-volume is used instead of a cube or sphere to accommodate large molecules with different aspect ratios. A heuristic for setting good values of the x-, y-, and z-dimensions is to use the geometric size of the large molecule plus 1-2 Å in each dimension. Although not used in this work, it is also possible to optimize N_{EX} and V_{EX} "on the fly" during a simulation to maximize the acceptance rate.

The acceptance criterion for a molecular exchange move that satisfies the detailed balance equation is written as

$$K(old \rightarrow new) = K(new \rightarrow old) \tag{5.1}$$

where $K(i \rightarrow j)$ is the flux of probability from state i to state j. The probability flux is equal to the product of the probability of finding the system in state i, the probability of generating a move that takes state i to state j, and the probability of accepting the move:

$$K(old \rightarrow new) = \mathcal{N}(old) \times \alpha(old \rightarrow new) \times acc(old \rightarrow new) \tag{5.2}$$

where, $\mathcal{N}(old)$ is the probability of finding the system in state *old*, $\alpha(old \rightarrow new)$ is the probability of generating a move that takes the system from state *old* to state *new*, and $acc(old \rightarrow new)$ is the probability of accepting the move that takes the system from state *old*

to state *new*. Based on the detailed balance Eq. 5.1, the ratio of the probability of accepting the move from $old \to new$ to that of its reverse move $new \to old$ is:

$$\frac{acc(old \to new)}{acc(new \to old)} = \frac{\mathcal{N}(new)}{\mathcal{N}(old)} \times \frac{\alpha(new \to old)}{\alpha(old \to new)} \tag{5.3}$$

The Gibbs ensemble partition function for N distinguishable molecules and regular Cartesian (unscaled) coordinates is

$$Q_G(N,V,T) = \frac{1}{\Lambda^{3N}} \sum_{n_1=0}^{N} \int_0^V \left[\left[\int_{v_1} exp\left[-\beta U\left(r_1^{n_1}\right)\right] dr_1^{n_1} \right] \right.$$
$$\left. \times \left[\int_{V-v_1} exp\left[-\beta U\left(r_2^{N-n_1}\right)\right] dr_2^{N-n_1} \right] dv_1 \right. \tag{5.4}$$

where, Λ is the thermal de Broglie wavelength, $N = n_1 + n_2$ is the total number of molecules in the system, n_1 is the number of molecules in box 1, n_2 is the number of molecules in box 2, $V = v_1 + v_2$ is to the total volume of the system, v_1 is the volume of box 1, v_2 is the volume of box 2, and r_b^i represents the coordinates of molecule i, in box b. The probability of finding a configuration with n_1 molecules in box 1 with volume v_1 and specific positions $r_1^{n_1}$ and $r_2^{N-n_1}$ is

$$\mathcal{N}\left(n_1, v_1, r_1^{n_1}, r_2^{N-n_1}\right) \propto exp\left\{-\beta\left[U\left(r_1^{n_1}\right) + U\left(r_2^{N-n_1}\right)\right]\right\} \tag{5.5}$$

In both the insertion and deletion moves, the ratio of the probability of being in the configuration *new* to the probability of being in the configuration *old* is simplified to

$$\frac{\mathcal{N}(new)}{\mathcal{N}(old)} = \frac{e^{-\beta(U_1(new)+U_2(new))}}{e^{-\beta(U_1(old)+U_2(old))}} \tag{5.6}$$

where $\beta = 1/k_B T$, $U_1(old)$, $U_2(old)$, $U_1(new)$, and $U_2(new)$ are the potential energies of the system in configuration *old* and configuration *new* in box 1 and box 2, respectively.

The probability of generating the *new* state, for both insertion and deletion of the large molecule, is given by the product of the probability of locating the center of the exchange sub-

volume at a particular point within the simulation box 1, the probability of choosing N_{EX} particular small molecules, the probability of choosing a particular large molecule, the probability of generating trial configurations for N_{EX} small molecules, and the probability of generating trial configurations for the large molecule,

$$\alpha(old \rightarrow new) = P_{sub-v}(old \rightarrow new) \times P_{pick-S}(old \rightarrow new) \times$$

$$P_{pick-L}(old \rightarrow new) \times P_{pos-S}(old \rightarrow new) \times P_{pos-L}(old \rightarrow new)$$

(5.7)

5.2.1 ME-2

For the large molecule insertion move, the geometric center of V_{EX} is placed on the centroid of a randomly selected small molecule in box 1. If the small molecule is monoatomic, the orientation of V_{EX} is assigned randomly, otherwise its z-axis is aligned with the backbone of the small molecule. For a large molecule deletion move, the geometric center of V_{EX} is located at the centroid of the selected large molecule in box 1 and its z-axis is aligned with the backbone of the large molecule. To improve acceptance rates for the MEMC move, multiple trial positions (j) and orientations (k) are performed.

Insertion of large molecule into box 1: The algorithm for the insertion of a large molecule into box 1 after the deletion of small molecule(s) is identical to ME-2 method described previously for grand canonical Monte Carlo[83]. Resolving the move requires accounting for the removal of the large molecule from box 2 and the insertion of small molecule(s) into box 2. An illustration of the large molecule insertion into box 1 in ME-2 algorithm is provided in Figure 5.1.
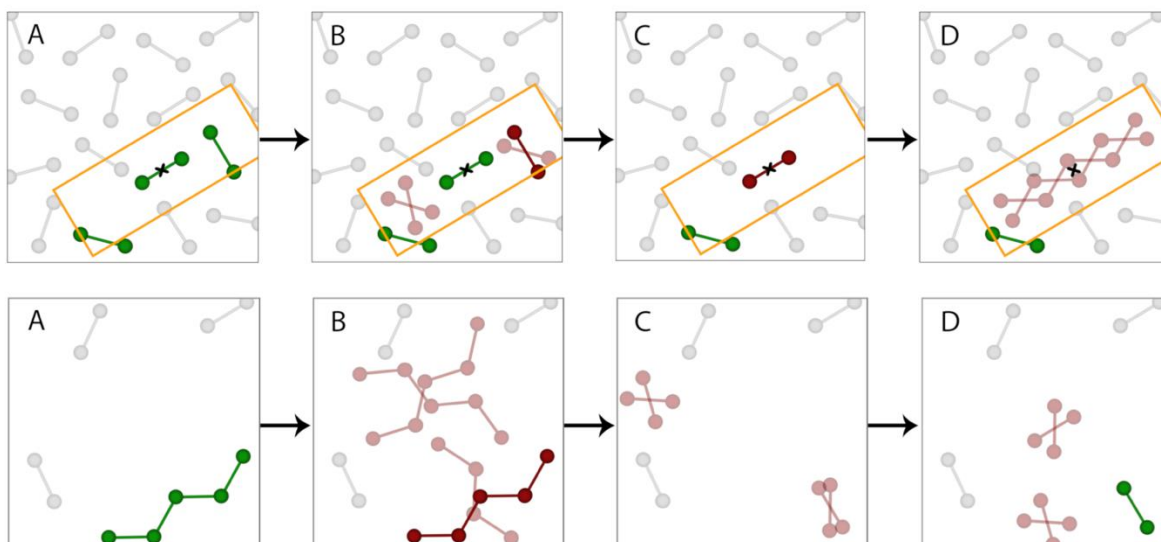
Figure 5.1: Schematic for the ME-2 algorithm for the transfer of a large molecule from box 2 (gas phase) into box 1 (liquid phase), and corresponding transfer of small molecules from box 1 into box 2. Selected or inserted molecule (green), trial position (light red), and actual position of the molecule (solid red). **Top row**, represents the exchange of two small molecules with one large molecule in box 1. The sub-volume is defined by the orange box. (A) Aligning the sub-volume z-axis with the backbone of a randomly selected small molecule, with geometric center placed at centroid of the selected small molecule, identifying the small molecules within the sub-volume, and randomly picking one small molecule for transfer. (B) Generating CBMC trials (3D rotation and centroid location) for the second small molecule, and then removing it. (C) Generating CBMC 2D rotational trials around the z-axis of the sub-volume for the first small molecule and then removing it. (D) Placing the large molecule's centroid at the geometric center of the sub-volume, aligning the backbone of the large molecule with the sub-volume z-axis, performing CBMC 2D rotational trials around the z-axis of the sub-volume, and inserting it to the sub-volume. **Bottom row**, represents the exchange of one large molecule with two small molecules in box 2. (A) Selecting a random large molecule. (B) Generating CBMC trials (3D rotation and centroid location) for the selected large molecules and then removing it. (C) Generating CBMC trials (3D rotation and centroid location) for the first small molecules and then inserting it. (D) Generating CBMC trials (3D rotation and centroid location) for the second small molecule and then inserting it.

The algorithm for doing this follows:

1. Select a large molecule out of N_2^L large molecules within the simulation box 2 with the probability of $1/N_2^L$.

2. Generate $j - 1$ random trial positions for the centroid of the selected large molecule within simulation box 2 (V_2). The original position of the centroid of the large molecule will be included as the jth term.

3. For each trial position p, generate k random trial orientations around the large molecule's centroid (except the jth centroid, where $k-1$ random trial orientations are generated, and the original orientation of the molecule will be included as the kth term). Trial orientations are generated keeping all internal degrees of freedom for the molecule fixed. The Rosenbluth weight is calculated as $W_{old}^L = \sum_{p=1}^{j} \sum_{r=1}^{k} exp(-\beta U_2^{p,r})$, where $U_2^{p,r}$ is the interaction energy of the large molecule in position p and orientation r with all other molecules in the simulation box 2.

4. Calculate the probability $P_{old}^L = \frac{exp(-\beta U_2^{j,k})}{W_{old}^L}$, where $U_2^{j,k}$ is the interaction energy of the large molecule at the original position and orientation with all other molecules in the simulation box 2. P_{old}^L is the probability of inserting the large molecule at its original configuration in the reverse move ($new \rightarrow old$). Then remove the large molecule from simulation box 2.

5. Repeat steps a→c for N_{EX} cycles ($i = 1, 2, ..., N_{EX}$) to insert the selected small molecules in box 2 with the probability of $N_{EX}!/V_2^{N_{EX}}$.

    a. Generate j random trial positions for the centroid of the ith small molecule within simulation box 2 (V_2).

    b. For each trial position, p, generate k random trial orientations around the molecule's centroid, and calculate the Rosenbluth weight $W_{i,new}^S = \sum_{p=1}^{j} \sum_{r=1}^{k} exp(-\beta U_2^{i,p,r})$, where $U_2^{i,p,r}$ is the interaction energy of the ith inserted small molecule at position p and orientation r with all the other molecules in box 2, including those added in the earlier cycles of the move.

c. Pick one of the generated trial configurations with the probability $P_{i,new}^S = \dfrac{exp\left(-\beta U_2^{i,p,r}\right)}{W_{i,new}^S}$

and insert the small molecule.

Deletion of large molecule from box 1: The algorithm for the deletion of a large molecule and subsequent insertion of small molecule(s) in box 1 is identical to ME-2 method described previously for simulations in the grand canonical ensemble[83]. Resolving the move requires accounting for the removal of small molecule(s) from box 2 and the insertion of the large molecule into box 2. An illustration of the large molecule deletion from box 1 in ME-2 algorithm is provided in Figure 5.2.
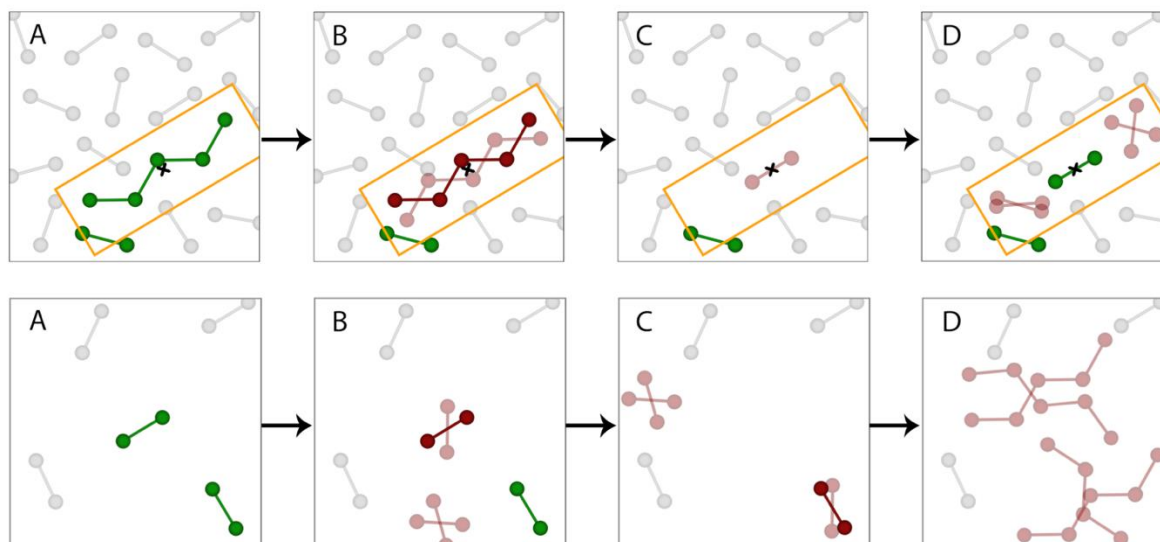
Figure 5.2: Schematic for the ME-2 algorithm for transfer of a large molecule from box 1 (liquid phase) into box 2 (gas phase) and transferring two small molecules from box 2 into box 1.. Selected or inserted molecule (green), trial position (light red), and actual position of the molecule (solid red). **Top row**, represents the exchange of one large molecule with two small molecules in box 1. The sub-volume is defined as the orange box. (A) Aligning the sub-volume with the backbone of the large molecule with geometric center placed at centroid of the large molecule and identifying the small molecules within the sub-volume. (B) Generating CBMC 2D rotational trials around the z-axis of the sub-volume and then removing the large molecule. (C) Placing the centroid of the first small molecule at the geometric center of the sub-volume, aligning the backbone of the small molecule with the z-axis of the sub-volume, generate the CBMC 2D rotational trials around the z-axis of the sub-volume, and then inserting it into the sub-volume. (D) Generating CBMC trials (3D rotation and centroid location) for the second small molecule and then inserting it into the sub-volume. **Bottom row**, represents the exchange of two small molecules with one large molecule in box 2. (A) Selecting two random small molecules. (B) Generating CBMC trials (3D rotation and centroid location) for the first small molecule and then removing it. (C) Generating CBMC trials (3D rotation and centroid location) for the second small molecule and then removing it. (D) Generating CBMC trials (3D rotation and centroid location) for the large molecules and then inserting it.

The algorithm for doing this follows:

1. Select N_{EX} small molecule(s) out of N_2^S small molecules in the simulation box 2 with the probability of $N_{EX}! (N_2^S - N_{EX})!/N_2^S!$.

2. Repeat steps a and b for N_{EX} cycles ($i = 1, 2, ..., N_{EX}$) to delete the selected small molecules from simulation box 2.

a. Generate $j - 1$ random trial positions for the centroid of the ith small molecule within simulation box 2 (V_2). The original position of the centroid of the ith small molecule will be included as the jth term.

b. For each trial centroid position p, generate k random trial orientations around the molecule's centroid (except the jth centroid, where $k - 1$ random trial orientations are generated and the original orientation of the molecule will be included as the kth term), and calculate the Rosenbluth weight $W_{i,old}^S = \sum_{p=1}^{j} \sum_{r=1}^{k} exp(-\beta U_2^{i,p,r})$, where $U_2^{i,p,r}$ is the interaction energy of the ith molecule to be removed in position p and orientation r with all other molecules in box 2, excluding those removed in the earlier cycles of the move. Finally, remove the molecule from simulation box 2. Calculate $P_{i,old}^S = \frac{exp(-\beta U_2^{i,j,k})}{W_{i,old}^S}$, where $U_2^{i,j,k}$ is the interaction energy of the ith small molecule at its original centroid position and orientation with all other molecules remaining in the simulation box 2. $P_{i,old}^S$ is the probability of inserting the ith small molecule back in its original configuration in the reverse move ($new \rightarrow old$).

3. Generate j random trial positions for the centroid of the selected large molecule within simulation box 2 (V_2). For each trial position p, generate k random trial orientations around the large molecule's centroid.

4. Calculate the Rosenbluth weight $W_{new}^L = \sum_{p=1}^{j} \sum_{r=1}^{k} exp(-\beta U_2^{p,r})$, where $U_2^{p,r}$ is the interaction energy of the inserted large molecule in position p and orientation r with all other molecules in simulation box 2.

5. Select one of the generated trial configurations with the probability $P_{new}^L = \frac{exp(-\beta U_2^{p,r})}{W_{new}^L}$ and insert the large molecule.

Based on the two algorithms described above, for the large molecule insertion move, the ratio of the probabilities for generating the move $new\ (N_1^L + 1, N_1^S - N_{EX};\ N_2^L - 1, N_2^S + N_{EX}) \rightarrow old\ (N_1^L, N_1^S; N_2^L, N_2^S)$ to that of the reverse move is:

$$\frac{\alpha(new \rightarrow old)}{\alpha(old \rightarrow new)} = \frac{\frac{1}{N_1^L + 1} \times \frac{N_{EX}! N_2^S!}{(N_2^S + N_{EX})!}}{\frac{1}{N_1^S} \times \frac{(N_{EX} - 1)! (N_{S,VEX} - N_{EX})!}{(N_{S,VEX} - 1)!} \times \frac{1}{N_2^L}} \times \frac{\frac{1}{V_2} \times \frac{(N_{EX} - 1)!}{V_{EX}^{N_{EX}-1}}}{\frac{N_{EX}!}{V_2^{N_{EX}}}} \times \prod_{i=1}^{N_{EX}} \left(\frac{P_{i,old}^S}{P_{i,new}^S}\right) \times \frac{P_{old}^L}{P_{new}^L} \qquad (5.8)$$

where, $N_{S,VEX}$ is the number of small kind molecules found within the the exchange sub-volume (V_{EX}). Simplifying Eq. 5.8 and substituting into Eq. 5.3, produces the acceptance criteria for the large molecule insertion and small molecule(s) deletion in box 1.

$$acc(old \rightarrow new) = min\left\{1\ , \frac{N_2^L N_1^S}{N_1^L + 1} \times \frac{(N_{S,VEX} - 1)! N_2^S!}{(N_2^S + N_{EX})! (N_{S,VEX} - N_{EX})!} \times \left(\frac{V_2}{V_{EX}}\right)^{N_{EX}-1} \times \prod_{i=1}^{N_{EX}} \left(\frac{W_{i,new}^S}{W_{i,old}^S}\right) \times \frac{W_{new}^L}{W_{old}^L}\right\} \qquad (5.9)$$

For the large molecule deletion move, the ratio of the probabilities for generating the move $new\ (N_1^L - 1, N_1^S + N_{EX};\ N_2^L + 1, N_2^S - N_{EX}) \rightarrow old\ (N_1^L, N_1^S; N_2^L, N_2^S)$ to that of the reverse move is:

$$\frac{\alpha(new \rightarrow old)}{\alpha(old \rightarrow new)} = \frac{\frac{1}{N_2^L + 1} \times \frac{1}{N_1^S + N_{EX}} \times \frac{(N_{EX} - 1)! N_{S,VEX}!}{(N_{S,VEX} + N_{EX} - 1)!}}{\frac{1}{N_1^L} \times \frac{N_{EX}! (N_2^S - N_{EX})!}{N_2^S!}} \times \frac{\frac{N_{EX}!}{V_2^{N_{EX}}}}{\frac{1}{V_2} \times \frac{(N_{EX} - 1)!}{V_{EX}^{N_{EX}-1}}} \times \prod_{i=1}^{N_{EX}} \left(\frac{P_{i,old}^S}{P_{i,new}^S}\right) \times \frac{P_{old}^L}{P_{new}^L} \qquad (5.10)$$

Simplifying Eq. 5.10 and substituting into Eq. 5.3, produces the acceptance criteria for the large molecule deletion and small molecule(s) insertion in box 1.

$$acc(old \rightarrow new) = min\left\{1\ , \frac{N_1^L}{(N_2^L + 1)(N_1^S + N_{EX})} \times \frac{N_{S,VEX}! N_2^S!}{(N_2^S - N_{EX})! (N_{S,VEX} + N_{EX} - 1)!} \times \left(\frac{V_{EX}}{V_2}\right)^{N_{EX}-1}\right.$$

$$\left. \times \prod_{i=1}^{N_{EX}} \left(\frac{W_{i,new}^S}{W_{i,old}^S}\right) \times \frac{W_{new}^L}{W_{old}^L}\right\} \qquad (5.11)$$

The energy difference between configuration new and old, does not appear directly in the acceptance criteria because their Boltzmann weight is already included in the probabilities used for selecting the position of the molecules.

For $N_{EX} = 1$, the acceptance criteria given in Eqs. 5.9 and 5.11, simplifies to that of the standard identity-exchange acceptance move[75, 168].

$$acc(N_1^L, N_1^S; N_2^L, N_2^S \to N_1^L + 1, N_1^S - 1;\ N_2^L - 1, N_2^S + 1)$$

$$= min\left\{ 1\ , \frac{N_2^L N_1^S}{(N_1^L + 1)(N_2^S + 1)} \times \frac{W_{new}^L}{W_{old}^L} \times \frac{W_{new}^S}{W_{old}^S} \right\} \tag{5.12}$$

$$acc(N_1^L, N_1^S; N_2^L, N_2^S \to N_1^L - 1, N_1^S + 1;\ N_2^L + 1, N_2^S - 1)$$

$$= min\left\{ 1\ , \frac{N_1^L N_2^S}{(N_2^L + 1)(N_1^S + 1)} \times \frac{W_{new}^L}{W_{old}^L} \times \frac{W_{new}^S}{W_{old}^S} \right\} \tag{5.13}$$

5.2.2 ME-3

The major difference between the ME-2 and ME-3 algorithms is that while ME-2 uses a rigid body insertion, in the ME-3 algorithm, the molecules to be exchanged are grown bead by bead using the coupled-decoupled configurational-bias Monte Carlo (CD-CBMC) algorithm[67]. The forward to reverse probability ratios for generating the large molecule insertion and the large molecule deletion moves are identical to those given in Eq. 5.8 and 5.10, respectively. The acceptance criteria for the ME-3 algorithm is identical to that of ME-2 given by Eq. 5.9 and 5.11. An illustration of the large molecule insertion and deletion in box 1 in ME-3 algorithm is provided in Appendix C, Figures C1 and C2, respectively.

5.3   Force Field Parameters

Calculations were performed with the Transferable Potentials for Phase Equilibria (TraPPE)[203, 204] and the Mie potentials of Potoff *et al.*[24, 124]. Both TraPPE and the Mie potentials use a similar potential function, which is presented in Eq. 2.3. Simulations of 1-octanol include electrostatic interactions that are modeled via partial charges. All non-bonded parameters used in this work are listed in Table 5.1, and were taken from their original sources without modification.

Table 5.1: Non-bonded parameters for alkanes, perfluoroalkanes, and 1-alcohols.

Model	Pseudo-atom	ε_i/k_b(K)	σ_i (Å)	q_i(e)	n_i
Mie-alkanes[24, 124]	CH$_4$	161.00	3.740	0.000	14
	CH$_3$	121.25	3.783	0.000	16
	CH$_2$	61.00	3.990	0.000	16
	CH (C$_N$ > 4)	14.00	4.700	0.000	16
	C (C$_N$ > 4)	1.20	6.200	0.000	16
Mie-perfluoroalkanes[24, 83]	CF$_3$	155.75	4.475	0.000	36
	CF$_2$	72.20	4.750	0.000	44
TraPPE-alkanes[67, 203]	CH$_4$	148.00	3.730	0.000	12
	CH$_3$	98.00	3.750	0.000	12
	CH$_2$	46.00	3.950	0.000	12
	CH	10.00	4.680	0.000	12
	C	0.50	6.400	0.000	12
TraPPE-alcohols[204]	CH$_3$-(OH)	98.00	3.750	0.265	12
	CH$_2$-(OH)	46.00	3.950	0.265	12
	O	93.00	3.020	-0.700	12
	H	0.00	0.000	0.435	12

All bonded parameters for alkanes, perfluoroalkanes, and 1-octanol, were taken from previous work[24, 67, 83, 124, 203, 204]. Fixed bond lengths were used to connect pseudo-atoms and are listed in Table 5.2. Equilibrium bond angles and bending constants are listed in Table 5.2.

Table 5.2: Equilibrium bond lengths, bond angles, and bending constants for alkanes, perfluoroalkanes, and alcohols.

Bond type	Bond length/ Å	Angle type	θ_0/degree	k_θ/ K-rad^{-2}
CF$_x$−CF$_y$	1.540	CF$_x$−CF$_2$−CF$_y$	114.00	62500
CH$_x$−CH$_y$	1.540	CH$_x$−CH$_2$−CH$_y$	114.00	62500
CH$_x$−C	1.540	CH$_x$−CH−CH$_y$	112.00	62500
CH$_x$−O	1.430	CH$_x$−C−CH$_y$	109.47	62500
O−H	0.945	CH$_x$−CH$_2$−O	109.47	50400
		CH$_x$−O−H	108.50	55400

The cosine series in dihedrals, accounts for the total rotational barrier, and no 1-4 Lennard-Jones interactions were included in the model. The dihedral parameters used in this work are listed in Table 5.3.

Table 5.3: Torsional parameters for alkanes, perfluoroalkanes, and alcohols.

Torsion	n	c_n/K	δ_n
CF_x—(CF_2)—(CF_2)—CF_y	0	-1577.68	0
	1	791.61	0
	2	333.65	0
	3	854.01	0
	4	349.25	0
	5	211.51	0
	6	117.66	0
	7	-83.44	0
CH_x—(CH_2)—(CH_2)—CH_y	1	355.03	0
	2	-68.19	180
	3	791.32	0
CH_x—(CH_2)—(CH)—CH_y	0	-251.06	0
	1	428.73	0
	2	-111.85	180
	3	441.27	0
CH_x—(CH_2)—(C)—CH_y	3	461.29	0
CH_x—(CH_2)—(CH_2)—O	1	176.62	0
	2	-53.34	180
	3	769.93	0
CH_x—(CH_2)—(O)—H	1	209.82	0
	2	-29.17	180
	3	187.93	0

5.4   Simulation Methodology

The molecular exchange Monte Carlo algorithms described in Section Methods were implemented for Gibbs ensemble simulations in the development version of GPU Optimized Monte Carlo (GOMC), which is available to the public via GitHub[170]. GOMC is an object-

oriented Monte Carlo simulation engine, capable of performing simulations in canonical, isobaric-isothermal, grand canonical ensembles, as well as Gibbs ensemble Monte Carlo. GOMC is designed for the simulation of complex molecular topologies, and supports a variety of potential functions, such as Lennard-Jones and Mie potentials. Coulomb interactions are also supported via the Ewald summation method[122]. GOMC is capable of parallel computation, either on multicore CPUs or GPUs.

Initial configurations were generated with Packmol[142], and Psfgen was used to generate coordinate (*.pdb) and connectivity (*.psf) files[143]. Parameters for the configurational-bias swap move were: 100 angle trials, 100 dihedral trials, 10 trial locations for the first site, and 8 trial locations for secondary sites. In calculations using the MEMC move, the first and last carbon atoms in a molecule were used to define the backbone of the large and small molecules. Non-bonded potentials were truncated at 10 Å and 14 Å for Mie[24, 124] and TraPPE[67, 203, 204] force fields, respectively, and analytical tail corrections were applied to the energy and pressure[124]. For all simulations with polar molecules, the real part of electrostatic potential was truncated at 14 Å and 120 Å for the liquid and vapor phase, respectively, and an Ewald convergence tolerance of 1×10^{-5} was used. For NVT-GEMC simulations, the pressure was calculated with frequency of 1×10^3 MCS.

5.4.1 Pressure-Composition Diagrams

The pressure-composition phase diagrams of methane+n-butane and perfluorobutane+n-butane were predicted using NPT-GEMC and NVT-GEMC simulations, respectively. NVT-GEMC simulations were used for the perfluorobutane+n-butane system due to the narrow range of pressures for which vapor-liquid phase coexistence exists at 260 K. Calculations were performed on systems containing 1000 molecules, with a move ratio of 38% displacements, 10% rotations, 2% volume transfers, 10% coupled-decoupled configurational-bias regrowth, 20% MEMC, and 20% coupled-decoupled configurational-bias (CD-CBMC)

molecule transfers. 4×10^7 Monte Carlo steps (MCS) were used for equilibration, followed by a data production period of 6×10^7 MCS. Statistical uncertainties were determined from three independent sets of simulations, where each simulation was initiated with a different random number seed. For the MEMC move, an exchange sub-volume of 8.8 Å x 8.8 Å x 11.8 Å was used[83].

5.4.2   Free Energies of Transfer

To calculate the Gibbs free energies of transfer for n-alkanes (C1-C8) from vapor phase to liquid 1-octanol, n-hexadecane, or 2,2,4-trimethylpentane, the initial vapor phase consisted of 580 methane, 50 ethane, 10 propane, 2 n-butane, 2 n-pentane, 2 n-hexane, 2 n-heptane, and 2 n-octane molecules, packed in a cubic box with side length of 300 Å. The liquid phase contained 240 1-octanol, 150 n-hexadecane or 250 2,2,4-trimethylpentane molecules, packed in cubic box with side length of 30 Å. Each phase was equilibrated for 1×10^7 MCS with isobaric-isothermal (NPT) simulations, with a move ratio of 38% displacements, 20% rotations, 2% volume transfers, 20% CD-CBMC regrowth, and 20% crankshaft[205, 206].

Starting from the equilibrated configurations, NPT-GEMC simulations were performed at 298 K and 1 atm pressure, with a move ratio of 29% displacements, 10% rotations, 1% volume transfers, 10% CD-CBMC regrowth, 10% crankshaft, 20% MEMC, and 20% CD-CBMC molecule transfer. Seven large-small molecule pairs were defined for the MEMC move, with an exchange ratio of 1:1: (ethane, methane), (propane, ethane), (n-butane, propane), (n-pentane, n-butane), (n-hexane, n-pentane), (n-heptane, n-hexane), and (n-octane, n-heptane). Pairs were chosen with equal probability to exchange the large molecule with the small molecule in the liquid phase. Each simulation was run for 1×10^8 Monte Carlo steps, and statistical uncertainties were determined from ten independent sets of simulations, where each simulation was initiated with a different random number seed. All molecule types were allowed

to be transferred between vapor and liquid phases, except for 1-octanol, n-hexadecane, and 2,2,4-trimethylpentane, due to the very low vapor pressure of these molecules at 298 K.

5.5   Results and Discussion

In this section, a number of examples are provided to validate the molecular exchange move in GEMC simulations, and to illustrate how the MEMC move can be used to significantly enhance acceptance rates for the molecule transfer move. Binary mixture phase diagrams were calculated for methane+n-butane and perfluorobutane+n-butane. Additional calculations were performed to calculate the Gibbs free energy of transfer for n-alkanes from vapor phase to liquid 1-octanol, n-hexadecane, and 2,2,4-trimethylpentane, to highlight the efficiency of the combination of MEMC and GEMC for the calculation of free energies of transfer.

5.5.1   Pressure-Composition Diagrams

In Figure 5.3, the pressure vs. composition diagram for methane+n-butane at 277 K, predicted using the ME-2 and the ME-3 algorithm, is compared with prior GCMC+histogram reweighting simulations[83] and experimental[181] data. Interactions between molecules were described with the Optimized Mie Potentials of Potoff *et al.*[24]. Calculations were performed with an exchange ratio of one n-butane with one methane. In addition to showing excellent agreement with experimental data, GEMC simulations with the ME-2 and ME-3 algorithm produced results that were statistically indistinguishable from prior simulations, validating the method. Additional calculations were performed with an exchange ratio of one n-butane to two methane molecules, and the resulting pressure-composition diagram is shown in Appendix C, Figure C3.

In Table 5.4, the average acceptance percentage for molecule transfers as a function of composition in liquid phase is presented for the methane+n-butane mixture at 277 K. When performing a one to one exchange, ME-3 was found to produce the largest improvement in

acceptance rates. Acceptance rates for the MEMC move were 3-10 times larger than configurational-bias swaps. The ME-2 algorithm produced acceptance rates that were 2-3 times configurational-bias swaps for $x_{methane} > 0.5$.

Because the ME-2 algorithm uses a rigid swap and the centroid of the large molecule is placed at the geometric center of the exchange sub-volume, only a fraction of the sub-volume is guaranteed to be empty. This is especially true when swapping molecules that differ greatly in size, such as methane and n-butane. In most of the ME-2 exchanges, it is likely that some atoms from the large molecule will have overlaps with existing molecules, lowering acceptance rates compared to ME-3. The ME-3 algorithm uses the same initial placement for the central atom as ME-2, but it grows the rest of the large molecule using configurational-bias. This allows the algorithm to find more energetically favorable configurations than are possible through a rigid molecule insertion, leading to greater acceptance rates for the exchange move in this system.

When performing a one to two exchange for methane+n-butane, ME-3 was found to produce up to a factor of three times improvement in acceptance rates at $x_{methane} = 0.95$, while the ME-2 algorithm produced acceptance rates similar to configurational-bias swaps. For all methane compositions, the acceptance rate for a one to two exchange is less than a one to one exchange due to the difficulty in finding two methane molecules within the exchange sub-volume.
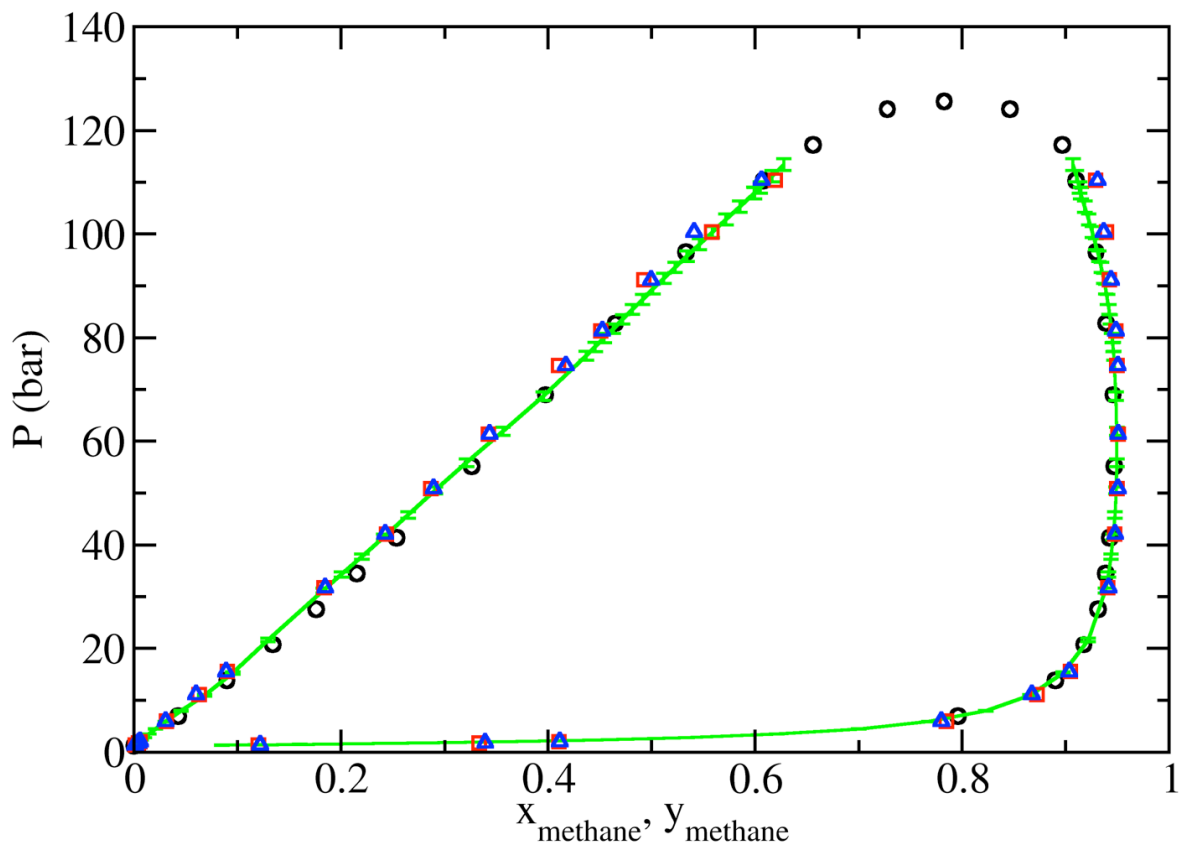
Figure 5.3: Pressure composition diagram for methane+n-butane at 277 K predicted from NPT-GEMC simulations using Mie potentials[24]. Experimental data (black circles)[181], GCMC+histogram reweighting[83] (green lines), ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles). Calculations were performed with an exchange ratio of one n-butane with one methane. The uncertainties in the predicted methane mole fractions were less than 0.01 and 0.004 in liquid and vapor phase, respectively.

In Figure 5.4, the pressure vs. composition diagram for perfluorobutane+n-butane at 260 K, predicted using GEMC simulations with the ME-2 and ME-3 algorithm is shown. Simulations were performed with the Mie potentials[24], which include an update to the perfluorocarbon force field[83]. The predictions of GEMC simulations are in close agreement with prior GCMC+histogram-reweighting calculations[83], validating the algorithm and its implementation in GOMC. Using standard Lorentz-Berthelot combining rules[120, 207] and no adjustable parameters for the cross interaction, very good agreement was achieved with experiment[186]. The largest deviation results from the limitation in the united-atom force field for perfluorobutane, which over-predicts the vapor pressure at 260 K by approximately 0.1 bar.
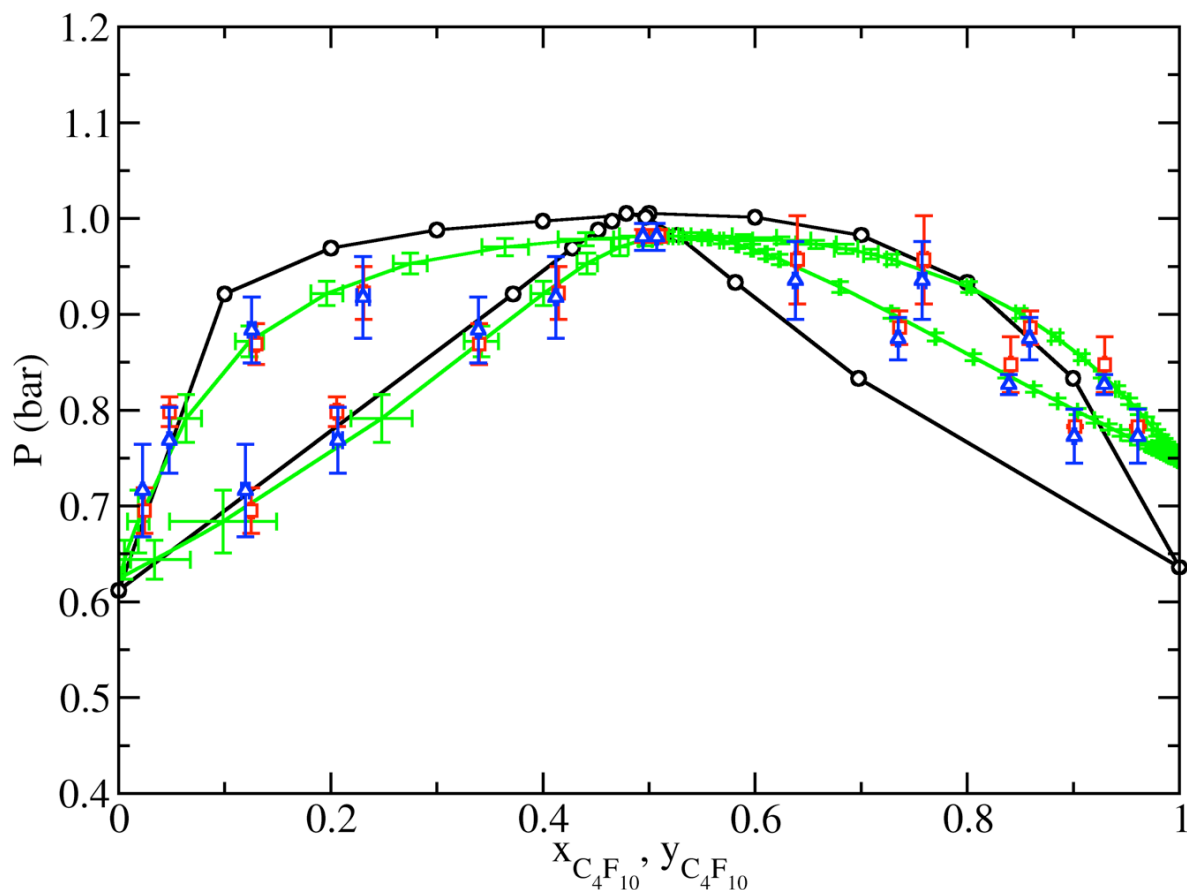
Figure 5.4: Pressure-composition diagram for perfluorobutane+n-butane at 260 K predicted from NVT-GEMC simulations using Mie potentials[24]. Experimental data (black circles)[186], GCMC+histogram reweighting (green lines)[83], ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles), with an exchange ratio of one perfluorobutane with one n-butane. The line connecting the experimental data points is provided as a guide to the eye.

In Table 5.4, the average acceptance percentages for molecule transfers as a function of composition in liquid phase are presented for the perfluorobutane+n-butane mixture at 260 K. Using coupled-decoupled configurational-bias (CD-CBMC) swaps, the probability of successfully inserting one perfluorobutane into the liquid phase, containing 14 mol%, 51 mol%, and 87 mol% of n-butane was 0.112%, 0.068%, and 0.018%, respectively. For the ME-2 algorithm, depending on composition, acceptance rates for the transfer of perfluorobutane were 47-204 times greater than CD-CBMC, while for ME-3, acceptance rates were 34-105 times greater than CD-CBMC. The MEMC move has approximately twice the computational cost of a standard CD-CBMC swap move, and therefore provides substantial improvements in

computational efficiency. For an extensive discussion of the computational efficiency of the MEMC algorithm readers are directed to Chapter 4. For this system, the ME-2 algorithm produces the largest acceptance rates because it works by aligning the backbone of perfluorobutane with the cavity left by the leaving n-butane and, in this case, the two molecules to be exchanged have similar size and shape. Acceptance rates were slightly lower for ME-3, since it grows the molecule using coupled-decoupled configurational-bias, without requiring the backbone of the molecule to be aligned with the cavity created by the molecule that was removed.

Table 5.4: Average acceptance percentages for molecule swaps of n-butane and perfluorobutane in GEMC simulations of methane+n-butane and perfluorobutane+n-butane, respectively.

Binary system	Sub-volume size (Å)	N_{EX}	$x_{C_4H_{10}}^{liquid}$	$x_{C_4H_{10}}^{vapor}$	CD-CBMC	ME-2	ME-3
methane+n-butane	-	1	0.50	0.05	2.56	3.15	7.68
			0.75	0.05	0.87	1.67	5.53
			0.95	0.13	0.45	1.24	4.34
	$8.8 \times 8.8 \times 11.8$	2	0.50	0.05	2.56	2.30	4.63
			0.75	0.05	0.87	1.12	2.88
			0.95	0.13	0.45	0.44	1.27
perfluorobutane+ n-butane	-	1	0.14	0.27	0.112	5.267	3.832
			0.51	0.49	0.068	4.696	3.024
			0.87	0.66	0.018	3.675	1.887

5.5.2 Free Energies of Transfer

Understanding the partitioning of compounds between phases is important for a wide variety of applications, such as drug design[110, 208, 209], design of separation processes[210], and prediction of environmental fate of toxic industrial chemicals[211, 212]. The most widely used partition coefficient describes the distribution of a solute between 1-octanol and water, which may be determined from the differences in the Gibbs free energies of hydration ΔG_{hyd} and solvation in 1-octanol ΔG_{solv}.

$$\log K_{OW} = \frac{\Delta G_{hyd} - \Delta G_{solv}}{2.303RT} \qquad (5.14)$$

Therefore, it is possible to determine log K_{ow}, and other partition coefficients, directly from computer simulations as long as a suitable methodology exists for the determination of ΔG.

The most common method to calculate free energy changes from atom-based computer simulations is to use molecular dynamics simulations coupled with techniques, such as thermodynamic integration (TI)[110, 213-215], free energy perturbation (FEP)[216, 217], or adaptive biasing force (ABF)[30, 105, 218, 219]. To achieve reliable sampling, these methods require the reaction coordinate to be divided into multiple smaller windows, where each window corresponds to a specific scaling of the Lennard-Jones and electrostatic interactions. Depending on the techniques used, and the level of accuracy desired, the number of discrete windows may vary from 16[220] to over 60[215, 221]. Typical simulation run lengths vary from 2-10 ns per window[215].

Alternatively, recognizing that the Gibbs free energy of transfer could be calculated from the average number density of the solute in each phase[222], Martin and Siepmann proposed an elegant and computationally efficient means for calculating free energies of transfer using Gibbs ensemble Monte Carlo simulations[77]

$$\Delta G_i^{transfer} = -RT \ln \left(\frac{<\rho_i^{liquid}>}{<\rho_i^{gas}>} \right)_{eq} \qquad (5.15)$$

where, R and T are the molar gas constant and absolute temperature in K, respectively, $<\rho_i^{liquid}>$ and $<\rho_i^{gas}>$ are the ensemble averaged number density (molecule/ Å^3) for solute i in liquid and gas phase at equilibrium, respectively. This methodology was used to determine air-water, air-octanol and water-octanol free energies of transfer for alkanes and alcohols with four or fewer carbons using the OPLS[116] and TraPPE[223] force fields.

A critical issue in the use of Gibbs ensemble Monte Carlo for the calculation of free energies of transfer is achieving sufficient numbers of molecule transfers between phases. In the past, this has been addressed by application of the "swatch" move, used extensively by Siepmann and co-workers[77], and more recently by inclusion of continuous fractional component methods[224]. In this section, the effectiveness of the combination of GEMC and the MEMC move is demonstrated through calculations of free energies of transfer of n-alkanes from vapor into liquid 1-octanol, n-hexadecane, and 2,2,4-trimethylpentane.

The octanol-air partition coefficient is used in environmental fate modeling[225], and has been shown to correlate well with air-soil[226, 227] and air-particle partitioning[228]. In Figure 5, the free energies of transfer for n-alkanes (C1-C8) from vapor to liquid 1-octanol at 298 K and 1 atm are shown. Tabulated values of the free energies are listed in Table 5.5. GEMC calculations were performed with the ME-2 and ME-3 algorithms, and the TraPPE force field[203, 204]. The predicted free energies of solvation are in close agreement with experiment[229] and prior calculations using molecular dynamics and the adaptive biasing force method[230]. The results also agree with prior molecular dynamics simulations using thermodynamic integration (TI) for methane through hexane, but differences of up to 0.7 kcal/mol in ΔG_{solv} were observed for n-octane[220]. The trend in the data suggests that the TI generated free energy data for longer alkanes in 1-octanol may not have been adequately converged.
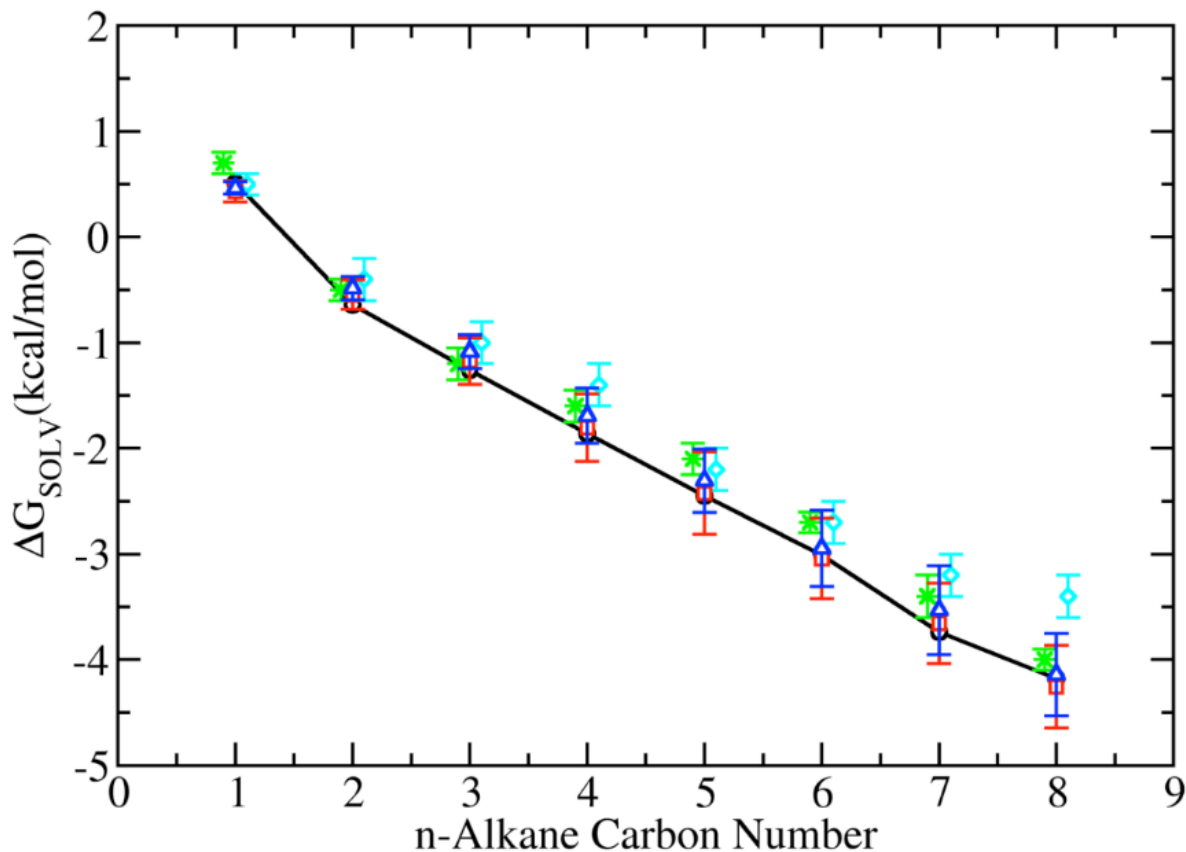
Figure 5.5: Free energy of solvation for n-alkanes in liquid 1-octanol predicted from NPT-GEMC simulations at 298 K and 1 atm using the TraPPE forcefield[203, 204]. Experimental data (black circles)[229], adaptive biasing force (green stars)[230], thermodynamic integration (cyan diamonds)[220], ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles). The line connecting the experimental data points is provided as a guide to the eye. The TI and ABF data points are shifted slightly along the x-axis for clarity.

Table 5.5: Free energies of transfer for n-alkanes from gas phase to liquid 1-octanol at 298 K and 1 atm calculated with the TraPPE force field[203, 204]. Number in parenthesis corresponds to the statistical uncertainties in the last digit determined from ten independent simulations.

	Free energy of solvation (kcal/mol)					
Solute \ Method	ME-2	ME-3	ABF[230]	TI[220]	GEMC[223]	Experiment[229]
methane	0.44(11)	0.46(06)	0.70(10)	0.5(1)	0.44(1)	0.51
ethane	-0.54(16)	-0.49(11)	-0.50(10)	-0.4(2)	-0.54(2)	-0.64
propane	-1.18(22)	-1.09(16)	-1.20(15)	-1.0(2)	-1.18(2)	-1.26
n-butane	-1.82(29)	-1.70(25)	-1.60(15)	-1.4(2)	-1.82(2)	-1.86
n-pentane	-2.42(35)	-2.31(30)	-2.10(15)	-2.2(2)	-	-2.45
n-hexane	-3.02(35)	-2.94(35)	-2.70(10)	-2.7(2)	-	-3.01
n-heptane	-3.63(37)	-3.52(41)	-3.40(20)	-3.2(2)	-	-3.74
n-octane	-4.25(40)	-4.13(39)	-4.00(10)	-3.4(2)	-	-4.18

The air-hexadecane partition coefficient provides a measurement of non-specific interactions between molecules and plays an important role as a compound descriptor used in linear solvation energy relationships (LSER). LSER models are used for prediction of solute partitioning in a variety of process, providing data that are needed for transport and environmental fate modeling[231, 232].

Free energies of solvation for n-alkanes (C1-C8) in liquid n-hexadecane ΔG_{C16} at 298 K and 1 atm, predicted using the TraPPE[203] and Mie[24] force fields, are shown in Figure 5.6. Tabulated numerical data are provided in Table 5.6. The predicted free energies of solvation using the Mie potential are in excellent agreement with experiment[229]. Predictions of the TraPPE force field are in close agreement with experiment for methane through n-butane, however, for longer alkanes TraPPE over-predicts ΔG_{C16} by up to 0.4 kcal/mol for n-octane, which is consistent with prior work[233].
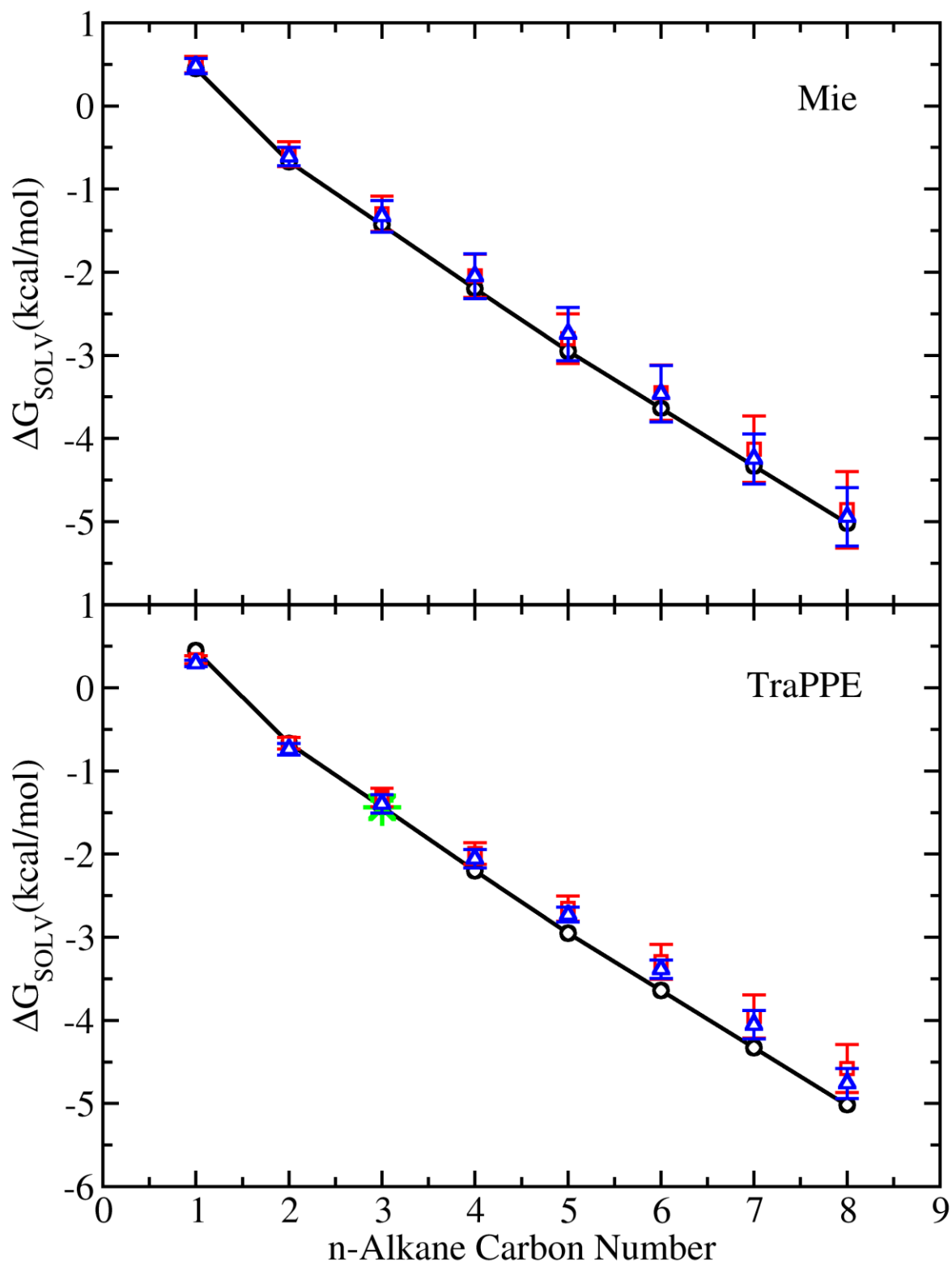
Figure 5.6: Free energies of solvation for n-alkanes in liquid n-hexadecane at 298 K and 1 atm predicted from NPT-GEMC simulations using TraPPE[203] and Mie[24] potentials. Experimental data (black circles)[229], thermodynamic integration (green stars)[234], ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles). The line connecting the experimental data points is provided as a guide to the eye.

Like the air-hexadecane partition coefficient, the air-2,2,4-trimethylpentane partition coefficient provides a measurement of non-specific solute-solvent interactions. In this case, calculations are performed to demonstrate the consistency of parameterization and transferability of potential parameters for the TraPPE and Mie force fields. Free energies of solvation for n-alkanes (C1-C8) in liquid 2,2,4-trimethylpentane at 298 K and 1 atm, predicted using the TraPPE[67] and Mie[124] force fields, are shown in In Figure 5.7. Tabulated numerical data are provided in Table 5.6. The predicted free energies of solvation using the Mie potential are in excellent agreement with experiment for all solutes[229], while TraPPE force field over-predicts the free energy of solvation by 0.5 kcal/mol for n-octane. The reported experimental value for the free energy of solvation for n-hexane does not follow the trend of other alkanes, and appears to be erroneous. The calculated free energy of solvation for n-hexane, using the SM5.42R/PM3 solvation model, was found to be -4.14 kcal/mol[229], which is consistent with the trends predicted by simulation. Considering that the SM5.42R/PM3 solvation model under-predicts the experimental solvation free energies of n-pentane and n-octane by 0.19 kcal/mol, it can be assumed that the experimental free energy of solvation for n-hexane should be around -3.95 kcal/mol, which is in exact agreement with the predictions of the Mie potentials.
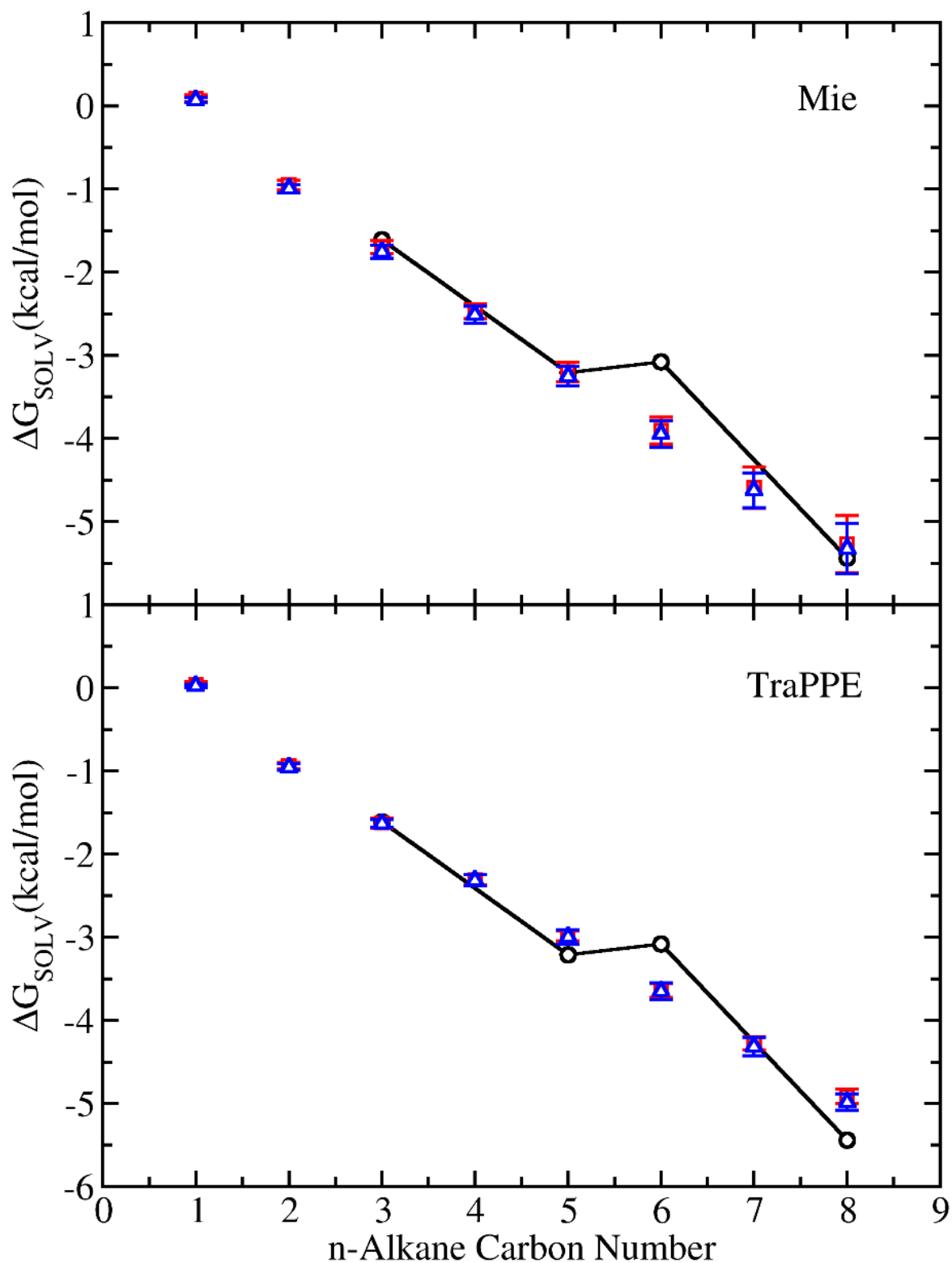
Figure 5.7: Free energies of solvation for n-alkanes in liquid 2,2,4-trimethylpentane at 298 K and 1 atm predicted from NPT-GEMC simulations using TraPPE[67] and Mie[124] force fields. Experimental data (black circles)[229], ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles). The line connecting the experimental data points is provided as a guide to the eye.

Table 5.6: The free energies of transfer for n-alkanes from gas phase to liquid n-hexadecane and 2,2,4-trimethylpentane at 298 K and 1 atm. Calculations were performed with the TraPPE[67, 203] and Mie[24, 124] potentials. Number in parenthesis corresponds to the statistical uncertainties in the last digit determined from ten independent simulations.

	Free energy of solvation (kcal/mol)									
Solvent	n-hexadecane					2,2,4-trimethylpentane				
Force field	TraPPE		Mie		Expt.[229].	TraPPE		Mie		Expt.
Solute	ME-2	ME-3	ME-2	ME-3	-	ME-2	ME-3	ME-2	ME-3	-
methane	0.34(05)	0.29(04)	0.50(10)	0.48(09)	0.45	0.04(03)	0.03(02)	0.10(04)	0.07(03)	-
ethane	-0.67(07)	-0.74(07)	-0.58(15)	-0.61(11)	-0.67	-0.94(04)	-0.95(04)	-0.95(06)	-1.00(05)	-
propane	-1.32(11)	-1.40(11)	-1.30(21)	-1.33(19)	-1.43	-1.62(06)	-1.63(05)	-1.70(08)	-1.75(08)	-1.61
n-butane	-1.99(13)	-2.06(11)	-2.05(26)	-2.05(27)	-2.20	-2.31(06)	-2.32(07)	-2.47(09)	-2.51(10)	-
n-pentane	-2.66(15)	-2.73(09)	-2.80(30)	-2.74(32)	-2.95	-2.98(06)	-3.00(09)	-3.20(12)	-3.25(12)	-3.21
n-hexane	-3.30(21)	-3.39(11)	-3.45(33)	-3.46(34)	-3.64	-3.64(08)	-3.65(10)	-3.90(16)	-3.95(16)	-3.08
n-heptane	-3.95(26)	-4.05(17)	-4.13(40)	-4.25(30)	-4.33	-4.28(08)	-4.32(11)	-4.59(25)	-4.62(21)	-
n-octane	-4.58(29)	-4.76(18)	-4.86(46)	-4.94(35)	-5.02	-4.92(09)	-4.98(10)	-5.27(34)	-5.33(30)	-5.44

In Table 5.7, the average acceptance percentage for insertion/deletion of n-alkane solutes in liquid 1-octanol, n-hexadecane, and 2,2,4-trimethylpentane phase is presented for coupled-decoupled configurational-bias swap, ME-2, and ME-3 methods using the TraPPE potential[67, 203, 204]. Additional data for the average acceptance percentage for insertion/deletion of n-alkane solutes in liquid n-hexadecane and 2,2,4-trimethylpentane phase, using Mie potential[24, 124], are presented in Appendix C, Table C1. As expected, the direct transfer of the solute from gas to liquid phase using the coupled-decoupled configurational-bias method decreases as the solute size increases. The methane transfer acceptance percentage in 1-octanol, n-hexadecane, and 2,2,4-trimethylpentane is 1.73%, 2.02%, and 5.41%, respectively, while the n-octane transfer acceptance percentage is near zero in each solvent. Due to the near zero acceptance rate for transferring solutes longer than n-butane, the insertion/deletion of these molecules depends completely on the MEMC move, which swaps n-butane for n-pentane, n-pentane for n-hexane, n-hexane for n-heptane, n-heptane for n-octane, and vice versa. For all solute exchange pairs, ME-2 algorithm acceptance percentages are 2-10X higher than ME-3 algorithm, except for methane-ethane exchange pair, where ME-

3 acceptance percentages are larger than ME-2. The greater acceptance rate for ME-2 vs. ME-3 can be attributed to the use of a rigid body insertion and rotation, which avoids the need to regrow the entire molecule. While regrowing the molecule helps find favorable regions for the molecule, growths can fail if the intramolecular geometric constants (angle bending and dihedral rotation) are not satisfied. Modification of the configurational-bias algorithm to use a Jacobian-Gaussian scheme for the generation of bond angles, for example, may lead to improved acceptance rates for the ME-3 algorithm[194].

Table 5.7: Average solute transfer acceptance percentages in GEMC simulations for mixtures of n-alkane +1-octanol, +n-hexadecane, and +2,2,4-trimethylpentane, using the TraPPE potential[67, 203, 204]. The coupled-decoupled configurational-bias swap acceptance percentages are presented for the small solute swap. The acceptance percentages for ME-2 and ME-3 are for exchanging a small solute with a large one.

Solvent	Solute (small)	Solute (large)	CD-CBMC	ME-2	ME-3
1-octanol	methane	ethane	1.7260	9.1609	13.5644
	ethane	propane	0.7913	7.1385	3.8629
	propane	n-butane	0.1575	3.5902	1.1005
	n-butane	n-pentane	0.0528	2.9461	0.6285
	n-pentane	n-hexane	0.0211	4.5830	0.6617
	n-hexane	n-heptane	0.0078	4.7709	0.6131
	n-heptane	n-octane	0.0026	3.0222	0.2565
	n-octane	-	0.0007	-	-
n-hexadecane	methane	ethane	2.0225	11.5058	19.8150
	ethane	propane	0.9813	9.4266	6.4551
	propane	n-butane	0.2083	4.7896	2.0233
	n-butane	n-pentane	0.0720	3.9912	1.1219
	n-pentane	n-hexane	0.0282	5.8574	1.0091
	n-hexane	n-heptane	0.0097	5.1392	0.6763
	n-heptane	n-octane	0.0026	2.6326	0.2237
	n-octane	-	0.0006	-	-
2,2,4-trimethylpentane	methane	ethane	5.4096	18.2207	28.1561
	ethane	propane	3.2777	15.3756	11.5313
	propane	n-butane	1.0048	8.6308	4.3540
	n-butane	n-pentane	0.4399	7.3336	2.5794
	n-pentane	n-hexane	0.2166	9.6328	2.0855
	n-hexane	n-heptane	0.0879	6.8203	1.1186
	n-heptane	n-octane	0.0268	2.7057	0.2887
	n-octane	-	0.0062	-	-

In Figure 5.8, the effect of run length on the statistical uncertainties of the free energies of transfer is shown as a function of number of carbon atoms for both the ME-2 and ME-3 methods. Calculations were performed with the TraPPE force field. In most cases, the ME-3 method produces lower statistical uncertainties for all solutes, despite having lower acceptance rates than the ME-2 method. While greater numbers of molecule identity exchanges occur with

ME-2, because ME-3 regrows the solute, the ME-3 method leads to faster sampling of phase space for linear molecules. Only when there is a large difference in acceptance rates does the performance of ME-2 surpass that of ME-3. An example of this is shown for alkane solvation in 2,2,4-trimethylpentane, where ME-2 produces slightly lower statistical uncertainties than ME-3. In this case, acceptance rates for ME-3 were 2-10 times lower than ME-2. For the case of solvation in 1-octanol, for methane to n-butane, the short (5×10^7 Monte Carlo steps) and long simulations (1×10^8 Monte Carlo steps) have similar statistical uncertainties. For molecules longer than butane, increasing the run length by a factor of two reduces the statistical error significantly.
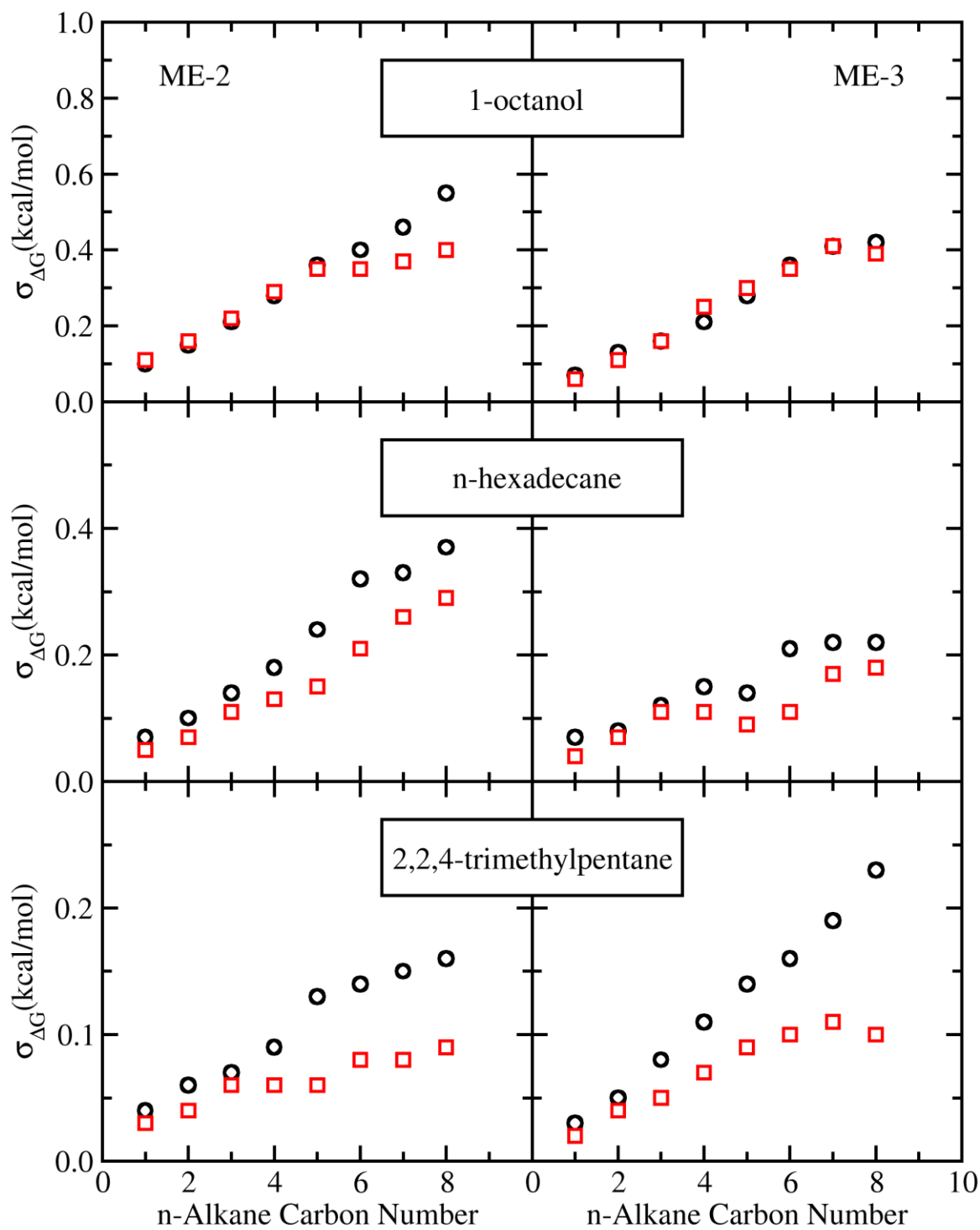
Figure 5.8: Standard deviation of predicted free energies of transfer for simulations of 5×10^7 Monte Carlo steps (black circles) and 1×10^8 Monte Carlo steps (red squares).

5.5.3 Evaluation of Computational Performance

To compare the efficiency of the GEMC-MEMC algorithm as implemented in GOMC with other methods for the calculation of free energies, the solvation free energy for ethane in 1-octanol and n-octane in 2,2,4-trimethylpentane were calculated using GOMC for Monte Carlo, and GROMACS version 2018.2 for thermodynamic integration in molecular dynamic simulations[41]. All calculations were performed on an Intel(R) i5-8600K 3.60GHz CPU. Calculations in GROMACS followed the protocol given by Garrido *et al.*[220], and used the same number of solvent molecules in the liquid phase as the Monte Carlo calculations. Lennard-Jones interactions were truncated at 1.0 nm and included dispersion corrections for the energy and pressure. Electrostatic interactions were calculated with the particle mesh Ewald with a converge tolerance of 1E-4. 16 discrete λ values were used, $\lambda \in$

$$\left\{ \begin{matrix} 0.0, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.65, \\ 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00 \end{matrix} \right\},$$ and NVT molecular dynamics simulations

of 5 ns in length were performed for each λ value at 298 K. Molecular dynamics simulations were run using 6 CPU-threads and required a total of 50 CPU hours for ethane-1-octanol ($\Delta G_{OA} = -0.54$ kcal/mol) and 35 CPU-hours for n-octane-2,2,4-trimethylpentane ($\Delta G_{224} = -5.00$ kcal/mol). These free energy results are in exact agreement with the predictions of GOMC for the TraPPE force field listed in Tables 5.5 and 5.6. The timing data for the molecular dynamics simulations correspond to the free energy calculations, only, and do not include the CPU time required to equilibrate the system. Calculations with GOMC used 4 threads for ethane+1-octanol, and 2 threads for n-octane+2,2,4-trimethylpentane. These calculations generated free energy data for all eight solutes from a single simulation, and required a total of 234 CPU hours for ethane in 1-octanol and 50 CPU hours for n-octane in 2,2,4-trimethylpentane. Considering that Monte Carlo calculations produced data for 8 solutes from a single simulation, while the MD simulations produced only a single data point, the Monte Carlo and molecular dynamics simulations show similar computational performance.

CHAPTER 6 EFFECT OF FLUORINATION ON THE PARTITIONING OF ALCOHOLS

6.1   Introduction

Perfluoroalkyl substances (PFAS) are a broad class of compounds where fluorine has been substituted for hydrogen on the alkyl chains. The most widely used and industrially relevant PFAS are surfactants, where fluorination of the alkyl tails renders them both hydrophobic and oleophobic, giving rise to unusual properties, such as exceptional chemical and thermal stability and very low interfacial tension at the air-water interface[85-87]. Owing to their unique properties, PFAS are used in a broad array of consumer applications, including coatings for non-stick cookware[88], grease-resistant paper[89], and stain resistant fabrics. Industrial applications include fire-fighting foams[90] and mist-suppressants in hard chrome plating[91].

The strength of the C-F bond, which contributes to the stability of fluorinated surfactants, also makes them extremely resistant to thermal, chemical, or photo degradation; experiments have shown that perfluorinated surfactants are highly resistant to biological degradation[92]. Numerous studies have shown widespread distribution of PFAS in the environment[93, 94]. As a result, PFAS are now considered to be a significant environmental threat[95].

Concerns about the environmental impact of PFAS lead to the phase-out of the two most common surfactants, perfluorooctanoic acid (PFOA) and perfluorooctanesulfonate (PFOS); however, the development of new fluorinated surfactants, some with reduced potential for bioaccumulation, is on-going[235, 236]. Analysis of fire sites where aqueous film forming foams (AFFF) had been used in Ontario, Canada, identified 103 different PFAS[237]. Fast atom bombardment and high resolution quadrupole-time-of-flight mass spectrometry performed on seven AFFF formulations used by the United States Military identified 10 unique

classes of compounds, with perfluoroalkyl chain lengths ranging from 4 to 12 carbon atoms[238]. The physicochemical properties, environmental fate, and toxicity of these compounds are largely unknown[238].

Environmental fate models rely on numerous physical property data, two of the most important of which are the Henry's law constant and the octanol-water partition coefficient, $\log K_{ow}$[96]. Given the breadth of PFAS chemistry and the lack of available experimental data, predictive methods are needed to fill these critical knowledge gaps. Prior work on the partitioning of fluorotelomer alcohols showed that common tools, such as EPISuite[97], CLOGP[98], SPARC[99] and COSMOTherm[100], produce a wide variety of results, with some predictions 2-5 orders of magnitude different than experiment[101].

Alternatively, atomistic computer simulations, combined with free energy methods such as thermodynamic integration[102], free energy perturbation[103, 104], or adaptive biasing force[30, 105], have been used with great success in the prediction of free energies of hydration and solvation in organic solvents for a wide variety of compounds[106-109]. While most work has focused on applications to drug[12, 13, 110] discovery, other calculations have focused on predicting the environmental fate of potentially toxic compounds, such as energetic materials[111, 112], ionic liquids[113], and fluorinated alcohols[114]. Additionally, computer simulations provide information on atomic-level structure, supporting the development of structure-property relationships.

While molecular dynamics simulations are widely used for the calculation of free energies of solvation, systems with large energy barriers to configurational and/or conformational change may exhibit biased sampling, leading to incorrect free energies if care is not taken[115]. On the other hand, Monte Carlo simulations allow the system to hop between states and in some cases, may offer conformational sampling advantages over molecular

dynamics. Free energies can be determined directly from Gibbs ensemble Monte Carlo simulations from the ratio of number densities of the solute in each phase[77, 84, 116]:

$$\Delta G_i^{transfer} = -RT \; ln \left(\frac{< \rho_i^{liquid} >}{< \rho_i^{gas} >} \right)_{eq} \tag{6.1}$$

where R and T are the molar gas constant and absolute temperature in K, respectively, and $< \rho_i^{liquid} >$ and $< \rho_i^{gas} >$ are the ensemble averaged number density (molecule/ Å^3) for solute i in liquid and gas phase at equilibrium, respectively.

Gibbs ensemble Monte Carlo provides a straightforward way of determining free energies of transfer as long as a sufficient number of successful exchanges of the solute between phases occurs, which usually requires the use of advanced configurational-bias sampling methods[83, 84, 116, 117]. For dense liquids with strong electrostatic interactions, obtaining adequately converged results for certain solutes may be challenging, even with state-of-the-art sampling algorithms for the molecule exchange move. The fluoro-alcohol systems of interest in this work present a perfect storm of sampling problems: the hydroxyl group has strong electrostatic interactions with the solvent (water or octanol) and it is difficult to find a favorably sized cavity to insert the bulky fluorinated alkyl tail. With enough intermediate states, nearly any molecule exchange between phases is possible[118], but if free energies of transfer are the quantity of interest, it may be more effective to perform standard thermodynamic integration or free energy perturbation. Therefore, this work describes the implementation of thermodynamic integration (TI) and free energy perturbation (FEP) methods into the Monte Carlo simulation engine GOMC[57], and the application of TI and FEP to determine the air-water, air-oil, air-octanol, and octanol-water partition coefficients for eight carbon alcohols with varying degrees of fluorination. Partitioning of fluorotelomer alcohols is of interest because they can degrade to form perfluorooctanoic acid (PFOA)[239]. The local solvation structure around C8 alcohols is determined and used to explain the impact

of fluorination of the alkyl tail on partitioning. The key founding of this work is presented in Chapter 7. Force Field parameters and additional results are provided in Appendix D. The result of this work has been published in Journal of Molecular Physics[119].

6.2 Force Field Parameters

Calculations were performed using the SPC water model[240], and the Transferable Potentials for Phase Equilibria (TraPPE) force field[187, 203, 204] to represent a variety of fluorinated analogues of 1-octanol and n-hexadecane, which are listed in Table 6.1. All non-bonded force field parameters are listed in Table 6.2.

Table 6.1: Fluorinated 1-octanol analogues studied in this work.

Molecular structure	Molecular formula	Molecular name
	$CH_3(CH_2)_7OH$	H8
	$CH_3(CH_2)_6CF_2OH$	F1H7
	$CH_3(CH_2)_5(CF_2)_2OH$	F2H6
	$CF_3(CF_2)_5(CH_2)_2OH$	H2F6
	$CF_3(CF_2)_6CH_2OH$	H1F7
	$CF_3(CF_2)_7OH$	F8

In TraPPE, a united-atom representation is used for all CF_x and CH_x groups; *i.e.* hydrogen or fluorine atoms bonded to carbon atoms are not represented explicitly and are,

instead, combined with carbon atoms to form a single interaction site or "pseudo-atom". Interactions between pseudo-atoms are described by pairwise-additive 12-6 Lennard-Jones potentials, combined with partial charges to represent Coulombic interactions. Non-bonded parameters for alkyl[203], perfluoro[187], and hydroxyl groups[204] were taken from the original TraPPE papers and are listed in Table 6.2. Parameters for unlike interactions were determined using the Lorentz-Berthelot combining rules[120, 121].

Table 6.2: Non-bonded parameters for alcohols, fluoroalcohols and fluorotelomer alcohols.

Group	ε/k_B (K)	σ (Å)	q_i
TraPPE-UA			
CH_3	98.0	3.75	0.0/0.265[*]
CH_2	46.0	3.95	0.0/0.265[*]
CF_3	87.0	4.36	0.0/0.265[*]
CF_2	27.5	4.73	0.0/0.265[*]
O (alcohol)	93.0	3.02	-0.700
H (alcohol)	0.0	0.0	0.435
SPC			
O	78.21	3.167	-0.820
H	0.0	0.0	0.410

[*]partial charges for the C_α bonded to oxygen.

United-atoms were connected with rigid bonds, for which the parameters are listed in the Table D1 of the Appendix D. Bond bending constants were taken from TraPPE[187, 203, 204], and are listed in the Table D1 of the Appendix D. Existing torsional potentials for the C-C-C-C backbone for n-alkanes and perfluoroalkanes in TraPPE were refit to use the form of equation (2.12) or taken from prior work[114]. Constants for all dihedral potentials are listed in the Table D2 of the Appendix D.

New Fourier coefficients for torsions in $CH_3(CH_2)_5(CF_2)_2OH$ (F2H6) and $CH_3(CH_2)_6CF_2OH$ (F1H7) were optimized to reproduce rotational barriers determined from relaxed potential energy scans generated from MP2/6-31+g(d,p) *ab initio* calculations. All *ab initio* calculations were performed in Gaussian 09[136].

6.3 Calculation of Solvation Free Energies

This section describes key details of the implementation of free energy perturbation and thermodynamic integration in GOMC. Free energy perturbation is discussed, first, followed by thermodynamic integration. Computational details for the calculations are given in Simulation Methodology.

In free energy perturbation (FEP)[103, 104], the free energy difference between two states A (*e.g.* non-interacting solute) and state B (*e.g.* fully interacting solute) is given by

$$\Delta G(A \rightarrow B) = -\frac{1}{\beta} \ln \langle \exp(-\beta \Delta U_{A,B}) \rangle_A \tag{6.2}$$

where $\Delta U_{A,B} = U_B - U_A$ is the energy difference between the system in state A and B, and $\langle \exp(-\beta \Delta U_{A,B}) \rangle_A$ is the ensemble average for simulation in state A. For most systems, there is limited phase-space overlap between state A and B, leading to poor convergence of the free energy. By constructing an artificial pathway through multistage sampling[241], satisfactory phase-space overlap can be achieved, greatly improving the accuracy and precision of the free energy calculation[242, 243]. Using the multistage sampling approach, the free energy difference between two states A and B, with $N - 2$ intermediate states given by[244]

$$\Delta G(A \rightarrow B) = -\frac{1}{\beta} \sum_{i=0}^{N-1} \ln \langle \exp(-\beta \Delta U_{i,i+1}) \rangle_i \tag{6.3}$$

where $\Delta U_{i,i+1} = U_{i+1} - U_i$ is the energy difference of the system between states i and $i+1$, and $\langle \exp(-\beta \Delta U_{i,i+1}) \rangle_i$ is the ensemble average for simulation performed in intermediate state i. A coupling parameter $0.0 \leq \lambda \leq 1.0$ is used to smoothly transform the simulated system between states A ($\lambda = 0.0$) and B ($\lambda = 1.0$), where

$$U_i = \lambda_i U_B + (1 - \lambda_i) U_A \tag{6.4}$$

Naive linear scaling of the intermolecular interactions with respect to λ produces a well-known numerical instability (end-point catastrophe) in the limit of $\lambda \to 0$ and $\lambda \to 1$ for Lennard-Jones potentials[245, 246], which can be avoided by shifting and scaling the Lennard-Jones potential via the soft-core scheme[247, 248]. Electrostatic interactions do not have the same numerical instability if a two-step transformation is applied[249], and it has been shown that it is computationally efficient to scale them linearly[250].

Therefore, in this work, soft-core scaling is used for the Lennard-Jones interactions, while linear scaling is used for the Coulombic interactions. Separate λ_{LJ} and λ_{Coul} were used to independently control the scaling of Lennard-Jones and Coulombic interactions, respectively. The energy of the solute interacting with the solvent is given by

$$U_i(r_{ij}) = \lambda_{LJ} U_{LJ}(r_{sc-ij}) + \lambda_{Coul} U_{Coul}(r_{ij}) \tag{6.5}$$

where

$$r_{sc-ij} = \left(\alpha(1 - \lambda_{LJ})^p \sigma_{ij}{}^6 + r_{ij}{}^6\right)^{1/6} \tag{6.6}$$

r_{sc-ij}, α, and p are the scaled distance, softness parameter, and soft-core power, respectively. To improve numerical convergence of the calculation, a minimum interaction diameter $\sigma_{min} = 3.0$ Å was defined for any atom with a diameter less than σ_{min}, $e.g.$ hydrogen atoms attached to oxygen in water or alcohols[249].

The effect of long-range corrections on predicted free energies were determined for Lennard-Jones and Coulombic interactions via a linear coupling with λ.

$$U_{LRC(LJ)} = \lambda_{LJ} \Delta U_{LRC(LJ)} \tag{6.7}$$

$$U_{LRC-Coul} = \lambda_{Coul}\left[\Delta U_{self} + \Delta U_{correction} + \Delta U_{reciprocal}\right] \tag{6.8}$$

where $\Delta U_{LRC(LJ)}$, $\Delta U_{reciprocal}$, ΔU_{self}, $\Delta U_{correction}$ are the change in the long-range correction energy due to adding a fully interacting solute to the solvent for both the Lennard-Jones and Coulombic interactions.

In thermodynamic integration, the free energy change is calculated from

$$\Delta G(A \rightarrow B) = \int\limits_{\lambda=0}^{\lambda=1} \langle \frac{dU}{d\lambda} \rangle_\lambda \, d\lambda \tag{6.9}$$

where $dU/d\lambda$ is the derivative of energy difference with respect to λ, and $\langle \frac{dU}{d\lambda} \rangle_i$ is the ensemble average for a simulation run at intermediate state λ. To calculate the free energy using thermodynamic integration, the derivative of the intermolecular energy with respect to λ must be evaluated for both the Lennard-Jones and Coulombic interactions of the solute with the solvent.

$$\frac{dU_{LJ}(r_{ij})}{d\lambda_{LJ}} = U_{LJ}(r_{sc-ij}) + \frac{\lambda_{LJ}p\alpha}{6}(1-\lambda_{LJ})^{p-1}\left(\frac{\sigma_{ij}^{6}}{r_{ij}^{5}}\right)F_{LJ}(r_{sc-ij}) \tag{6.10}$$

$$\frac{dU_{Coul}(r_{ij})}{d\lambda_{Coul}} = U_{Coul}(r_{ij}) \tag{6.11}$$

$$F_{LJ}(r_{ij}) = -\frac{dU_{LJ}(r_{ij})}{dr_{ij}} = \frac{4\varepsilon_{ij}}{r_{ij}}\left[12\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{6.12}$$

The derivative of the long-range correction energies with respect to λ is given by

$$\frac{dU_{LRC(LJ)}}{d\lambda_{LJ}} = \Delta U_{LRC(LJ)} \tag{6.13}$$

$$\frac{dU_{LRC-Coul}}{d\lambda_{Coul}} = \Delta U_{self} + \Delta U_{correction} + \Delta U_{reciprocal} \tag{6.14}$$

6.4   Simulation Methodology

6.4.1   Free Energy Calculations

The free energy calculations described in previous Section, were implemented in the development version of the open-source Monte Carlo simulation engine GOMC[57], which is available to the public via GitHub[170]. To calculate the free energy of solvation/hydration, all intermediate λ states were equilibrated independently in the canonical ensemble (NVT) for $5\text{x}10^6$ Monte Carlo steps (MCS) at 298 K, followed by a $3\text{x}10^7$ MCS isobaric-isothermal (NPT) ensemble simulation at 1 bar and 298 K. Production data were taken from a subsequent $5\text{x}10^7$ MCS NPT simulation, which used the final configuration of the prior NPT simulation as the initial configuration. For production runs, all λ states were simulated independently in parallel. During the production run, the change in energy ($\Delta U_{i,j}$) between the current lambda state and all other lambda states, and the derivative of potential with respect to lambda ($dU_{Coul}/d\lambda_{Coul}$, $dU_{LJ}/d\lambda_{LJ}$), were evaluated and stored for post-simulation analysis every $5\text{x}10^3$ MCS. A sample of GOMC free energy output is provided in Table D3 of the Appendix D. The implementation of free energy methods into GOMC was validated through calculations of free energies of solvation for various n-alkanes in 1-octanol. A comparison with prior calculations performed with NPT-Gibbs ensemble Monte Carlo simulations[84] is provided in Table D4 of the Appendix D, and shows that all methods produce free energies that are within 0.1 kcal/mol of each other.

To calculate the free energy of solvation in water and 1-octanol, 23 intermediate lambda states, as shown in Figure 1, were used

$$\lambda_{coul,LJ} \in \left\{ \begin{array}{c} (0.0, 0.0), (0.0, 0.05), (0.0, 0.1), (0.0, 0.15), (0.0, 0.2), \\ (0.0, 0.25), (0.0, 0.3), (0.0, 0.35), (0.0, 0.4), (0.0, 0.45), \\ (0.0, 0.5), (0.0, 0.6), (0.0, 0.7), (0.0, 0.8), (0.0, 0.9), \\ (0.0, 1.0), (0.2, 1.0), (0.4, 1.0), (0.6, 1.0), (0.7, 1.0), \\ (0.8, 1.0), (0.9, 1.0), (1.0, 1.0) \end{array} \right\}$$

while 16 intermediate states were used to calculate the free energies of solvation in n-hexadecane.

$$\lambda_{coul,LJ} \in \left\{ \begin{array}{c} (0.0, 0.0), (0.0, 0.05), (0.0, 0.1), (0.0, 0.15), (0.0, 0.2), \\ (0.0, 0.25), (0.0, 0.3), (0.0, 0.35), (0.0, 0.4), (0.0, 0.45), \\ (0.0, 0.5), (0.0, 0.6), (0.0, 0.7), (0.0, 0.8), (0.0, 0.9), (0.0, 1.0) \end{array} \right\}$$
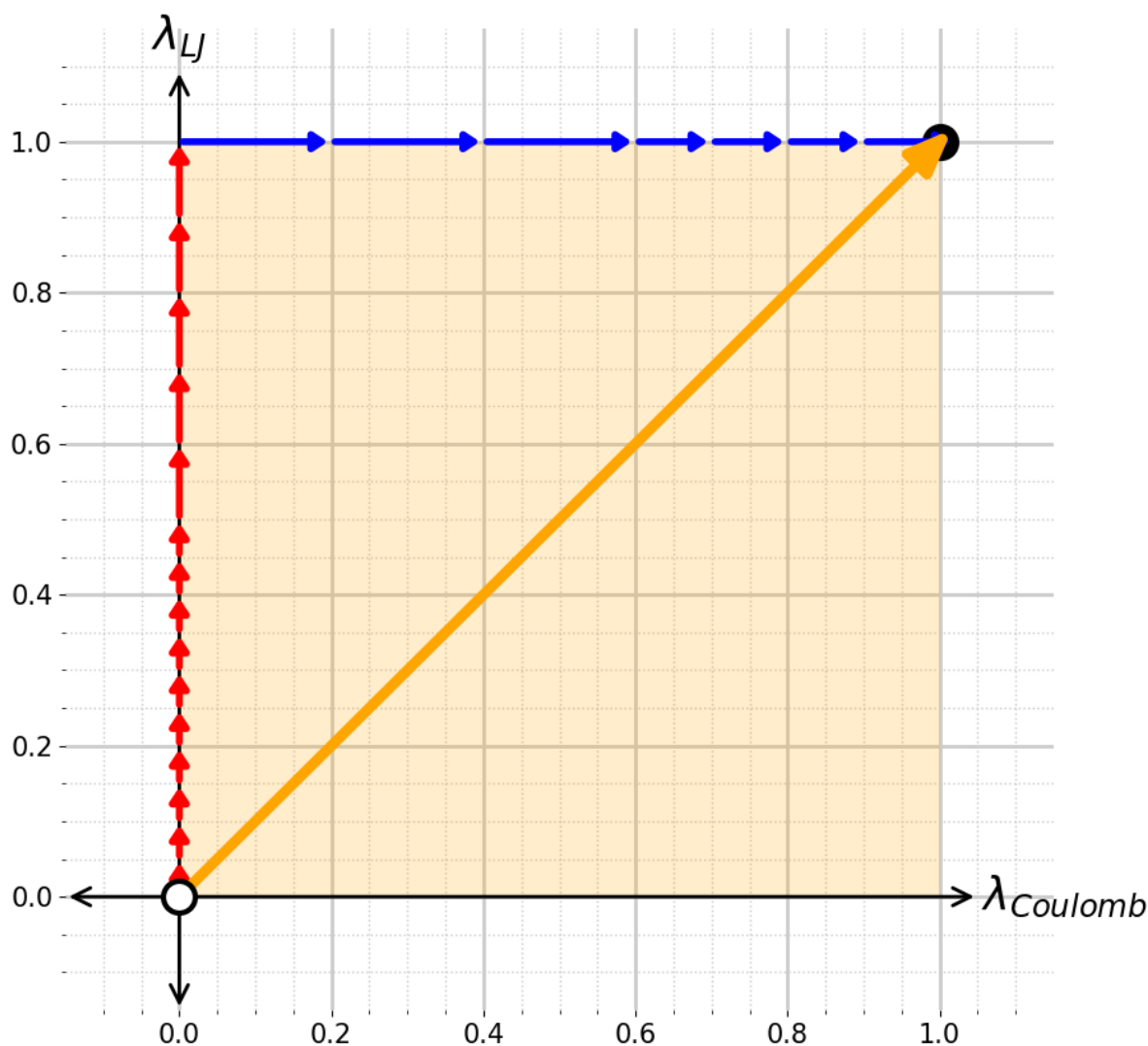
Figure 6.1: The transformation pathway starting from non-interacting solute (0.0, 0.0) to fully interacting solute (1.0, 1.0) in λ vector space, which is shown as an orange square on the Cartesian plane formed by the axes $\lambda_{Coulomb}$ and λ_{LJ}, which control the solute Coulombic and Lennard-Jones interactions, respectively. Intermediate states are denoted by the arrowheads.

While it is possible to alter the Lennard-Jones and Coulomb interactions simultaneously, recent work suggests it is more efficient to first turn on the full Lennard-Jones

interactions before scaling the Coulombic interactions[250, 251]. For liquid phase systems containing 1-octanol or water, the λ vectors were defined to turn on the full Lennard-Jones interaction, first, before introducing Coulombic interactions between the solute and the solvent, as shown in Figure 1, to avoid the direct interaction of atoms with "naked" charges[249, 252]. The soft-core parameters used for Lennard-Jones interactions were, $\alpha = 0.5$, $p = 2$, and $\sigma_{min} = 3.0$ [249, 253].

A variety of methods were used to analyze the resulting data, including thermodynamic integration (TI)[254], Bennett acceptance ratio (BAR)[255], and multistate Bennett acceptance ratio (MBAR)[256], as implemented in the software alchemlyb[257] and alchemical-analysis[258]. A parser for GOMC output was implemented for both alchemlyb and the alchemical-analysis. Since alchemical-analysis is no longer supported by its authors, the GOMC parser for it was stored in a separate GitHub repository[77].

To determine the free energy of solvation/hydration accurately, the data points used in the calculation must be sampled at equilibrium conditions and be uncorrelated. Several techniques have been developed[259, 260] to detect uncorrelated samples; both alchemlyb[257] and alchemical-analysis[258] use an autocorrelation time analysis, as implemented in pymbar[256]. In autocorrelation time analysis, the autocorrelation function $C_A(i)$ is determined for a data point i in a given data series (in this work $dU/d\lambda$), and the autocorrelation time (τ) is calculated as the integral of $C_A(i)$[261]. Once the autocorrelation time (τ) is obtained, the gth element of the data series is treated as an uncorrelated sample, where $g = 1 + 2\tau$. In pymbar, a data point is defined as statistically independent if $C_A(i) = 0$; however, the autocorrelation function becomes noisy as $C_A(i) \to 0$, making it difficult to rigorously determine uncorrelated samples. In practice, pymbar provides a conservative estimate of uncorrelated data, and tends to under-predict the number of uncorrelated samples.

In addition to using only uncorrelated samples, care must be taken to ensure that data used in the free energy calculation are collected from simulations that have reached equilibrium. Prior molecular dynamics simulations have shown, for example, challenges in converging liquid phase densities and free energies of solvation in 1-octanol[262]. In this work, NPT simulations of 3×10^7 MCS were used to equilibrate the system at each λ_i prior to the production run, ensuring stability of the density during free energy calculations, as shown in Figure S2 for perfluorooctanol in 1-octanol. Once free energy data were collected, convergence of the data were assessed by calculating free energies of hydration/solvation in both the forward and reverse directions with alchemical-analysis[258]. In the forward direction, the free energy was calculated using data in the order in which they were collected, while in the "reverse" direction, the free energy was calculated from the data ordered in the reverse of which it was collected. As shown in Figure 6.2 for F2H6, the forward and reverse calculations match within the statistical uncertainty of the data, suggesting convergence of the free energy calculations[221, 258]. Free energies were calculated from simulation data using a variety of thermodynamic integration methods (trapezoidal rule (TI) and cubic spline (TI-CUBIC)), and free energy perturbation techniques (Bennett acceptance ratio (BAR) and multi-state Bennett acceptance ratio (MBAR)). MBAR results are discussed in the body of the paper, while results for TI and BAR may be found in Table D5 of the Appendix D. For simulations that have high quality sampling, and sufficient overlap between energy difference distributions, it is expected that all methods will produce similar results. As shown in Figure 6.3, good agreement for all intermediate states was achieved with all methods.

Additional insight is provided by the overlap matrix, as shown in Figure 6.4. The overlap matrix quantifies the overlap of the $\Delta U_{i,j}$ distributions between neighboring intermediate states (i and j) and gives the probability of observing a sample from state i in state j, which can be used to detect intermediate states with insufficient overlap. In this case, the data

shown in Figure 6.4 conform to the recommendations of Klimovich *et al.*[258]. Neighboring

states along the main diagonal have overlap values significantly above the recommended value

of 0.03, indicating sufficient overlap between states has been achieved to obtain reliable free

energy predictions.
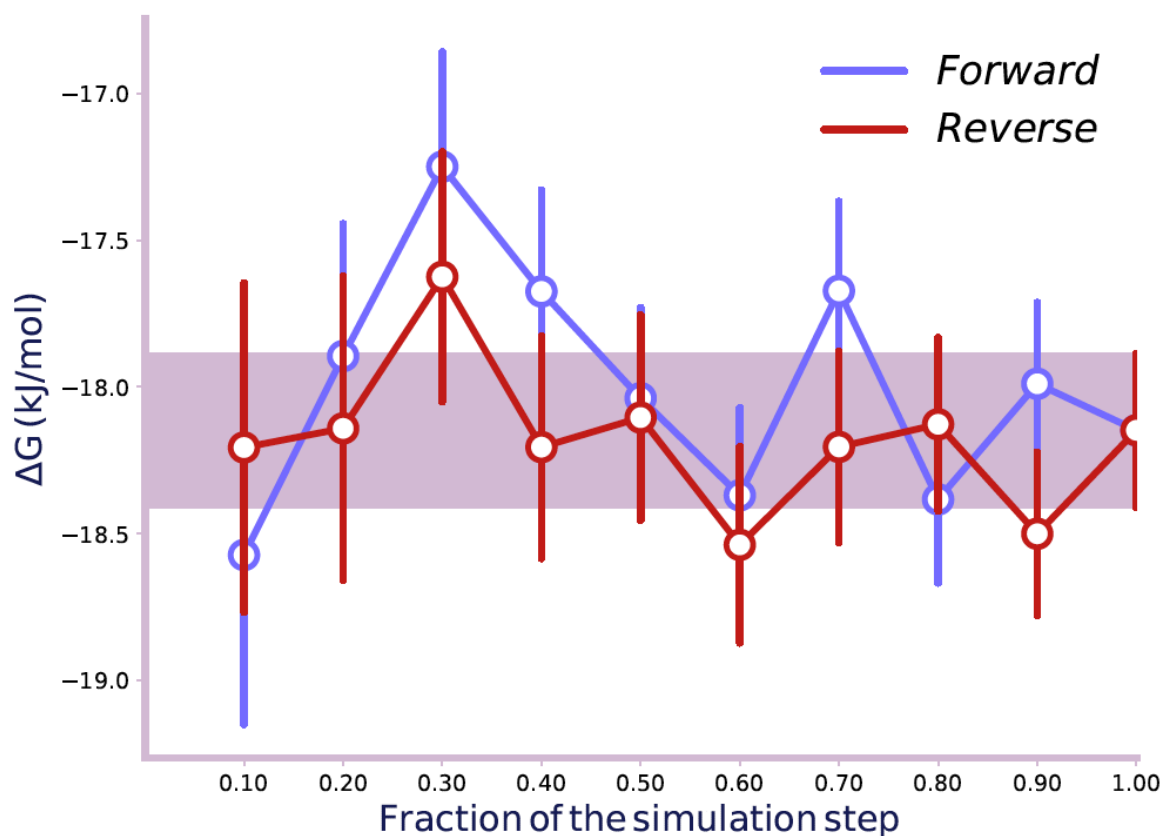
Figure 6.2: Solvation free energy for F2H6 in n-hexadecane plotted as a function of simulation steps. The agreement between the forward and reverse calculation is within the standard error bar (purple bar), indicating convergence of the free energy simulations.
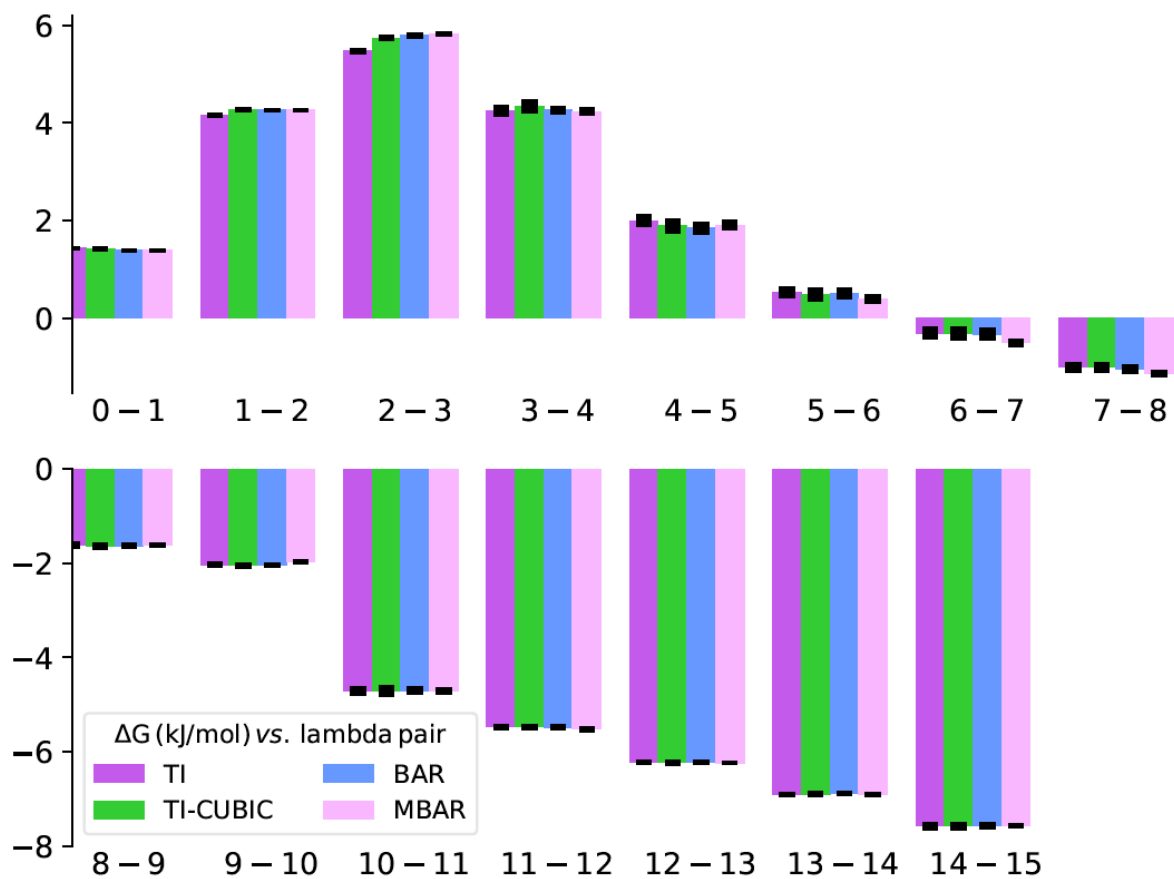
Figure 6.3: Intermediate free energy differences for solvation of F2H6 in n-hexadecane, calculated by a variety of thermodynamic integration and free energy perturbation techniques.

λ	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	.41	.34	.21	.05												
1	.34	.32	.25	.09	.01											
2	.21	.25	.29	.21	.04											
3	.05	.09	.21	.37	.21	.05	.01									
4		.01	.04	.21	.38	.23	.09	.03	.01							
5				.05	.23	.32	.22	.11	.05	.02	.01					
6				.01	.09	.22	.26	.20	.12	.06	.03	.01				
7					.03	.11	.20	.22	.18	.13	.08	.03	.01	.01		
8					.01	.05	.12	.18	.20	.18	.14	.07	.03	.02	.01	.01
9						.02	.06	.13	.18	.19	.18	.11	.06	.03	.02	.02
10						.01	.03	.08	.14	.18	.19	.15	.10	.06	.04	.03
11							.01	.03	.07	.11	.15	.18	.15	.12	.10	.08
12								.01	.03	.06	.10	.15	.18	.17	.15	.14
13								.01	.02	.03	.06	.12	.17	.19	.20	.20
14									.01	.02	.04	.10	.15	.20	.23	.24
15									.01	.02	.03	.08	.14	.20	.24	.28

Figure 6.4: Overlap matrix for the solvation of F2H6 in n-hexadecane.

6.4.2 Monte Carlo Simulations

NVT ensemble simulations were performed with a move ratio of 50% displacements, 20% rotations, 20% coupled-decoupled configurational-bias (CD-CBMC) regrowth[67], and 10% crankshaft[205, 206]. Parameters for the configurational-bias regrowth move were 100 angle trials, 50 dihedral trials, and 10 trial locations for grown pseudo-atoms. NPT ensemble simulations were performed with similar move ratios, except for the addition of 1% volume changes, while the percentage of displacement moves was reduced to 49%. Non-bonded potentials were truncated at 14 Å[187, 203, 204] and analytical tail corrections were applied to

the energy[124]. For simulations with electrostatic interactions, the real space part of electrostatic potential was truncated at 14 Å and an Ewald convergence tolerance of 1×10^{-5} was used[263].

During grand canonical and Gibbs ensemble Monte Carlo simulations, molecule swap moves are frequently used to sample phase-space. Intra-box swap moves may also be used to enhance the sampling of phase-space in NVT and NPT ensemble simulations. For polar molecules, where an atom has a naked charge, such as hydrogen in many alcohol and water models, during a swap move it is possible to a place opposing charges in close proximity. This produces very large negative energies that overwhelm the repulsive component of the Lennard-Jones potential, leading to the sampling of unphysical states. A common workaround is to introduce a hard inner cut-off and reject any trial moves that bring atom centers closer than 1 Å[264]. Using a hard inner cut-off in free energy simulations, however, produces incorrect sampling of the solvent structure in the limit of $\lambda \rightarrow 0$, leading to inaccurate free energies. Therefore, intra-box swap moves were not used during free energy simulations.

Liquid phase systems contained one solute in a solvent box of 200 1-octanol, 150 n-hexadecane, or 1000 water molecules. Initial cubic box sizes were selected to produce densities that were close to equilibrium, with a side length of 37.6, 41.6, and 31.3 Å for 1-octanol, n-hexadecane, and water, respectively. Initial configurations were generated with Packmol[142], and Psfgen was used to generate coordinate (*.pdb) and connectivity (*.psf) files[143].

Radial distribution functions for solute-solvent systems were determined by performing a 5×10^6 MCS equilibration in the canonical ensemble at 298 K, followed by 7×10^7 MCS NPT ensemble simulation at 1 bar and 298 K, where production data were taken from the last 5×10^7 MCS of the simulation. Atomic coordinates for all atoms in the system were stored every 2.5×10^3 MCS. Radial distribution functions were calculated from saved configurations with the gofr tool in VMD[265].

6.5 Results and Discussion

6.5.1 Free Energies of Hydration

Free energies of hydration predicted by simulation for each solute in SPC water are listed in Table 6.3. From trends in the data, where $\Delta G_{\text{water}}(\text{H8}) < \Delta G_{\text{water}}(\text{H2F6}) < \Delta G_{\text{water}}(\text{H1F7}) < \Delta G_{\text{water}}(\text{F8})$, it was hypothesized that fluorination near the head group has the greatest impact on the solubility of fluorinated alcohols in water. This was confirmed through free energy calculations for two additional molecules: F1H7 and F2H6. For each of these molecules, C_α (F1H7) or C_α and C_β (F2H6) were fluorinated, while the remaining carbon atoms were CH_2 or CH_3 groups. Fluorinating C_α produces a 1.5 kcal/mol increase (less negative) in the free energy of hydration compared to 1-octanol, while fluorination of C_β and C_α produces only an additional 0.1 kcal/mol change in ΔG_{water}. This free energy change, due to fluorination of only C_α, accounts for almost half of the difference in ΔG_{water} between 1-octanol and perfluorooctanol.

Table 6.3: Calculated free energies of hydration and solvation for alcohols predicted with the MBAR method, with a comparison to experimental data. Numbers in parenthesis correspond to the uncertainty in the last digit.

Molecule	$\rho_{liq}(298\,K)$	ΔG_{C16}(kcal/mol)		$\Delta G_{1-octanol}$(kcal/mol)		ΔG_{water}(kcal/mol)		$\log K_{ow}$	
		Sim.	Exp.	Sim.	Expt.	Sim.	Expt.	Sim.	Expt.
$CH_3(CH_2)_7OH$(H8)	826(5)	-5.15(5)	-6.3[229]	-8.6(2)	-8.13[229]	-2.9(2)	-4.09[266]	4.2(2)	3.0[267]
$CF_3(CF_2)_5(CH2)_2OH$ (H2F6)	1743(13)	-4.16(7)	-4.0(1)[268]	-7.1(2)	-7.2(3)[268] -6.01[269]	-1.7(2)	-0.76(3)[268], -2.01[269], 0.50[270]	4.0(2)	4.7(3)[101]
$CF_3(CF_2)_6CH_2OH$ (H1F7)	1847(14)	-4.10(7)		-6.0(2)		-1.6(2)		3.2(2)	
$CF_3(CF_2)_7OH$(F8)	1897(15)	-3.32(7)		-5.2(2)		0.0(3)		3.8(2)	
$CH_3(CH_2)_6CF_2OH$ (F1H7)	971(7)	-4.38(6)		-6.1(2)		-1.4(2)		3.4(2)	
$CH_3(CH_2)_5(CF_2)_2OH$ (F2H6)	1124(7)	-4.34(6)		-5.7(2)		-1.3(2)		3.2(2)	

Further insight into the role fluorination near the hydroxyl group plays in the altering the free energy of hydration, was obtained by calculating the separate Lennard-Jones and Coulombic contributions to the free energy for each solute, which are listed in Table 6.4. For solutes where C_α is hydrogenated, the Coulombic contribution to the free energy is between -6.5 and -6.1 kcal/mol, while for solutes where C_α is fluorinated, the Coulombic contribution is reduced to -5.6 to -5.4 kcal/mol. This provides evidence that fluorination of C_α reduces hydrogen bonding of the solute with water compared to 1-octanol. H1F7 and F1H7 have similar ΔG_{water} of -1.6 and -1.4 kcal/mol, respectively, which is a result of competing changes in the Lennard-Jones and Coulombic interactions. Compared to 1-octanol, fluorinating the last seven carbon atoms (H1F7), increases the Lennard-Jones contribution to ΔG_{water} by 0.9 kcal/mol, while the Coulombic interaction is decreased by 0.8 kcal/mol. Fluorinating C_α only (F1H7) results in a 1.1 kcal/mol decrease in the Coulombic contribution, with a 0.4 kcal/mol increase in the Lennard-Jones contribution to the free energy of hydration.

Table 6.4: Contribution of Lennard-Jones and Coulombic energy to the free energies of hydration/solvation predicted by MBAR[257]. Numbers in parenthesis correspond to the uncertainty in the last digit.

Molecule	$\Delta G_{1-octanol}$(kcal/mol)			ΔG_{water}(kcal/mol)		
	LJ	Coulomb	Total	LJ	Coulomb	Total
$CH_3(CH_2)_7OH$ (H8)	-4.84(7)	-3.8(2)	-8.6(2)	3.6(2)	-6.47(9)	-2.9(2)
$CF_3(CF_2)_5(CH2)_2OH$ (H2F6)	-3.85(7)	-3.3(2)	-7.1(2)	4.8(2)	-6.48(9)	-1.7(2)
$CF_3(CF_2)_6CH_2OH$ (H1F7)	-3.5(1)	-2.6(2)	-6.0(2)	4.4(2)	-6.08(9)	-1.6(2)
$CF_3(CF_2)_7OH$ (F8)	-3.06(9)	-2.1(1)	-5.2(2)	5.4(3)	-5.30(8)	0.0(3)
$CH_3(CH_2)_6CF_2OH$ (F1H7)	-4.15(6)	-1.9(1)	-6.1(2)	4.0(2)	-5.4(1)	-1.4(2)
$CH_3(CH_2)_5(CF_2)_2OH$ (F2H6)	-3.84(6)	-1.8(2)	-5.7(2)	4.3(2)	-5.59(9)	-1.3(2)

It should be noted that a key difference between this work and past calculations with NAMD[37] for the same molecules and models is that in this work long-range corrections for Lennard-Jones interactions are included in the free energy calculation, whereas, in past work, they were not[114]. In preliminary calculations, the contribution of long-range corrections to

the free energy of hydration for these molecules was found to be approximately -1.0 to -0.8 kcal/mol, which is consistent with prior calculations for n-alkanes[271]. Accounting for this difference in the treatment of long-range corrections to the Lennard-Jones interactions brings the results shown in Table 6.3 in good agreement with prior calculations[114]. Inclusion of long-range corrections for the Lennard-Jones interactions substantially improves the agreement of ΔG_{water} predictions of simulation with experiment for 1-octanol, but makes agreement with the most reliable experimental data for H2F6 worse[268].

To further understand how fluorination near the head group affects the solubility of alcohols in water, radial distribution functions (RDF) were calculated for O(solute)-O(solvent) and C_α(solute)-O(solvent), and are presented in Figure 6.5. For all molecules, for O(solute)-O(solvent) interactions, a peak is observed at approximately 2.75 Å corresponding to hydrogen bonding between water and the solute. Peak heights varied, depending on the degree of fluorination near the hydroxyl group. Similar peak heights were observed for the O-O RDF for 1-octanol and H2F6 interacting with water, while a slightly lower peak height was observed for H1F7. The lowest peak heights were observed for perfluorooctanol, F1H7, and F2H6, which all have a fluorinated α carbon. For the C_α(solute)-O(water) radial distribution functions, perfluorooctanol, F1H7, and F2H6 all show similar behavior with a first peak at approximately 3.9 Å, while the first peak in the RDF for 1-octanol, H1F7, and H2F6 occurs at 3.7 Å. These results for the O-O and C_α-O RDFs are consistent with prior calculations with the OPLS-AA force field[114], and show clearly that fluorination of C_α creates steric hindrance to solute-solvent hydrogen bond formation, strongly impacting on hydration free energies. These results are consistent with the work of Dalvi and Rossky, which concluded for perfluoroalkanes, that increased hydrophobicity was due to the increased volume occupied by fluorine compared to hydrogen atoms[272].
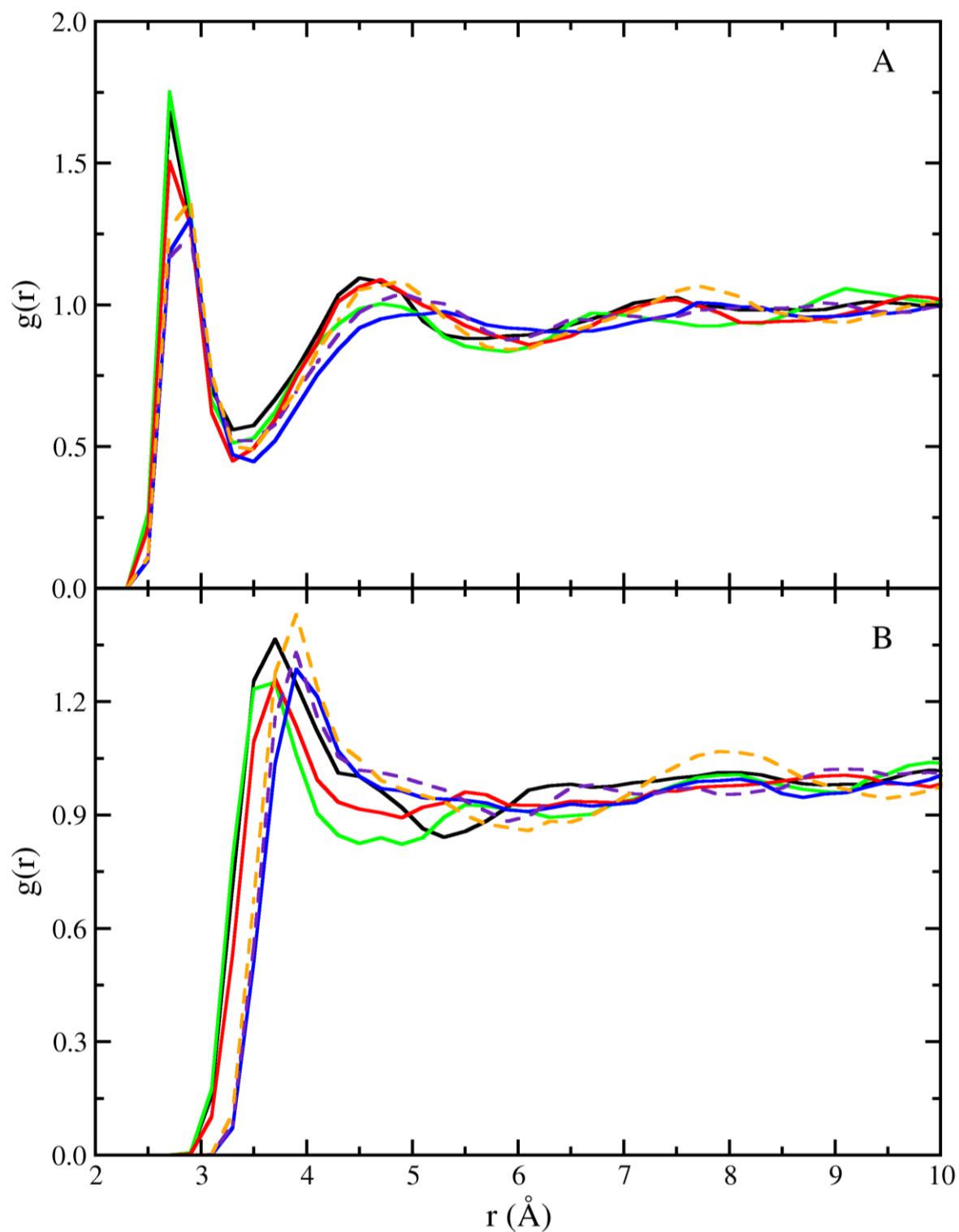
Figure 6.5: Radial distribution function for solute interactions with water: (A) O(solute)-O(water) and (B) C_α(solute)-O(water). Data are represented by: octanol (solid black line), H2F6 (solid green line), H1F7 (solid red line), and perfluorooctanol (solid blue line), F1H7 (dashed orange line), and F2H6 (dashed indigo line).

6.5.2 1-octanol Free Energies of Solvation

Free energies, predicted by the TraPPE-UA force field for each solute in 1-octanol, are listed in Table 6.3 and individual contributions of Lennard-Jones and Coulombic interactions to solvation free energies are listed in Table 6.4. Free energies of solvation for 1-octanol and H2F6 in 1-octanol were found to be in excellent agreement with experiment, with errors of 0.47 and 0.1 kcal/mol, respectively. Calculated free energies of solvation show a monotonic increase (become less negative) as fluorination of the alkyl tail increases. This is similar to the phenomena observed for hydration free energies, though fluorination of the alkyl tail has a larger impact on solvation free energies in octanol than in water, as evidenced by the calculated octanol-water partition coefficients for all fluorinated alcohols being lower than that of 1-octanol, despite also having lower hydration free energies.

The peak height in radial distribution functions for O(solute) with O(1-octanol), shown in Figure 6.6, follow a similar trend as the solvation free energies. The largest peak height was observed for 1-octanol in 1-octanol, while the lowest peak height was for perfluorooctanol. These results suggest that C_α fluorination state plays a significant role in in the predicted free energy, since fluorination near the hydroxyl group sterically hinders the solvent's ability to form hydrogen bonds with the solute. These results were confirmed by additional free energy calculations performed for F1H7 and F2H6. Fluorination of both the α and β carbons (F2H6) produces a free energy of solvation that is within 0.5 kcal/mol of perfluorooctanol, while fluorinating only the α carbon produces a free energy of solvation that is similar to H1F7.

Fluorination of C_α produced a marked decrease in the Coulombic contribution to the free energy. For F1H7, F2H6, and perfluorooctanol, the Coulombic contribution varied from -2.1 to -1.8 kcal/mol, compared to -3.8 kcal/mol for 1-octanol. Unlike solvation in water, fluorination of C_β and later carbons also impacted the hydrogen bonding of solutes with 1-octanol. Coulombic contributions to the free energy decrease with increasing fluorination,

regardless of position on the alkyl tail; for H2F6 ΔG_{Coul} = -3.3 kcal/mol, while for H1F7 $\Delta G_{Coul} = -2.6$ kcal/mol. Radial distribution functions for C_α(solute)-O(solvent) interactions show decreased height of the first peak going from 1-octanol to H2F6 and H1F7. While both water and 1-octanol form complex hydrogen bonded networks, the alkyl tail of 1-octanol creates additional constraints on the microstructures that may form. Adding bulky fluorine atoms to the alkyl tail of solutes, beyond C_α and C_β, appears to be capable of creating steric hindrance to hydrogen bond formation between the solute and the 1-octanol solvent.
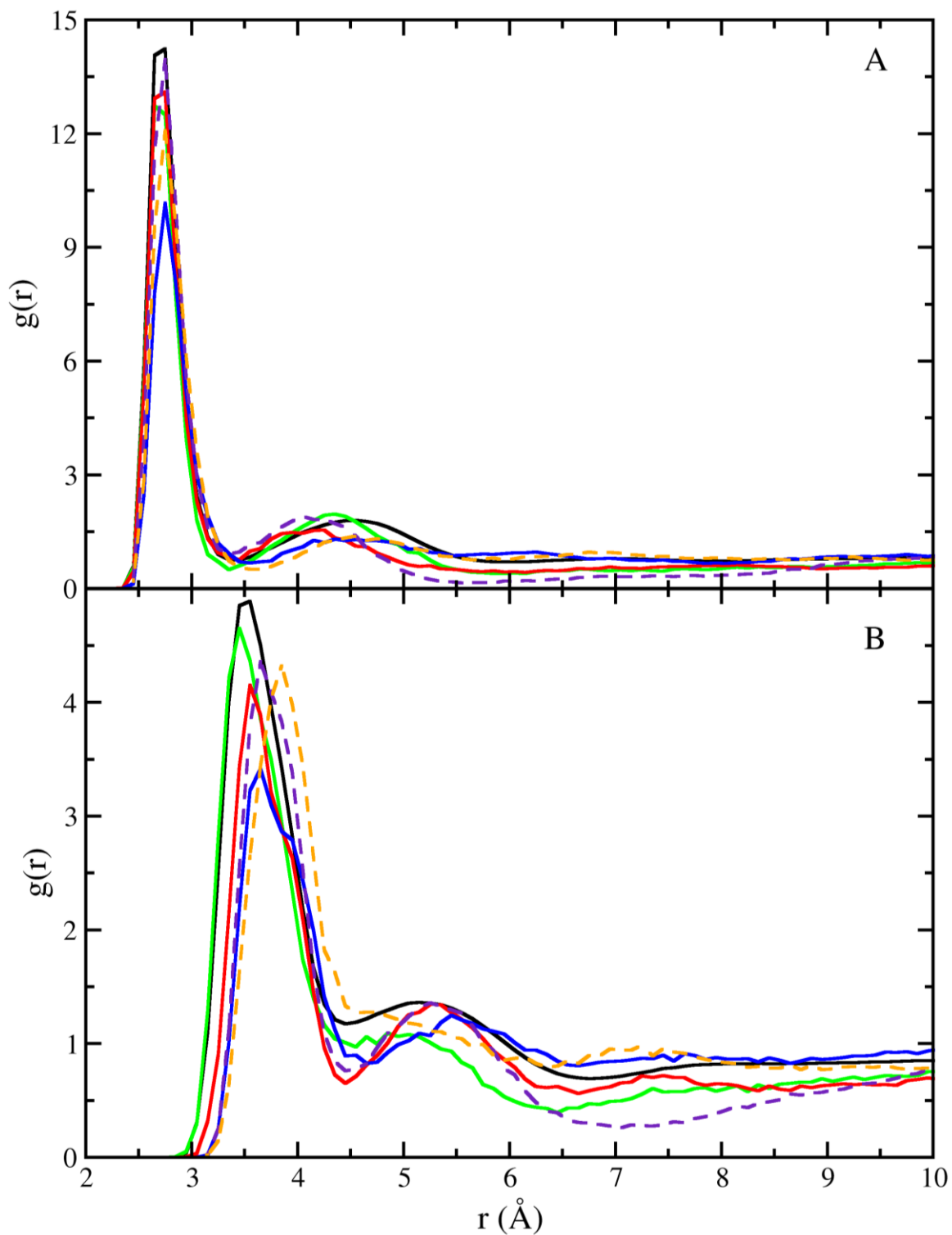
Figure 6.6: Radial distribution function for solute interactions with 1-octanol: (A) O(solute)-O(1-octnaol) and (B) C_α(solute)-O(1-octanol). Data are represented by: octanol (solid black line), H2F6 (solid green line), H1F7 (solid red line), and perfluorooctanol (solid blue line), F1H7 (dashed orange line), and F2H6 (dashed indigo line).

6.5.3 n-hexadecane Free Energies of Solvation

The air-hexadecane partition coefficient provides a measurement of non-specific interactions between molecules and plays an important role as a compound descriptor used in linear solvation energy relationships (LSER). LSER models are used for prediction of solute partitioning in a variety of process, providing data that are needed for transport and environmental fate modeling[231, 232]. Additionally, water-hexadecane partition coefficients are used to model lipophilic systems, such as the core of lipid bilayers[273, 274]. Predicting solvation free energies of fluorinated 1-octanol analogues in n-hexadecane provides additional insight into the role of fluorine in altering Lennard-Jones interactions between the solute and organic solvents, without the complications of hydrogen bonding present in the solvent 1-octanol.

Free energies predicted by the TraPPE-UA force field for each solute in n-hexadecane are listed in Table 6.3. Experimental data for these compounds is limited to 1-octanol and H2F6. For H2F6, simulations predicted $\Delta G_{C16} = -4.16$ kcal/mol, which is in close agreement with the experimental value of -4.0 kcal/mol from Goss *et al.*[268]. For 1-octanol, simulations predict $\Delta G_{C16} = -5.15$ kcal/mol vs. the experimental value of -6.3 kcal/mol.

Interestingly, the data follow the same trend with increasing fluorination as the free energies of hydration $\left(\Delta G_{C16}(\text{H8}) < \Delta G_{C16}(\text{H2F6}) < \Delta G_{C16}(\text{H1F7}) < \Delta G_{C16}(\text{F8})\right)$, despite the absence of specific hydrogen bonding interactions. Fluorinating only C_α (F1H7) produces a 0.77 kcal/mol increase in the free energy of solvation compared to 1-octanol, while fluorination of C_β and C_α (F2H6) produces only an additional 0.04 kcal/mol change in ΔG_{C16}. The free energy change due to fluorination of only C_α accounts for almost half of the difference in ΔG_{C16} between 1-octanol and perfluorooctanol. Fluorination of C_β and later carbons (H2F6) produces only an additional 0.18 kcal/mol change in ΔG_{C16}, as compared to F2H6.

Radial distribution functions for each solute interacting with n-hexadecane are presented in Figure 6.7. For O(solute)-CH_x(n-hexadecane), 1-octanol and H2F6 have similar behavior, while, for all other solutes, the first peak is slightly lower and shifted to larger distances, illustrating the additional space occupied by the fluorine atoms near the hydroxyl group. For the CH_x or CF_x(solute)-CH_x(n-hexadecane) radial distribution functions, the most highly fluorinated molecules, H1F7 and perfluorooctanol, display similar behavior, while a reduction in the number of fluorine atoms (*i.e.* F1H7 and F2H6), causes the first peak in the RDF to shift to smaller distances.
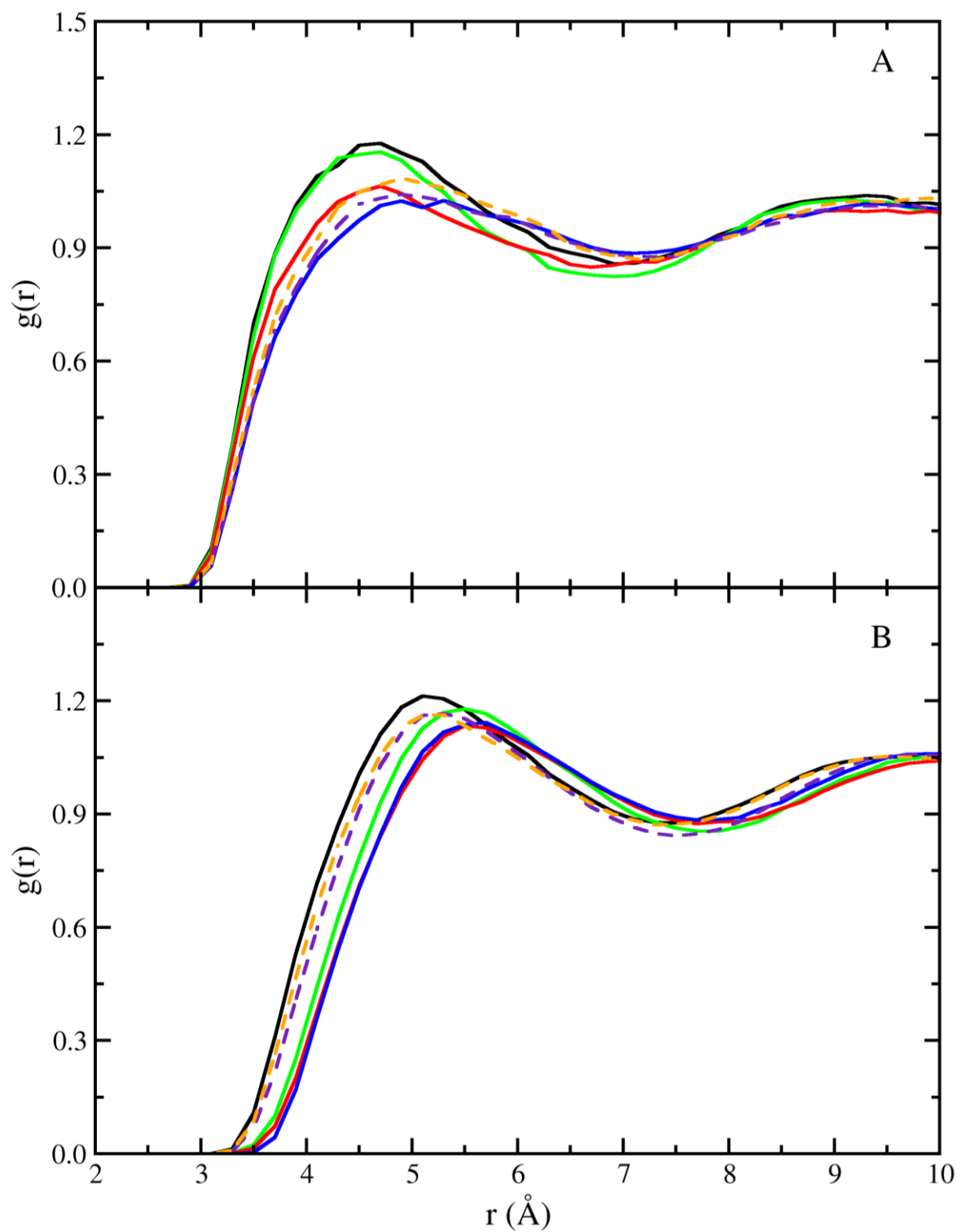
Figure 6.7: Radial distribution function for solute interactions with n-hexadecane: (A) interaction of O(solute)-CH$_x$(n-hexadecane) and (B) CH$_x$ or CF$_x$(solute)-CH$_x$(n-hexadecane). Data are represented by: octanol (solid black line), H2F6 (solid green line), H1F7 (solid red line), and perfluorooctanol (solid blue line), F1H7 (dashed orange line), and F2H6 (dashed indigo line).

CHAPTER 7 CONCLUSIONS

7.1   Mie Potential For Alkynes

In Chapter 3, the Mie potentials were extended to alkynes. Through an exhaustive search of parameter space, it was determined that it was not possible to transfer C(sp) parameters from 1-alkynes to 2-alkynes, and unique C(sp) parameters for these compounds were developed. Predicted critical properties and liquid structure show the expected convergence of alkyne properties to *n*-alkanes with increasing chain length. Overall, the predictions of simulation for 1-alkynes were found to be in reasonable agreement with experiment and correlations with the notable exception of 1-hexyne. Saturated liquid densities for 1-hexyne were under-predicted with an AAD of 3%, while saturated liquid densities for all other 1-alkynes were within 1-2% AAD of experiment. The non-monotonic behavior of the deviation between simulation and the DIPPR correlations suggests possible inconsistencies in the correlations, although resolving this issue will require new experimental measurements to be performed.

Transferability of the Mie potentials was further evaluated through simulations of the binary mixtures of propadiene+propyne, propene+propyne, and propadiene+propyne. The phase behavior of propadiene+propyne was in close agreement with experiment. Predictions for mixtures of propyne with propene or propane under-predicted mixture vapor pressures indicating that unlike molecule interactions were over-predicted by 2-4%.

7.2   Molecular Exchange Monte Carlo in GCMC Simulation

In Chapter 4, three variants of the molecular exchange method were developed, which could be used to evaluate the efficiency of various aspects of the algorithms. Locating the exchange sub-volume randomly (ME-1) was found to have the lowest efficiency, since frequently, no small molecules were found in the sub-volume that could be used for the

molecular exchange, resulting in immediate rejection of the move. The ME-1 method is suitable only for systems that are very dilute with respect to the concentration of the large molecule. By identifying a small molecule at random first, placing the center of the sub-volume at the geometric center of the small molecule (ME-2), and aligning the backbone of the large molecule to be inserted with the small molecule to be removed, acceptance rates for the exchange move increased substantially. For water, the acceptance efficiency of the ME-2 method was found to be nearly 40 times greater than standard configurational-bias insertions, while for 2,2,4-trimethylpentane a 410 times improvement in acceptance efficiency was achieved. In the latter case, this was due to the use of a rigid-body insertion in ME-2, which eliminated the need to regrow the molecule in place. Finally, the inclusion of coupled-decoupled configurational-bias methods[67] to grow sections of the molecule from a central atom (ME-3) placed at the center of the sub-volume resulted in the greatest improvement in statistical efficiency compared to standard configurational-bias insertions for linear molecules without strong directional interactions. Improvements in efficiency of up to 200 times were observed for the perfluorobutane+n-butane system.

The algorithms presented in this work are notable because they were designed to work for any molecular topology over a wide range of compositions. Substantial performance gains were observed for ME-2 and ME-3 for all systems and compositions studied. As shown through the various case studies, however, each method has its strengths and weaknesses. For linear non-polar molecules, ME-3 is generally more efficient than ME-2, while ME-2 offers better performance for small polar molecules, such as water, and highly branched molecules. Each algorithm has been implemented, and is now available, in the open-source Monte Carlo simulation engine GOMC, which is available to the public at GitHub[170].

7.3   Molecular Exchange Monte Carlo in GEMC Simulation

In Chapter 5, the molecular exchange Monte Carlo (MEMC) method has been adapted for use in Gibbs ensemble Monte Carlo simulations. Calculations of pressure-composition diagrams for methane+n-butane and perfluorobutane+n-butane show exact agreement with prior grand canonical Monte Carlo simulations[83]. The combination of GEMC and MEMC was used to predict the free energies of transfer for n-alkanes in 1-octanol, n-hexadecane and 2,2,4-trimethylpentane. In comparison to more traditional methods for the calculation of free energies of solvation (thermodynamic integration in molecular dynamics), the GEMC-MEMC method shows similar computational efficiency. The GEMC-MEMC method has some potential advantages over molecular dynamics simulations for calculation of free energies of solutes which have large energy barriers between conformers. These solutes require either biased sampling techniques or very long molecular dynamics simulations to sample all relevant states[115]. In the GEMC-MEMC, the coupled-decoupled configurational-bias algorithm allows the simulation to rapidly jump between minimum energy conformers, leading to faster sampling of the relevant conformational space.

Free energy calculations for alkanes-1-octanol were performed with only the TraPPE force field, and were in excellent agreement with prior simulations and experimental data. For n-alkane solvation in n-hexadecane and 2,2,4-trimethylpentane, simulations were performed with both the TraPPE and Mie potentials. The Mie potentials were found to offer superior performance compared to TraPPE, being in close agreement with experimental data for all solutes from methane to n-octane. TraPPE displayed good agreement with experiment for n-alkane solutes with four or fewer carbons, but for larger n-alkanes, TraPPE under-predicted free energies of transfer with the difference increasing with solute size.

7.4   Partitioning of Fluorinated Alcohols

In Chapter 6, free energies of solvation in water, 1-octanol, and n-hexadecane were calculated with Monte Carlo simulations in the isobaric-isothermal ensemble for a variety of fluorinated analogues of 1-octanol. The combination of SPC water and TraPPE-UA were found to provide a good qualitative reproduction of experimental data.

Davli and Rossky concluded that the molecular basis for hydrophobicity exhibited by perfluoroalkanes was due to the larger volume occupied by fluorine compared to hydrogen atoms[272]. Similarly, this work has shown that the larger volume of fluorine atoms compared to hydrogen leads to the oleophobic behavior of fluoroalcohols. Fluorination of the α and β carbons was found to have the greatest impact on the free energy of hydration and the free energy of solvation in 1-octanol. The addition of fluorine atoms to the alpha and beta carbons creates a steric hindrance to hydrogen bonding between the solute and the solvent. In 1-octanol and n-hexadecane, subtle effects of fluorination of methyl groups further away from the hydroxyl group on hydrogen bonding were observed. Down-chain fluorination increases the volume occupied by the solute, while intramolecular geometrical constraints and barriers to dihedral rotation limit the ability of 1-octanol to reorient to form hydrogen bonds with the solute. In n-hexadecane, reductions in the free energy of solvation with fluorination are largely due to increases volume occupied by fluorine atoms and their lower energy density.

APPENDIX A

In this section, additional results with their numerical values for Mie potential in alkynes are provided.
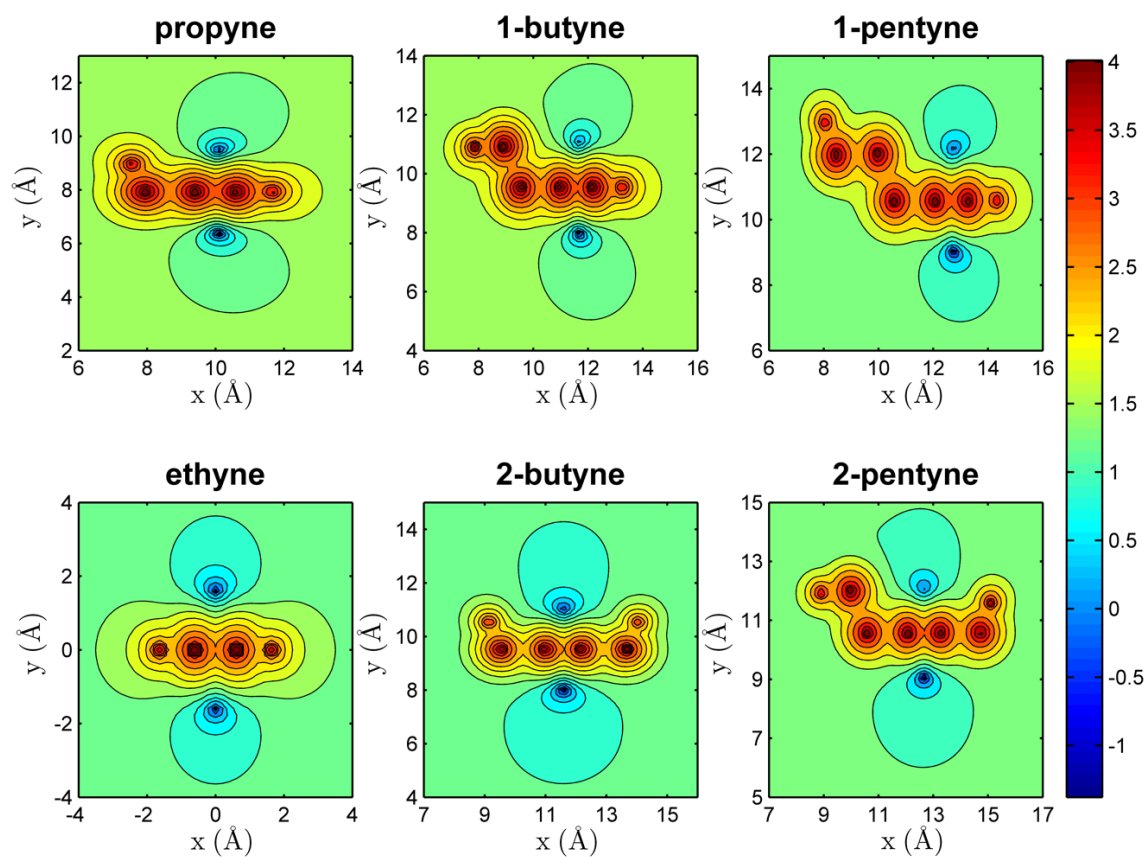
Figure A1: Electrostatic potential isosurfaces (ESP) determined from HF/6-31+g (d, p) ab initio calculations using Gaussian 09[136] in the plane formed by carbons in the molecule's back bone.
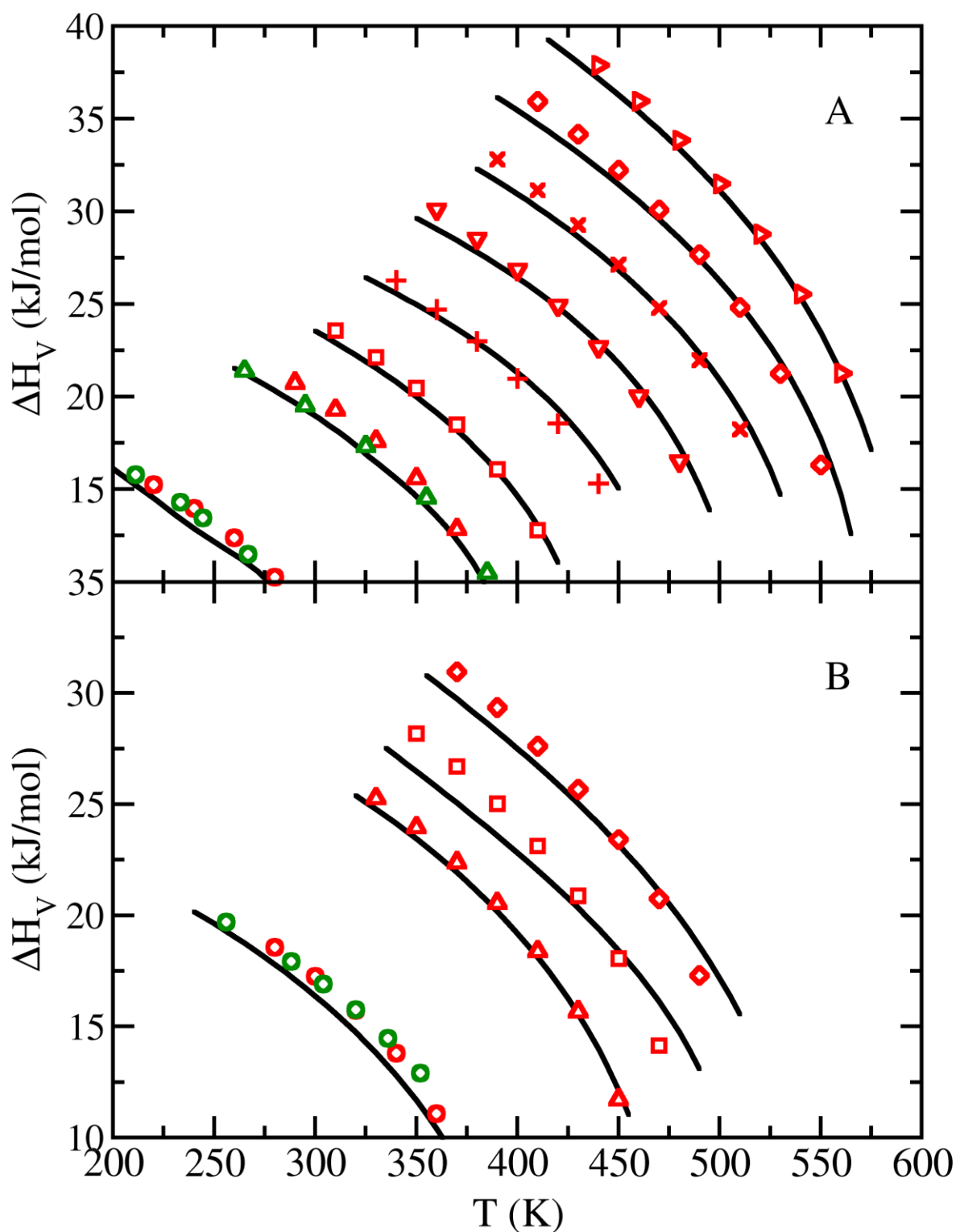
Figure A2: Heats of vaporization predicted by the optimized Mie potentials (red symbols) and 2CLJQ model (green symbols)[61] compared to experiment or correlation (solid line)[149, 151, 154] for alkynes and propadiene. Figure S2A: ethyne (circles); propyne (triangles up); 1-butyne (squares); 1-pentyne (plus); 1-hexyne (triangles down); 1-heptyne (crosses); 1-octyne (diamonds); 1-nonyne (triangles right); Figure S2B: propadiene (circles); 2-butyne (triangles up); 2-pentyne (squares); 2-hexyne (diamonds).
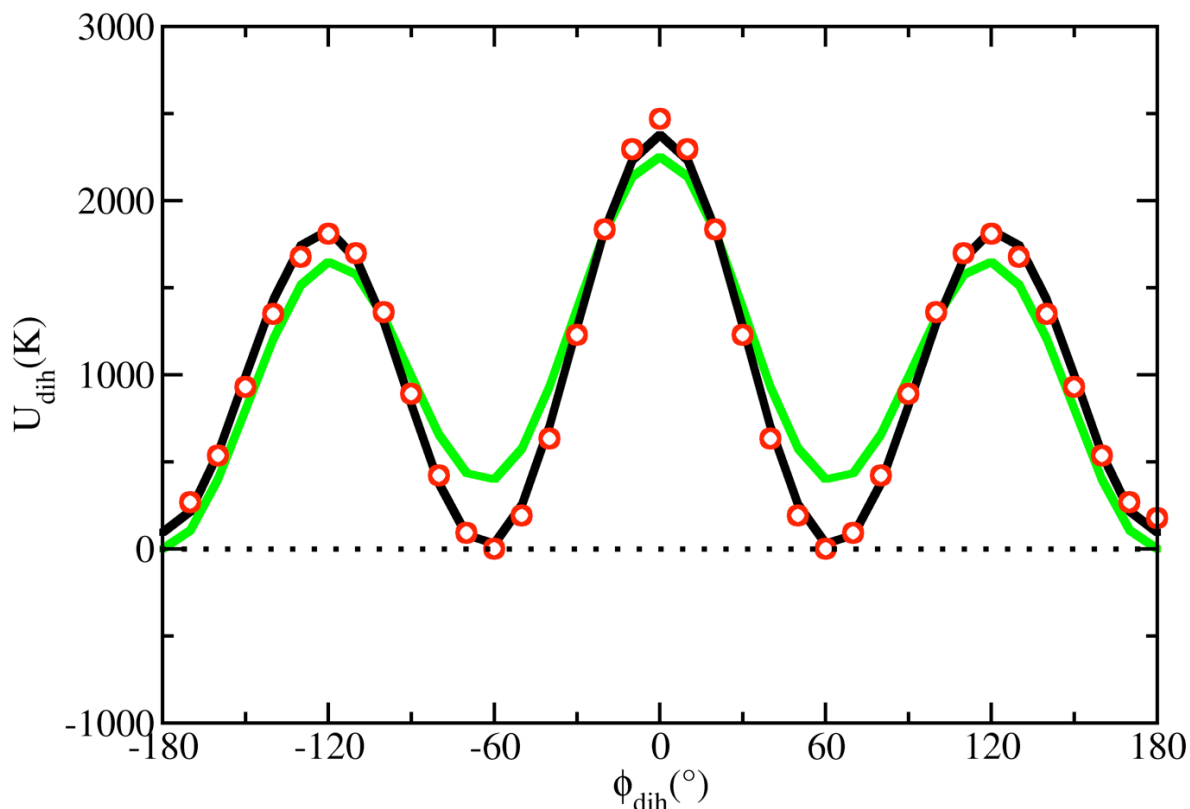
Figure A3: Rotational barrier for the CHx–CH2–CH2–C(sp) torsion in alkynes. Predictions of MP2/aug-cc-PVTZ ab initio calculations (red circles); fit of cosine series (black solid line); OPLS cosine series for CHx–CH2–CH2– CHx torsion[139, 140] (green solid line).

Effect of electrostatic interactions on the phase behavior and structure of short alkynes:

Alkynes have a small dipole moment of approximately 0.7D, while ethyne has a quadrupole moment of 20.4×10^{-40} C-m^2[157]. To understand the impact of neglecting the small dipole and quadrupole moments on the predictive capability of the force field, additional parameters were optimized for ethyne and propyne models that included point charges. Partial charges for ethyne and propyne were determined from a CHELPG[275, 276] analysis of *ab initio* potential energy surfaces, generated with MP2 theory and the aug-cc-PVTZ basis set. *Ab initio* calculations were performed with Gaussian 09[136]. The resulting partial charges produce quadrupole and dipole moments that are in close agreement with experiment. Partial charges were placed on the nuclei of the hydrogen and carbon atoms. Then, Lennard-Jones parameters were optimized to reproduce experimental vapor pressures and saturated liquid densities. The resulting partial charges and Lennard-Jones parameters are listed in Table A1.

The hydrogen-carbon bond length was set at 1.062 Å, which was taken from prior empirical force fields[62] and *ab initio* calculations[138].

Table A1: Non-bonded parameters for ethyne and propyne force fields with electrostatic interactions.

molecule	pseudo-atom	$\varepsilon_i/k_b(K)$	σ_i (Å)	n_i	q_i
ethyne	H	0.0	0.0	-	0.26
	CH(sp)	87.50	3.590	16	-0.26
propyne	H	0.0	0.0	-	0.30
	CH(sp)	87.50	3.590	16	-0.40
	C(sp)	188.00	2.960	16	0.10
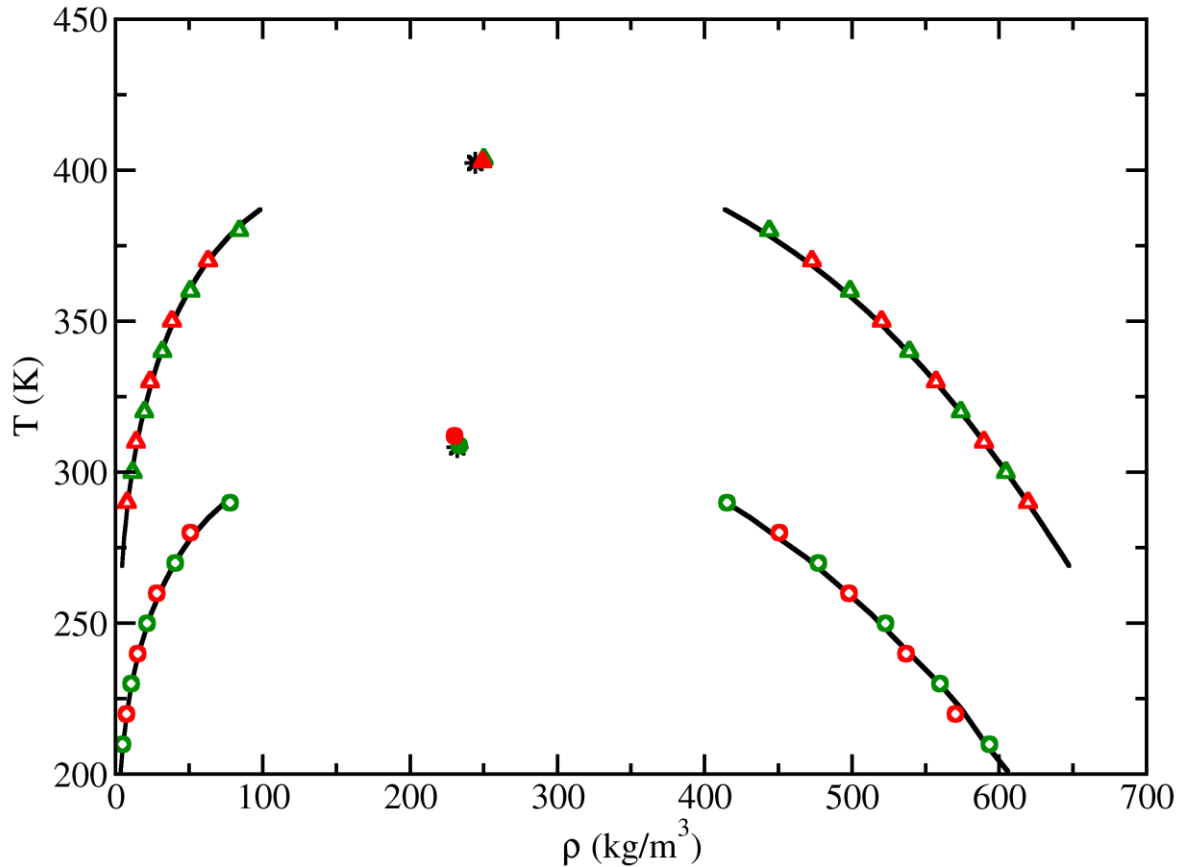
Figure A4: Vapor-liquid coexistence curves predicted by the optimized Mie potentials without electrostatic interactions (red symbols) and with electrostatic interactions (green symbols), compared to experiment (solid line)[149] for ethyne and propyne. Experimental critical points (black stars)[151, 155] and predictions of simulation (filled symbols). Predictions of simulation are represented as ethyne (circles) and propyne (triangles).
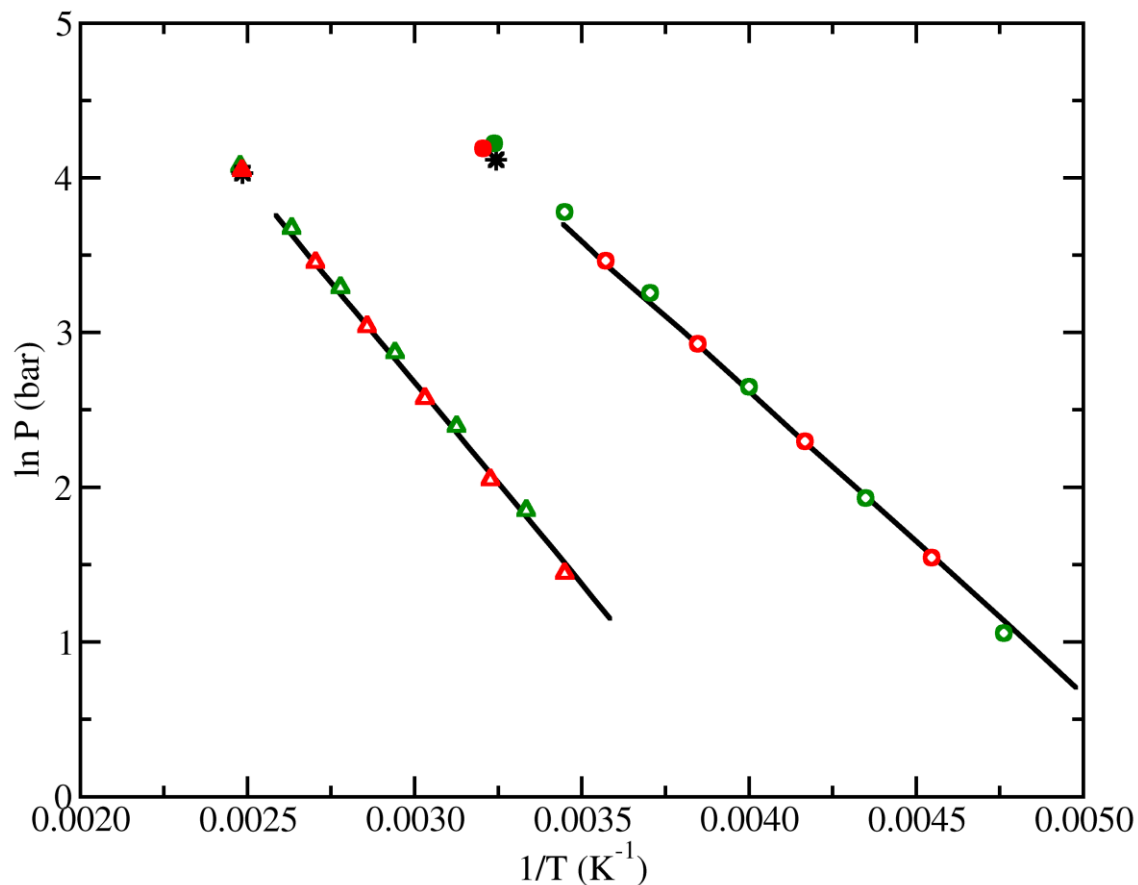
Figure A5: Clausius-Clapeyron plots predicted by the optimized Mie potentials without electrostatic interactions (red symbols) and with electrostatic interactions (green symbols), compared to experiment (solid line)[149] for ethyne and propyne. Experimental critical points (black stars)[151, 155] and predictions of simulation (filled symbols). Predictions of simulation are represented as ethyne (circles) and propyne (triangles).
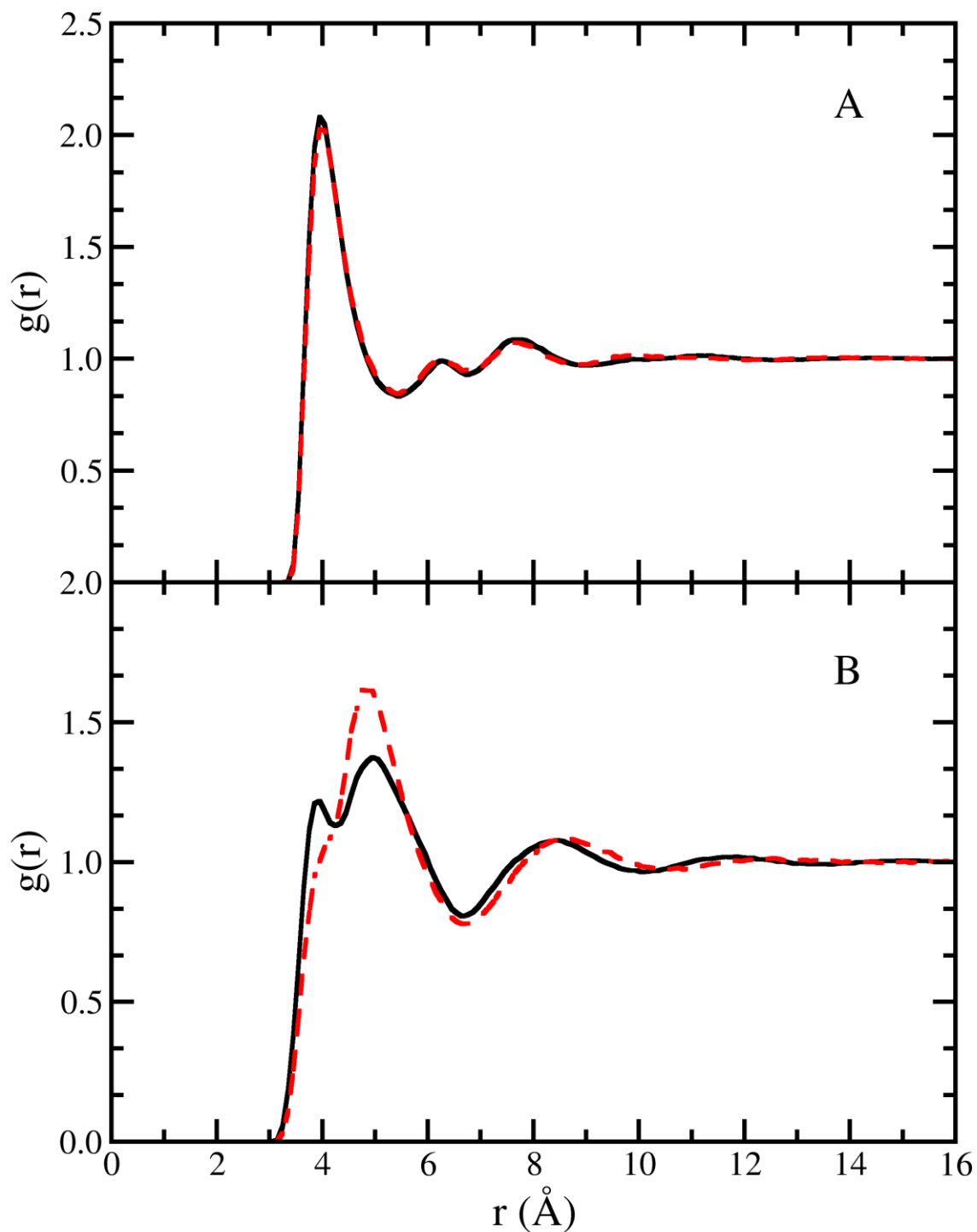
Figure A6: Comparison of radial distribution functions for propyne force fields with (dashed red) and without (solid black) explicit electrostatic interactions. Figure A6A: radial distribution functions for CH3-CH3 interactions; Figure A6B: radial distribution functions for C-C interactions.
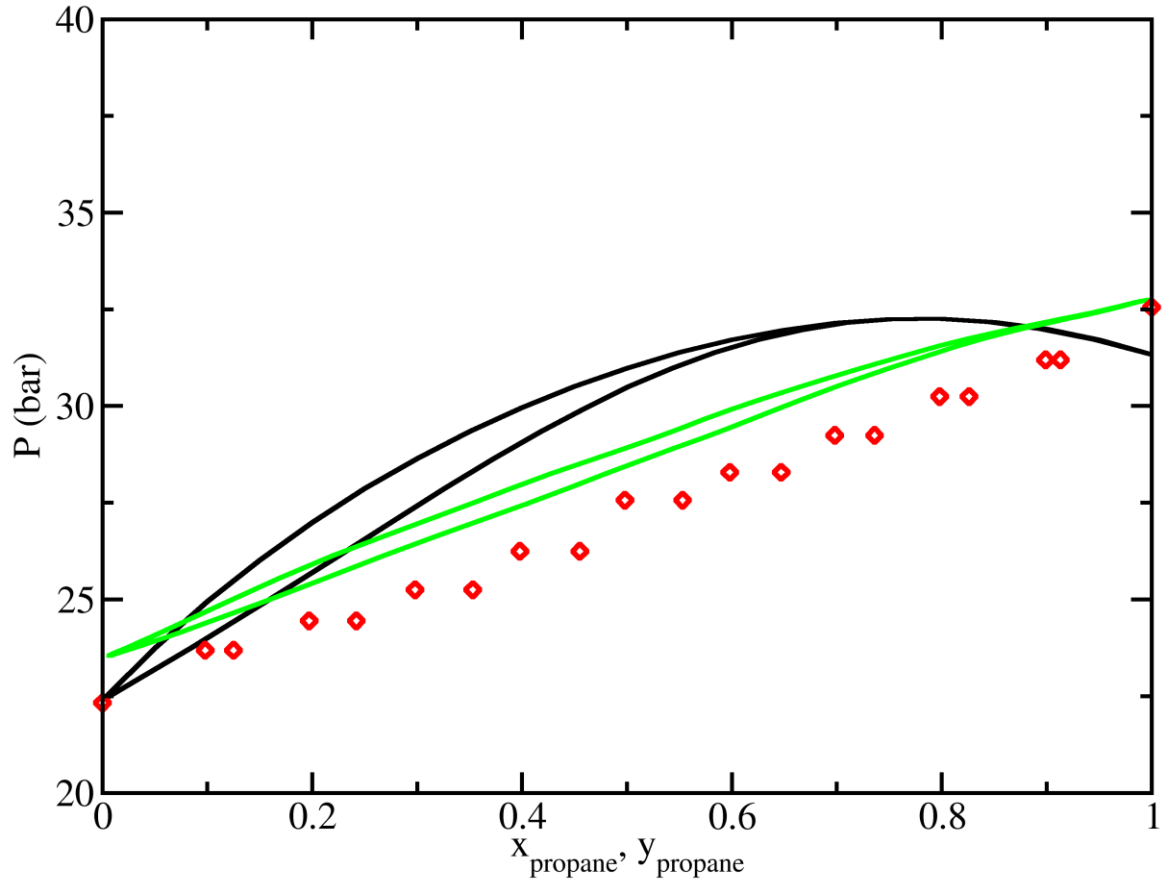
Figure A7: Pressure-composition diagram for propane+propyne at 353.15 K. Data are represented by: Experiment (black lines)[159], Mie potentials without electrostatic interactions (red symbols) and with electrostatic interactions (green lines).

Table A2: Vapor-liquid coexistence data predicted by the optimized Mie potentials for ethyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
290	0.420(1)	0.0693(2)	40.84(6)	8.81(2)	0.636(3)
280	0.4505(5)	0.05068(3)	32.00(8)	10.26(1)	0.706(2)
270	0.4755(5)	0.03767(6)	24.69(7)	11.40(1)	0.760(3)
260	0.4980(5)	0.02795(3)	18.68(6)	12.38(1)	0.805(3)
250	0.5182(6)	0.02054(2)	13.81(6)	13.23(2)	0.843(4)
240	0.5366(8)	0.01483(4)	9.95(5)	13.97(3)	0.875(6)
230	0.5537(5)	0.01047(3)	6.95(5)	14.64(2)	0.904(8)
220	0.570(1)	0.00718(2)	4.69(5)	15.26(4)	0.93(1)

Table A3: Vapor-liquid coexistence data predicted by the optimized Mie potentials for propyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
380	0.4417(9)	0.0821(2)	38.39(6)	10.99(2)	0.593(2)
370	0.4728(2)	0.0626(2)	31.63(5)	12.84(2)	0.658(3)
360	0.4982(1)	0.0486(1)	25.84(4)	14.34(1)	0.712(3)
350	0.5200(3)	0.03808(9)	20.89(3)	15.586(9)	0.755(2)
340	0.5394(7)	0.02985(5)	16.68(2)	16.65(2)	0.792(2)
330	0.5569(7)	0.02329(4)	13.14(2)	17.60(3)	0.824(2)
320	0.5734(8)	0.01801(3)	10.19(2)	18.46(4)	0.852(2)
310	0.590(1)	0.01375(3)	7.77(2)	19.28(5)	0.878(3)
300	0.605(1)	0.01035(2)	5.80(1)	20.04(4)	0.901(3)
290	0.620(1)	0.00765(1)	4.24(1)	20.73(4)	0.921(3)

Table A4: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 1-butyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
410	0.4459(4)	0.0762(3)	29.58(5)	12.795(7)	0.616(4)
400	0.4728(5)	0.0594(2)	24.65(3)	14.61(1)	0.674(3)
390	0.4953(4)	0.0471(1)	20.37(2)	16.083(9)	0.721(2)
380	0.5155(2)	0.03759(6)	16.68(1)	17.36(1)	0.760(2)
370	0.5341(4)	0.02994(5)	13.51(1)	18.50(3)	0.793(2)
360	0.5513(7)	0.02373(4)	10.80(1)	19.53(3)	0.823(2)
350	0.5672(8)	0.01866(4)	8.52(1)	20.46(4)	0.849(2)
340	0.5823(5)	0.01452(3)	6.62(1)	21.33(2)	0.872(3)
330	0.5964(4)	0.01116(2)	5.06(1)	22.13(2)	0.893(3)
320	0.6099(8)	0.00845(2)	3.79(1)	22.87(4)	0.912(3)
310	0.623(1)	0.00629(2)	2.780(8)	23.57(5)	0.928(4)

Table A5: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 2-butyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
450	0.434(2)	0.096(2)	36.3(1)	11.7(1)	0.55(1)
440	0.466(1)	0.0747(5)	30.73(5)	13.90(4)	0.609(5)
430	0.491(1)	0.0592(1)	25.88(4)	15.67(3)	0.662(2)
420	0.5120(6)	0.04754(8)	21.65(3)	17.11(3)	0.705(2)
410	0.5311(5)	0.03838(5)	17.96(3)	18.38(3)	0.743(2)
400	0.5490(5)	0.03097(4)	14.77(2)	19.52(3)	0.776(2)
390	0.5654(3)	0.02491(4)	12.03(2)	20.54(2)	0.806(2)
380	0.5809(2)	0.01992(4)	9.69(2)	21.49(2)	0.832(2)
370	0.5958(5)	0.01582(3)	7.70(2)	22.38(2)	0.856(2)
360	0.6100(9)	0.01244(1)	6.04(1)	23.20(3)	0.878(2)
350	0.6229(8)	0.00967(1)	4.67(1)	23.95(3)	0.897(3)
340	0.6346(7)	0.00742(2)	3.55(1)	24.62(3)	0.915(4)
330	0.6464(7)	0.00561(2)	2.65(1)	25.26(3)	0.930(5)

Table A6: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 1-pentyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
450	0.434(2)	0.086(1)	27.32(4)	13.1(1)	0.58(1)
440	0.4623(5)	0.0669(4)	23.03(2)	15.31(6)	0.641(4)
430	0.4855(4)	0.0531(2)	19.31(2)	17.08(3)	0.693(2)
420	0.5055(6)	0.04260(7)	16.07(2)	18.55(2)	0.736(2)
410	0.5236(6)	0.03433(4)	13.26(2)	19.82(3)	0.772(2)
400	0.5403(6)	0.02764(3)	10.84(2)	20.98(3)	0.803(2)
390	0.5559(5)	0.02215(3)	8.77(2)	22.03(2)	0.831(2)
380	0.5705(2)	0.01763(3)	7.01(1)	22.99(1)	0.857(2)
370	0.5841(7)	0.01391(3)	5.53(1)	23.88(3)	0.880(3)
360	0.597(1)	0.01086(2)	4.30(1)	24.71(6)	0.901(3)
350	0.610(1)	0.00838(1)	3.29(1)	25.51(5)	0.919(3)
340	0.6229(7)	0.006371(9)	2.47(1)	26.28(3)	0.935(4)

Table A7: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 2-pentyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
470	0.4446(5)	0.0825(5)	27.52(7)	14.14(5)	0.581(5)
460	0.4728(6)	0.0651(3)	23.32(5)	16.28(3)	0.638(4)
450	0.4961(4)	0.0522(2)	19.65(4)	18.056(8)	0.686(3)
440	0.5162(3)	0.0422(1)	16.44(3)	19.559(9)	0.726(3)
430	0.5345(4)	0.03418(8)	13.65(2)	20.88(2)	0.761(3)
420	0.5511(5)	0.02767(7)	11.23(2)	22.06(3)	0.791(3)
410	0.5663(7)	0.02231(6)	9.14(2)	23.13(4)	0.819(3)
400	0.5805(7)	0.01789(4)	7.37(1)	24.10(4)	0.844(3)
390	0.5942(6)	0.01424(3)	5.87(1)	25.02(4)	0.866(3)
380	0.6074(6)	0.01123(2)	4.61(1)	25.89(3)	0.886(3)
370	0.6201(8)	0.008755(7)	3.573(9)	26.70(4)	0.904(3)
360	0.6322(9)	0.006740(6)	2.724(9)	27.46(5)	0.920(3)
350	0.6439(9)	0.005113(8)	2.040(8)	28.18(5)	0.934(4)

Table A8: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 1-hexyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
490	0.426(1)	0.0891(3)	25.33(8)	14.10(4)	0.573(3)
480	0.4542(4)	0.0701(2)	21.63(7)	16.51(2)	0.635(3)
470	0.4773(4)	0.0564(1)	18.37(6)	18.44(2)	0.685(3)
460	0.4971(3)	0.04596(7)	15.51(6)	20.03(2)	0.725(3)
450	0.5149(3)	0.03759(7)	12.99(5)	21.42(2)	0.759(4)
440	0.5312(5)	0.03072(7)	10.80(4)	22.68(3)	0.789(4)
430	0.5466(6)	0.02503(6)	8.89(4)	23.85(4)	0.816(4)
420	0.5611(5)	0.02028(5)	7.25(3)	24.92(4)	0.841(5)
410	0.5748(4)	0.01633(4)	5.84(3)	25.92(3)	0.863(5)
400	0.5878(4)	0.01303(3)	4.66(3)	26.86(2)	0.882(6)
390	0.6001(7)	0.01030(3)	3.66(2)	27.73(4)	0.900(7)
380	0.612(1)	0.00805(3)	2.83(2)	28.54(7)	0.915(8)
370	0.623(1)	0.00621(3)	2.16(2)	29.33(7)	0.93(1)
360	0.635(1)	0.00472(3)	1.62(2)	30.10(6)	0.94(1)

Table A9: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 2-hexyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
500	0.437(1)	0.0853(5)	24.90(6)	14.97(4)	0.577(5)
490	0.4641(4)	0.06757(8)	21.27(6)	17.30(1)	0.635(2)
480	0.4865(4)	0.05462(7)	18.08(5)	19.18(1)	0.681(3)
470	0.5057(7)	0.04461(8)	15.28(5)	20.75(3)	0.720(3)
460	0.5232(8)	0.03654(8)	12.81(5)	22.14(4)	0.753(4)
450	0.5396(5)	0.02990(7)	10.66(5)	23.41(2)	0.783(4)
440	0.5550(4)	0.02439(5)	8.80(4)	24.59(2)	0.810(5)
430	0.5694(3)	0.01980(3)	7.19(4)	25.67(2)	0.834(5)
420	0.5831(4)	0.01597(2)	5.81(4)	26.67(2)	0.856(6)
410	0.596(1)	0.01278(2)	4.64(3)	27.61(6)	0.875(7)
400	0.608(1)	0.01013(3)	3.66(3)	28.50(9)	0.893(8)
390	0.621(1)	0.00794(3)	2.85(3)	29.35(8)	0.91(1)
380	0.6329(4)	0.00614(4)	2.18(2)	30.19(3)	0.92(1)
370	0.6443(6)	0.00469(4)	1.64(2)	30.96(3)	0.94(1)

Table A10: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 1-heptyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
520	0.4295(8)	0.0859(8)	22.25(5)	15.74(8)	0.576(7)
510	0.4550(4)	0.0680(4)	19.08(4)	18.24(5)	0.636(5)
500	0.4767(2)	0.0550(2)	16.29(4)	20.27(3)	0.685(4)
490	0.49567(9)	0.0450(1)	13.83(3)	21.97(2)	0.725(3)
480	0.5127(1)	0.0370(1)	11.66(3)	23.46(3)	0.759(4)
470	0.5282(2)	0.0305(1)	9.76(3)	24.79(3)	0.788(4)
460	0.5426(3)	0.02503(9)	8.11(3)	26.00(3)	0.814(4)
450	0.5564(3)	0.02046(6)	6.67(3)	27.13(3)	0.838(5)
440	0.5700(3)	0.01663(4)	5.44(3)	28.23(3)	0.860(5)
430	0.5831(2)	0.01341(2)	4.38(3)	29.27(1)	0.879(6)
420	0.5955(3)	0.01071(2)	3.49(2)	30.24(3)	0.897(7)
410	0.6072(5)	0.00848(2)	2.74(2)	31.14(4)	0.913(8)
400	0.6182(5)	0.00663(3)	2.13(2)	31.99(4)	0.93(1)
390	0.6290(3)	0.00512(3)	1.62(2)	32.81(2)	0.94(1)

Table A11: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 1-octyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
550	0.418(1)	0.0893(4)	20.84(6)	16.31(8)	0.563(4)
540	0.4446(7)	0.0713(4)	17.96(5)	18.96(6)	0.618(5)
530	0.4672(5)	0.0576(3)	15.42(4)	21.24(5)	0.669(5)
520	0.4866(4)	0.0472(2)	13.17(4)	23.16(5)	0.711(4)
510	0.5039(2)	0.0389(2)	11.19(3)	24.82(3)	0.746(4)
500	0.5199(2)	0.0322(1)	9.44(3)	26.31(1)	0.777(4)
490	0.5346(4)	0.02657(8)	7.90(2)	27.67(2)	0.805(4)
480	0.5483(1)	0.02187(6)	6.57(2)	28.92(1)	0.829(4)
470	0.5613(5)	0.01791(5)	5.41(2)	30.08(4)	0.851(4)
460	0.5738(4)	0.01459(4)	4.41(1)	31.18(4)	0.871(4)
450	0.5858(1)	0.01180(4)	3.56(1)	32.22(1)	0.889(5)
440	0.5975(5)	0.00946(3)	2.84(1)	33.22(3)	0.905(5)
430	0.6088(6)	0.00751(3)	2.241(9)	34.17(4)	0.919(5)
420	0.6197(4)	0.00590(3)	1.742(8)	35.07(2)	0.932(6)
410	0.6300(2)	0.00457(2)	1.334(6)	35.93(2)	0.943(6)

Table A12: Vapor-liquid coexistence data predicted by the optimized Mie potentials for 1-nonyne.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
570	0.4294(9)	0.0814(8)	17.73(6)	18.50(9)	0.571(7)
560	0.4521(7)	0.0646(3)	15.27(5)	21.27(5)	0.630(5)
550	0.4724(6)	0.0523(2)	13.11(4)	23.60(3)	0.681(4)
540	0.4904(5)	0.0430(1)	11.20(4)	25.54(3)	0.721(4)
530	0.5067(5)	0.0355(1)	9.51(3)	27.23(4)	0.754(4)
520	0.5219(4)	0.0294(1)	8.03(3)	28.77(3)	0.784(5)
510	0.5362(3)	0.02435(9)	6.73(3)	30.18(3)	0.810(5)
500	0.5496(3)	0.02007(7)	5.60(2)	31.49(3)	0.833(5)
490	0.5621(3)	0.01647(6)	4.61(2)	32.70(2)	0.854(5)
480	0.5741(3)	0.01344(5)	3.77(2)	33.84(2)	0.873(6)
470	0.5856(3)	0.01089(4)	3.05(2)	34.92(2)	0.890(6)
460	0.5967(2)	0.00875(3)	2.44(1)	35.95(2)	0.905(6)
450	0.6075(2)	0.00697(3)	1.93(1)	36.93(2)	0.919(7)
440	0.6180(3)	0.00549(2)	1.50(1)	37.89(3)	0.931(8)
430	0.6283(4)	0.00427(2)	1.157(8)	38.81(4)	0.942(8)
420	0.6383(5)	0.00328(2)	0.876(7)	39.70(4)	0.951(9)

Table A13: Vapor-liquid coexistence data predicted by the optimized Mie potentials for propadiene.

T (K)	ρ_l (g/cm^3)	ρ_v (g/cm^3)	P (bar)	ΔH_v (kJ/mol)	Z
360	0.4561(4)	0.0704(2)	33.19(3)	11.08(2)	0.631(2)
350	0.4831(4)	0.0543(1)	27.26(2)	12.58(2)	0.691(2)
340	0.5058(4)	0.04256(6)	22.17(2)	13.79(1)	0.738(1)
330	0.5260(2)	0.03347(4)	17.82(1)	14.819(6)	0.777(1)
320	0.5443(3)	0.02623(3)	14.13(2)	15.72(2)	0.811(2)
310	0.5613(7)	0.02041(2)	11.04(2)	16.53(3)	0.841(2)
300	0.577(1)	0.01571(2)	8.48(2)	17.26(4)	0.867(2)
290	0.592(1)	0.01192(2)	6.39(2)	17.93(4)	0.890(3)
280	0.6065(9)	0.00890(2)	4.71(2)	18.56(3)	0.911(4)

Table A14: Deviation of vapor pressures and saturated liquid densities predicted by the optimized Mie potentials from experiment and correlations[149, 151, 154].

compound	$\%_{ERR} P$				$\%_{ERR} \rho_L$			
	min.	max.	avg.	med.	min.	max.	avg.	med.
ethyne	0.37	2.40	1.20	1.16	0.17	1.39	0.77	0.74
propyne	0.08	6.67	2.60	2.00	0.00	0.86	0.31	0.13
1-butyne	0.64	3.01	2.12	2.34	1.58	2.99	2.05	1.88
2-butyne	0.00	3.28	1.52	1.42	0.13	0.66	0.36	0.32
1-pentyne	6.07	8.19	7.57	7.93	0.19	2.45	1.25	1.14
2-pentyne	10.85	31.85	21.48	21.58	0.78	5.98	2.15	1.38
1-hexyne	1.35	4.22	3.39	3.86	2.54	3.81	3.01	2.91
2-hexyne	15.67	31.91	23.78	23.81	1.27	5.96	2.64	2.03
1-heptyne	0.11	4.34	3.08	3.83	1.47	2.62	1.75	1.69
1-octyne	0.03	3.00	2.04	2.44	1.36	2.61	1.59	1.46
1-nonyne	0.51	2.77	1.34	0.99	0.89	1.05	0.96	0.96
propadiene	2.29	10.50	6.17	5.94	0.12	1.03	0.40	0.22

Table A15: Absolute error in critical temperature (TC), pressure (PC), density (ρC), and normal boiling point (TNBP), predicted by the optimized Mie potentials, compared to experiment and correlations[154, 155, 158].

Compound	$\%_{ERR} T_C$	$\%_{ERR} P_C$	$\%_{ERR} \rho_C$	$\%_{ERR} T_{NBP}$
ethyne	1.22	7.55	0.82	0.65
propyne	0.08	1.74	1.67	0.49
1-butyne	0.77	0.75	5.37	0.91
2-butyne	0.37	2.99	3.38	0.09
1-pentyne	1.37	3.77	0.34	0.47
2-pentyne	4.41	1.90	1.64	0.56
1-hexyne	0.60	0.92	3.51	0.04
2-hexyne	4.29	2.97	0.04	1.13
1-heptyne	0.11	3.48	2.14	0.19
1-octyne	0.31	4.56	3.24	0.07
1-nonyne	0.50	5.45	4.19	0.31
propadiene	1.63	2.76	2.07	0.71

Table A16: Selected phase coexistence data for propane(1)+propyne(2) predicted by NVT Gibbs ensemble Monte Carlo simulations for the optimized Mie potentials. The maximum uncertainty in the mole fractions is 0.009.

278.15 K			303.15 K			328.15 K			353.15 K		
P (bar)	x_1	y_1	P (bar)	x_1	y_1	P (bar)	x_1	y_1	P (bar)	x_1	y_1
2.80(2)	0.000	0.000	6.2(1)	0.000	0.000	12.4(2)	0.000	0.000	22.3(1)	0.000	0.000
3.56(4)	0.199	0.332	6.9(1)	0.099	0.160	13.2(1)	0.099	0.141	23.7(3)	0.098	0.125
3.77(1)	0.299	0.455	7.31(2)	0.198	0.294	14.06(2)	0.198	0.266	24.4(3)	0.197	0.242
4.10(4)	0.399	0.563	8.01(7)	0.299	0.417	14.9(3)	0.297	0.378	25.3(1)	0.298	0.353
4.33(5)	0.499	0.654	8.43(1)	0.398	0.519	15.56(7)	0.396	0.482	26.2(2)	0.398	0.455
4.62(3)	0.599	0.733	8.84(1)	0.498	0.613	16.28(4)	0.497	0.580	27.6(1)	0.498	0.553
4.90(2)	0.699	0.807	9.43(9)	0.598	0.701	17.0(1)	0.597	0.672	28.3(2)	0.598	0.647
5.12(5)	0.799	0.874	9.90(6)	0.699	0.782	17.67(6)	0.698	0.758	29.2(1)	0.698	0.736
5.46(1)	0.900	0.940	10.27(3)	0.799	0.857	18.54(2)	0.798	0.842	30.3(1)	0.798	0.826
5.71(2)	1.000	1.000	10.68(2)	0.899	0.930	19.0(1)	0.899	0.921	31.2(1)	0.899	0.913
			11.08(1)	1.000	1.000	19.87(2)	1.000	1.000	32.6(2)	1.000	1.000

Table S17: Selected phase coexistence data for propene(1)+propyne(2) predicted by NVT Gibbs ensemble Monte Carlo simulations for the optimized Mie potentials. The maximum uncertainty in the mole fractions is 0.009.

278.15 K			303.15 K			328.15 K			353.15 K		
P (bar)	x_1	y_1	P (bar)	x_1	y_1	P (bar)	x_1	y_1	P (bar)	x_1	y_1
2.8(2)	0.000	0.000	6.48(9)	0.000	0.000	12.4(1)	0.000	0.000	22.15(6)	0.000	0.000
4.0(2)	0.198	0.373	7.10(6)	0.099	0.181	13.36(6)	0.098	0.159	23.9(2)	0.098	0.137
4.18(6)	0.299	0.509	7.9(1)	0.198	0.331	14.61(9)	0.197	0.292	25.30(9)	0.195	0.262
4.69(6)	0.399	0.604	8.56(9)	0.298	0.454	15.92(6)	0.297	0.412	27.1(2)	0.296	0.375
4.80(1)	0.499	0.691	9.34(4)	0.398	0.556	16.9(2)	0.397	0.515	28.7(3)	0.395	0.477
5.4(1)	0.599	0.766	9.98(5)	0.498	0.648	17.94(8)	0.497	0.608	30.3(3)	0.495	0.573
5.72(5)	0.699	0.832	10.69(4)	0.598	0.730	19.15(9)	0.597	0.696	31.79(1)	0.595	0.663
6.15(4)	0.799	0.892	11.33(3)	0.698	0.805	20.06(4)	0.697	0.777	33.8(1)	0.695	0.749
6.58(6)	0.900	0.948	12.00(5)	0.799	0.873	21.3(1)	0.798	0.854	34.9(3)	0.796	0.834
7.04(3)	1.000	1.000	12.74(4)	0.899	0.938	22.7(2)	0.899	0.928	36.4(2)	0.898	0.917
			13.4(1)	1.000	1.000	23.43(4)	1.000	1.000	38.06(3)	1.000	1.000

Table A18: Selected phase coexistence data for propadiene(1)+propyne(2) predicted by NVT Gibbs ensemble Monte Carlo simulations for the optimized Mie potentials. The maximum uncertainty in the mole fractions is 0.009.

278.15 K			328.15 K			353.15 K		
P (bar)	x_1	y_1	P (bar)	x_1	y_1	P (bar)	x_1	y_1
6.3(3)	0.000	0.000	12.6(2)	0.000	0.000	22.37(6)	0.000	0.000
6.7(3)	0.099	0.138	12.91(9)	0.099	0.128	23.22(4)	0.099	0.119
6.9(1)	0.199	0.266	13.47(8)	0.199	0.248	23.9(3)	0.198	0.232
7.22(8)	0.299	0.380	13.9(2)	0.299	0.359	24.5(2)	0.298	0.340
7.6(1)	0.399	0.480	14.80(6)	0.399	0.459	24.96(8)	0.398	0.443
7.9(2)	0.499	0.579	15.23(3)	0.499	0.558	25.88(8)	0.498	0.541
8.36(3)	0.599	0.668	15.69(9)	0.599	0.651	26.5(2)	0.598	0.636
8.4(1)	0.699	0.757	16.33(3)	0.699	0.741	27.03(9)	0.699	0.728
8.86(3)	0.800	0.842	16.3(1)	0.799	0.830	27.8(1)	0.799	0.820
8.94(9)	0.900	0.920	16.8(3)	0.900	0.916	28.4(1)	0.900	0.910
9.22(9)	1.000	1.000	16.93(3)	1.000	1.000	28.99(1)	1.000	1.000

APPENDIX B

In this section, the detailed computational procedures, mathematical methods of Molecular Exchange Monte Carlo (MEMC) move, and additional results are provided.

Defining the exchange sub-volume vectors and transformation matrix T_{VEX}:

An exchange sub-volume is a rectangular cuboid defined by three mutually orthogonal vectors a, b, and c. Vector c is either defined by the backbone orientation of the selected molecule or randomly defined according to a uniform distribution. For a given vector c, vectors a and b are generated based on the following *Gram-Schmidt* algorithm.

1- Set a and b to two independent vectors, such as i and j. (if c was in the same plane as a and b, set either of a or b to k).

2- $e_3 = \frac{c}{|c|}$

3- $b = b - (b.e_3)e_3$

4- $e_2 = \frac{b}{|b|}$

5- $a = a - (a.e_3)e_3 - (a.e_2)e_2$

6- $e_1 = \frac{a}{|a|}$

where $|a|$ is the norm of vector a, and $(a.b)$ represent scalar product of the two vectors.

To perform MEMC operations such as, counting the number of small molecules in sub-volume V_{EX}, inserting small molecules in V_{EX}, and aligning small and large molecules backbones with z-axis of the sub-volume, we need to define a new coordinate system based on the three unit vectors e_1, e_2, and e_3. To transform the coordinates from the simulation box reference frame to the one defined by e_1, e_2, and e_3, we apply the transformation matrix T_{VEX}^{-1} and for the inverse transformation we apply T_{VEX} as defined below:

$$T_{VEX} = \begin{bmatrix} e_{11} & e_{21} & e_{31} \\ e_{12} & e_{22} & e_{32} \\ e_{13} & e_{23} & e_{33} \end{bmatrix} \tag{B1}$$

$$T_{VEX}^{-1} = T_{VEX}^{T} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \tag{B2}$$

Defining a 2D random rotation matrix R_z about the z-axis of the sub-volume:

In an MEMC move, the backbone of the molecule is aligned with e_3 (z-axis of the sub-volume). To perform random rotation around the backbone, a rotation matrix R_z is defined according to the following procedure:

1- Set θ to a random number between 0 and 1.

2- $\theta = \theta \times 2 \times \pi$

3- $\theta = \theta - \pi$

$$R_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{B3}$$

Defining a 3D random rotation matrix R_s:

In the MEMC move, to perform rotation on a sphere uniformly, the fast random rotation matrices algorithm by Arvo is used. To construct the rotation matrix, perform the following steps.

1- Set θ to a random number between 0 and 1.

2- $\theta = \theta \times 2 \times \pi$

3- $\theta = \theta - \pi$

4- Set φ to a random number between 0 and 1.

5- $\varphi = \varphi \times 2 \times \pi$

6- Set r to a random number between 0 and 1.

7- Construct the 2D rotation R_z, using θ.

8- Define v as

$$v = \begin{bmatrix} \sqrt{r} \sin \varphi \\ \sqrt{r} \cos \varphi \\ \sqrt{1-r} \end{bmatrix} \tag{B4}$$

9- Defining the *Householder matrix* $H = I - 2\,v\,v^T$

10- The final rotation matrix can be expressed as

$$R_S = -H\,R_z = 2\,v\,v^T R_z - R_z \tag{B5}$$

Defining the random orientation vector c for the exchange sub-volume V_{EX}:

To generate a random orientation for the exchange sub-volume V_{EX}, we generate the vector c according to the following algorithm:

1- $c = k$

2- Construct the 3D rotation matrix R_s

3- $c = R_s\,c$

Finding the number of small molecules within the sub-volume V_{EX}:

To count the number of small molecules inside the V_{EX}, with the geometric center defined as vector r_c and dimensions of $w \times w \times l$, the following steps are performed. Repeat steps 1-3 for all the small molecules within the simulation box.

1- Calculate the minimum image distance between the geometric center of the sub-volume and centroid of the molecule: $\Delta r = r_c - r_{centroid}$.

2- Transform the vector to the sub-volume coordinate system: $\Delta r' = T_{VEX}^{-1}\,\Delta r$

3- If $\Delta r_1' < 0.5w$ and $\Delta r_2' < 0.5w$ and $\Delta r_3' < 0.5l$, the molecule is located within the sub-volume.

Finding a random location for centroid of small molecule, within the sub-volume V_{EX}:

1- Set u_1, u_2, and u_3 to a random number between 0 and 1, independently.

2- $x_{centroid} = u_1 \times w - 0.5\,w$, $y_{centroid} = u_2 \times w - 0.5\,w$, $z_{centroid} = u_3 \times l - 0.5\,l$

3- Transform the centroid coordinate vector $r_{centroid}$, to the sub-volume coordinate system:

$$r'_{centroid} = T_{VEX}\, r_{centroid}$$

4- Shift the $r'_{centroid}$ to the geometric center of the sub-volume r_c: $r''_{centroid} = r'_{centroid} + r_c$

Generate Rotational trial around centroid:

1- Construct the 3D rotation matrix R_s

2- Repeat the following steps, for all atoms in the molecule ($i = 0, 1, \dots, n$)

    a. Shift the atom i to the origin with respect of its centroid: $r'_i = r_i - r_{centroid}$

    b. Rotate the atom i around origin: $r''_i = R_s\, r'_i$

    c. Shift the atom i back to its location: $r'''_i = r''_i + r_{centroid}$

Generate Rotational trial around the molecule's backbone (aligned with z-axis of the sub-volume):

       To generate the rotational trial around backbone of the molecule, the molecule's backbone must be aligned with predefined sub-volume V_{EX} system coordinate, T_{VEX}. To align the molecule with the V_{EX}, the transformation matrix of molecule system coordinate T_M is defined as follow:

1- Shift the molecule coordinates to the origin with respect to its centroid.

2- Calculate the minimum image vector of two specific atoms of the molecule Δr that represent the orientation of the molecule's backbone.

3- Set c to this vector: $c = \Delta r$

4- Construct transformation matrix T_M of the molecule using the *Gram-Schmidt* algorithm.

5- Transform the molecule coordinates to the simulation box coordinate system, where c is aligned with the z-axis. Repeat the following step for all atoms in the molecule ($i = 0, 1, \dots, n$)

    a. $r'_i = T_M^{-1}\, r_i$

Once the molecule coordinates are transformed, rotational trials around the z-axis are generated, molecule coordinates are transformed to V_{EX} system coordinate, and shifted to the geometric center of the sub-volume r_c, as follows:

6- Construct the 2D rotational matrix R_Z.

7- Repeat the following step for all atoms in the molecule ($i = 0, 1, ..., n$)

    a. $r_i'' = R_Z r_i'$

    b. $r_i''' = T_{VEX} r_i''$

    c. $r_i''' = r_i''' + r_c$

Forcefield:

Mie potential has been optimized for noble gases[126, 127], linear and branched alkane[24, 124], n-alkyne[64]. All non-bonded parameters used in this work are listed in Table B1.

Table B1: Non-bonded parameters for n-alkanes, perfluoro-alkanes[24], branched alkanes[124], and SPC/E water[188].

Pseudo-atom	$\varepsilon_i/k_b(K)$	σ_i (Å)	n_i	q_i
CH$_4$	161.00	3.740	14	0.00
CH$_3$	121.25	3.783	16	0.00
CH$_2$	61.00	3.990	16	0.00
CH ($C_N > 4$, S/L)	14.00	4.700	16	0.00
C ($C_N \leq 4$, S/L)	1.45	6.100	16	0.00
C ($C_N > 4$, S/L)	1.20	6.200	16	0.00
CF$_3$	155.75	4.475	36	0.00
CF$_2$	72.20	4.750	44	0.00
O	78.21	3.167	12	-0.8476
H	0.00	0.00	0.00	0.4238

Fixed bond lengths for n-alkanes, perfluoro-alkane[24], branched alkanes[124], and SPC/E water[188] were used to connect pseudo-atoms and are listed in Table B2. Equilibrium bond angles and force constants are listed in Table B2.

Table B2: Bonded parameters for n-alkanes, perfluoro-alkane[24], branched alkanes[124], and SPC/E water[188].

Bond type	Bond length (Å)	Angle type	θ_0(degree)	k_θ/k_b (K.rad^{-2})
CH_2-CH_3	1.54	$CH_3-CH_2-CH_2$	114	31250
CH_2-CH_2	1.54	$CH_2-CH_2-CH_2$	114	31250
CH_2-CH_2	1.54	$C-CH_2-CH$	114	31250
$CH-CH_3$	1.54	$CH_3-CH-CH_3$	112	31250
$CH-CH_2$	1.54	$CH_3-CH-CH_2$	112	31250
$C-CH_3$	1.54	CH_3-C-CH_3	109.47	31250
$C-CH_2$	1.54	CH_3-C-CH_2	109.47	31250
CF_2-CF_3	1.54	$CF_3-CF_2-CF_2$	114	31250
CF_2-CF_2	1.54	$CF_2-CF_2-CF_2$	114	31250
$O-H$	1.00	$H-O-H$	109.47	Fixed

Dihedral parameters are listed in Table B3. Fourier constants for alkanes were taken from OPLS-UA[139, 140] and for perfluoroalkanes, more accurate seven term cosine series were used.

Table B3: Torsional parameters for n-alkanes, perfluoro-alkane[24], branched alkanes[124], and SPC/E water[188].

torsion	n	c_n/k_b (K)	δ_n
CH_x—(CH_2)—(CH_2)—CH_2	1	335.03	0
	2	-68.19	π
	3	791.32	0
CH_x—(CH_2)—(CH)—CH_y	0	-251.06	0
	1	428.73	0
	2	-111.85	π
	3	441.27	0
CH_x—(CH_2)—(C)—CH_y	3	461.29	0
CF_x—(CF_2)—(CF_2)—CF_y	0	-1577.68	0
	1	791.61	0
	2	333.65	0
	3	854.01	0
	4	349.25	0
	5	211.51	0
	6	117.66	0
	7	-83.44	0

Additional Results:

In this section, the numerical results and additional data are provided.

Table B4: Selected phase coexistence data for perfluorobutane(1)+n-butane(2) predicted by grand canonical Monte Carlo simulations using Mie potentials[24]. Uncertainty in data are presented by the numbers in parenthesis.

P (bar)	x_1	y_1
0.68(3)	0.02(1)	0.10(5)
0.87(2)	0.12(1)	0.34(2)
0.92(1)	0.20(2)	0.40(1)
0.97(1)	0.36(2)	0.472(8)
0.981(5)	0.50(3)	0.517(5)
0.980(3)	0.59(3)	0.553(3)
0.963(5)	0.709(7)	0.610(1)
0.928(6)	0.802(2)	0.680(1)
0.855(4)	0.908(3)	0.806(1)
0.747(3)	1.000	1.000

Table B5: Vapor-liquid coexistence data predicted from GCMC+histogram reweighting simulations using ME-2 method for SPC/E water.

T (K)	ρ_l (kg/m^3)	ρ_v (kg/m^3)	P (bar)	ΔH_v (kJ/mol)
600	535(3)	90(2)	102.7(7)	15.6(2)
580	615(2)	52.1(8)	74.7(2)	21.4(2)
560	675(2)	33.0(1)	53.54(9)	25.76(3)
540	720(1)	21.51(6)	37.54(6)	29.09(3)
520	756(1)	14.11(5)	25.71(5)	31.77(4)
500	787(1)	9.17(4)	17.12(4)	34.01(3)
480	818(1)	5.86(2)	11.04(2)	36.01(6)
460	846(2)	3.64(1)	6.84(2)	37.97(8)
440	871(1)	2.173(5)	4.05(2)	39.69(4)
420	896(2)	1.241(3)	2.26(1)	41.31(6)
400	917(2)	0.669(4)	1.19(1)	42.92(4)
380	935(1)	0.336(3)	0.576(5)	44.29(5)
360	953(2)	0.155(1)	0.255(3)	45.59(3)
340	969(1)	0.065(1)	0.101(1)	46.74(4)
320	984(1)	0.024	0.035	47.94(4)
300	996(2)	0.007	0.010	49.19(5)
280	1007(5)	0.002	0.003	50.4(1)

The acceptance rate of inserting or removing neopentane was 68 times lower than the acceptance rate for exchanging neopentane with 2,2,4-trimethylpentane and vice versa via the ME-2 algorithm. This shows that insertion of neopentane is the rate limiting step in the process. In order to improve the acceptance rate for insertions of neopentane, CBMC angle and dihedral trials were increased to 500, and the number of CBMC trials for the first atom and remaining atoms were increased to 16 and 10, respectively. In Table B6, a detailed comparison is presented for the acceptance rates for direct swaps of neopentane and 2,2,4-trimethylpentane, MEMC moves, effective acceptance rates and effective acceptance rates per CPU time.

Table B6: Comparison of acceptance rates for swaps of the impurity molecule (neopentane), identity exchange via the MEMC algorithm, and swaps performed with standard configurational-bias Monte Carlo for 2,2,4-trimethylpentane.

T (K)	$\%P_{Imp-acc}$	$\%P_{Switch-acc}$		$\%P_{Effective-acc}$		$\%P_{acc}$	Effective acceptance per CPU time (1/s)			Relative acceptance efficiency	
	swap	ME-2	ME-3	ME-2	ME-3	CBMC	CBMC	ME-2	ME-3	ME-2	ME-3
280	0.017	1.19	0.05	0.039	0.099	0.0002	0.0005	0.244	0.042	520.2	90.0
330	0.168	2.60	0.18	0.183	0.077	0.0023	0.0047	1.312	0.250	278.7	53.3
390	1.30	5.89	0.71	0.878	0.393	0.036	0.0745	5.504	1.023	73.9	13.7
450	6.05	10.26	1.56	3.466	1.067	0.338	0.745	22.40	2.779	30.1	3.73
510	18.0	21.76	3.43	7.887	2.337	1.510	3.520	49.02	5.576	13.9	1.58
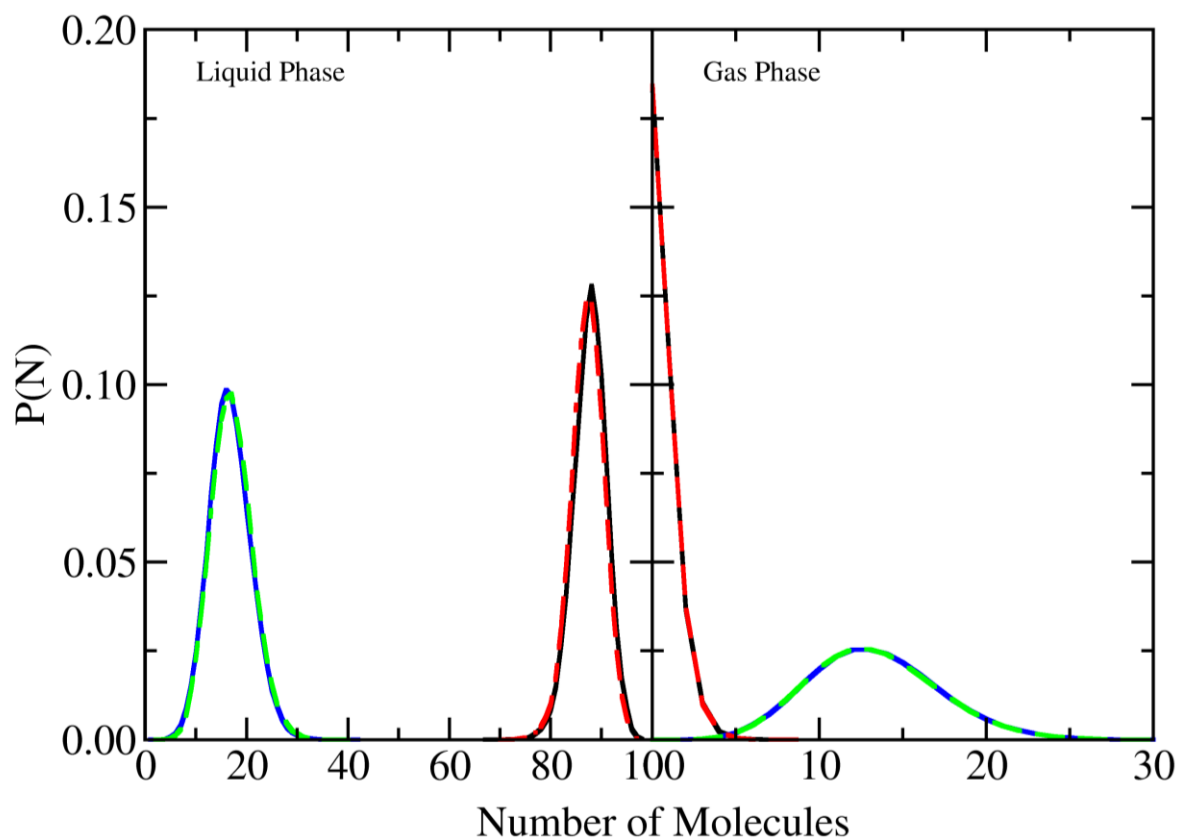
Figure B1: Probability distributions predicted from gas $(\mu_{butane} = -2960, \mu_{methane} = -2000)$ and liquid $(\mu_{butane} = -2840, \mu_{methane} = -2000)$ phase GCMC simulations of methane+n-butane at 277 K. Solid lines denote the probability distributions for n-butane (black) and methane (blue) using standard configurational-bias insertions and deletions. Dashed lines denote the probability distributions for n-butane (red) and methane (green) using the ME-1 algorithm.
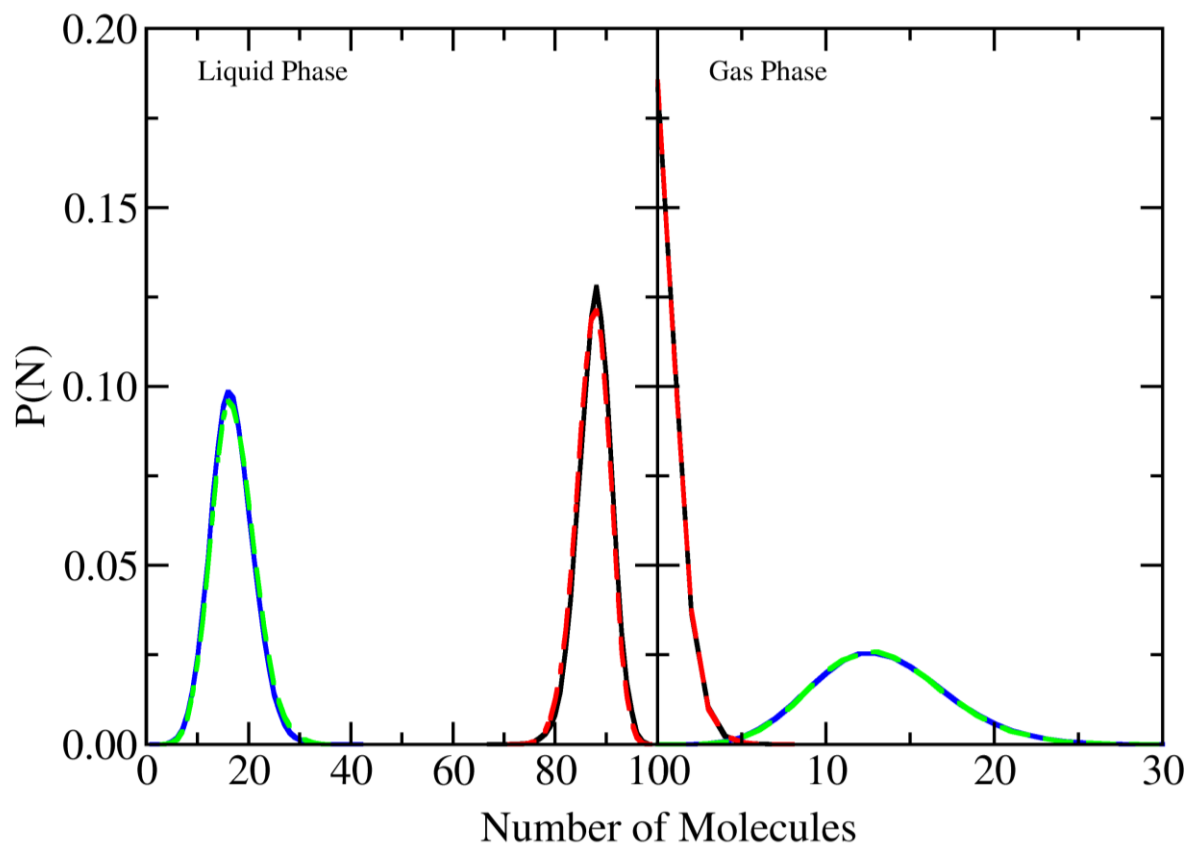
Figure B2: Probability distributions predicted from gas $(\mu_{butane} = -2960, \mu_{methane} = -2000)$ and liquid $(\mu_{butane} = -2840, \mu_{methane} = -2000)$ phase GCMC simulations of methane+n-butane at 277 K. Solid lines denote the probability distributions for n-butane (black) and methane (blue) using standard configurational-bias insertions and deletions. Dashed lines denote the probability distributions for n-butane (red) and methane (green) using the ME-2 algorithm.
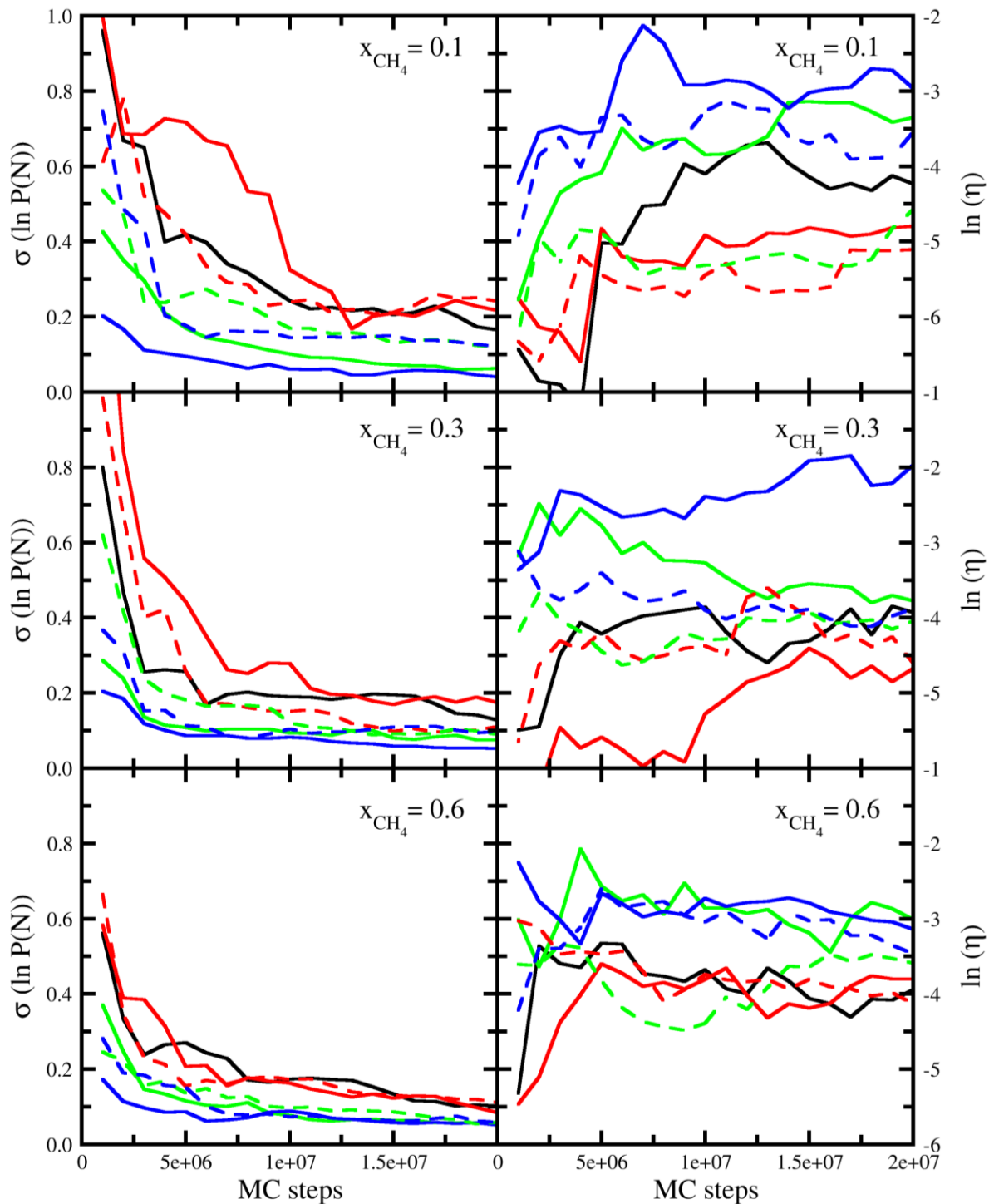
Figure B3: Efficiency and standard deviation in methane+n-butane binary mixture at 255 K. Lines represent the efficiency and uncertainty in n-butane distribution probability; standard CBMC method (black), ME-1 (red), ME-2 (green), and ME-3 (blue). MEMC move with exchanging one n-butane with one methane represented by solid lines, exchanging one n-butane with two methane molecules represented by dashed lines.

Figure B4: Molecule probability distribution in methane+n-butane binary mixture system at $x_{methane} = 0.1$ and 255 K. Lines in magenta, green, blue, red, and black represent the probability distribution of n-butane after 1, 5, 10, 15, and 20 million MC steps, respectively. (A) represent probability distribution using standard insertion and deletion with coupled-decoupled CBMC technique, (B), (C), and (D) represent probability distribution using ME-1, ME-2, and ME-3 method with exchanging ratio of one methane with one n-butane, respectively.
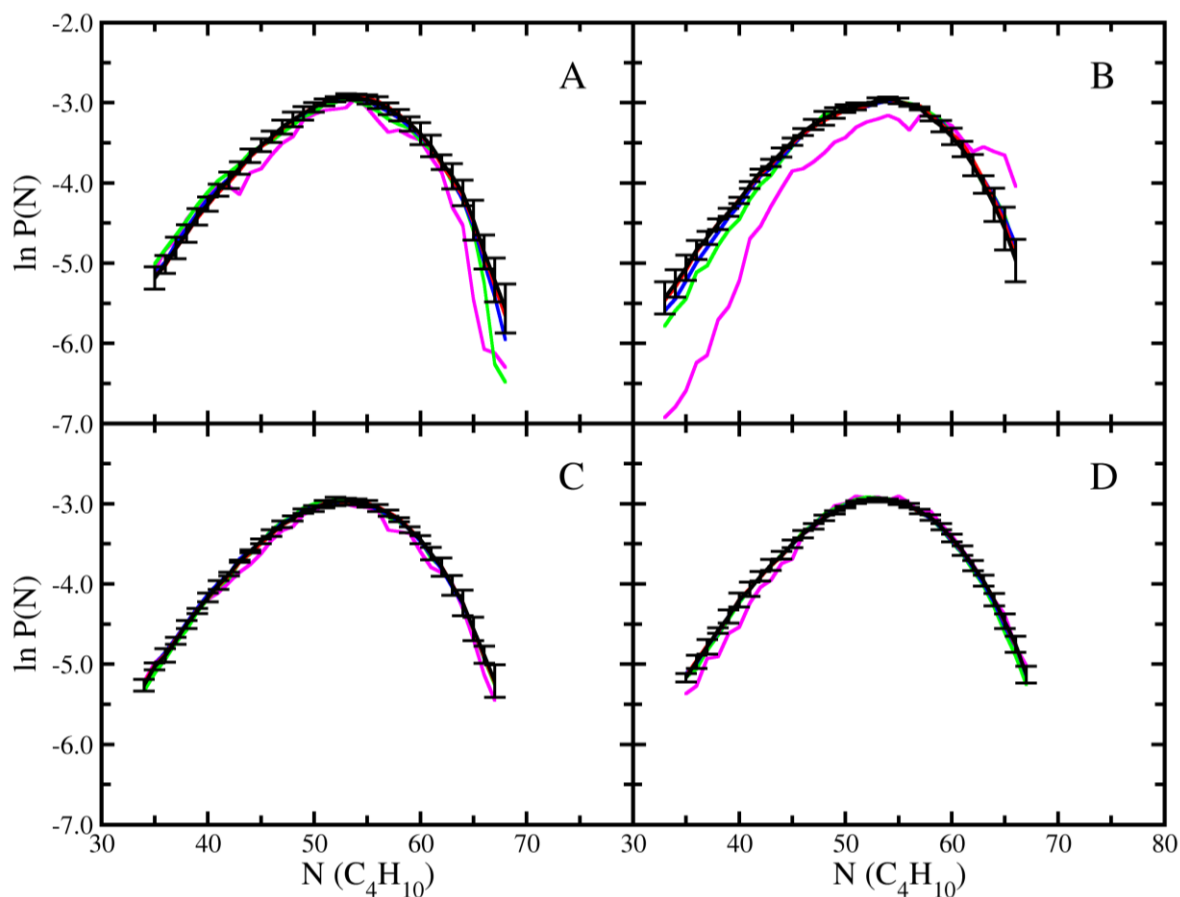
Figure B5: Molecule probability distribution in methane+n-butane binary mixture system at $x_{methane} = 0.6$ and 255 K. Lines in magenta, green, blue, red, and black represent the probability distribution of n-butane after 1, 5, 10, 15, and 20 million MC steps, respectively. (A) represent probability distribution using standard insertion and deletion with coupled-decoupled CBMC technique, (B), (C), and (D) represent probability distribution using ME-1, ME-2, and ME-3 method with exchanging ratio of one methane with one n-butane, respectively.
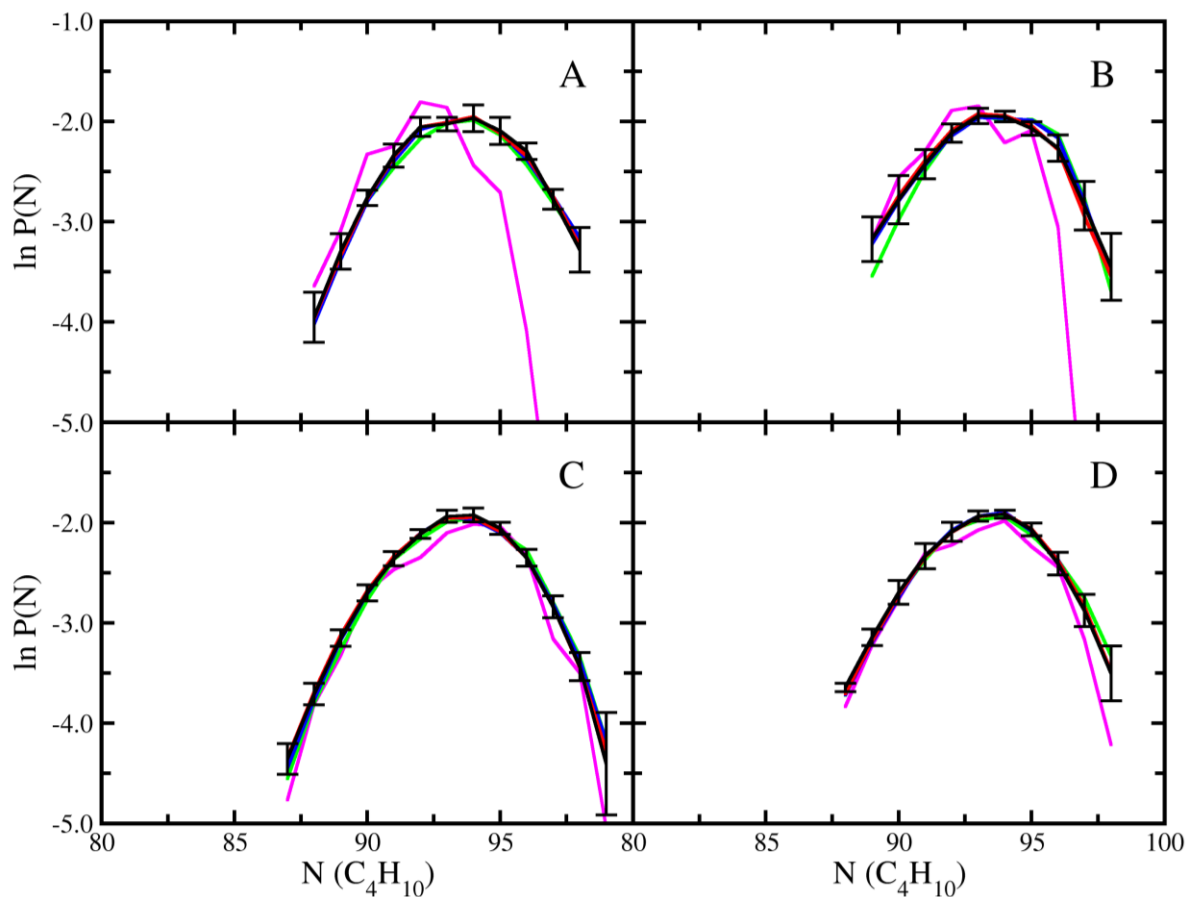
Figure B6: Molecule probability distribution in methane+n-butane binary mixture system at $x_{methane} = 0.1$ and 255 K. Lines in magenta, green, blue, red, and black represent the probability distribution of n-butane after 1, 5, 10, 15, and 20 million MC steps, respectively. (A) represent probability distribution using standard insertion and deletion with coupled-decoupled CBMC technique, (B), (C), and (D) represent probability distribution using ME-1, ME-2, and ME-3 method with exchanging ratio of two methane molecules with one n-butane, respectively.
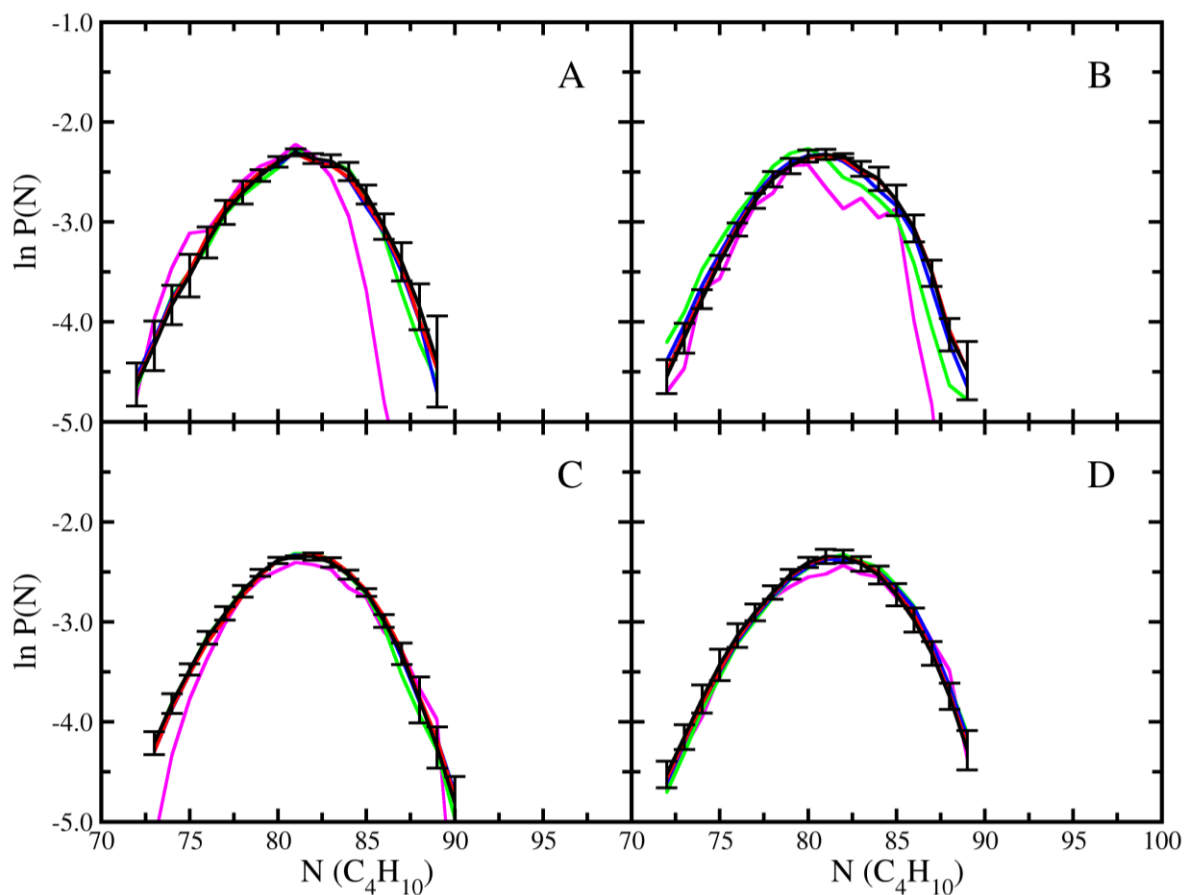
Figure B7: Molecule probability distribution in methane+n-butane binary mixture system at $x_{methane} = 0.3$ and 255 K. Lines in magenta, green, blue, red, and black represent the probability distribution of n-butane after 1, 5, 10, 15, and 20 million MC steps, respectively. (A) represent probability distribution using standard insertion and deletion with coupled-decoupled CBMC technique, (B), (C), and (D) represent probability distribution using ME-1, ME-2, and ME-3 method with exchanging ratio of two methane molecules with one n-butane, respectively.
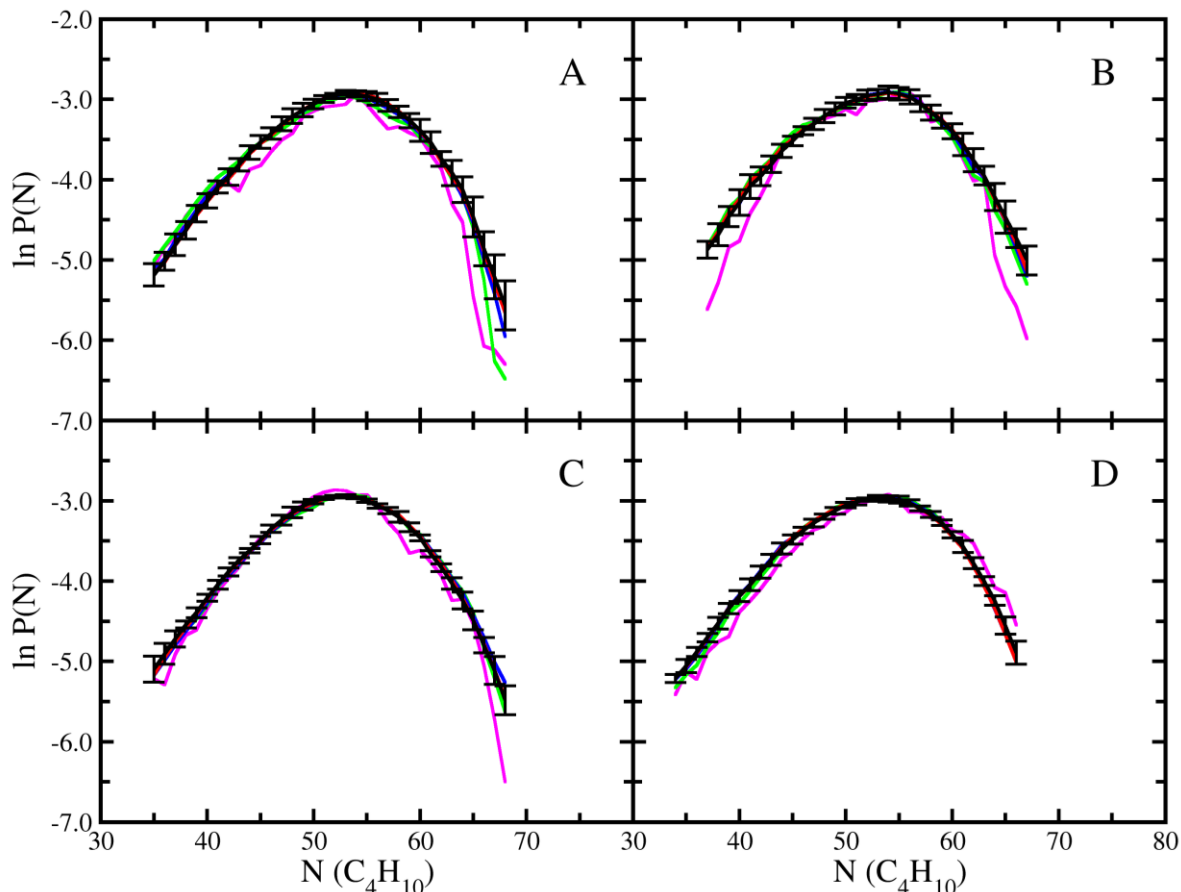
Figure B8: Molecule probability distribution in methane+n-butane binary mixture system at $x_{methane} = 0.6$ and 255 K. Lines in magenta, green, blue, red, and black represent the probability distribution of n-butane after 1, 5, 10, 15, and 20 million MC steps, respectively. (A) represent probability distribution using standard insertion and deletion with coupled-decoupled CBMC technique, (B), (C), and (D) represent probability distribution using ME-1, ME-2, and ME-3 method with exchanging ratio of two methane molecules with one n-butane, respectively.

In Figure B9, the effect of CBMC parameters on perfluorobutane insertion/deletion acceptance and acceptance efficiency of standard CBMC and ME methods are provided. For perfluorobutane+butane with an exchange ratio of 1, both ME-2 and ME-3 are independent from first site atom trials, while ME-1 and standard CBMC are dependent to this variable. The maximum acceptance 0.09% and acceptance efficiency 0.91 (1/sec) for standard CBMC is achieved at 18 trials for the first atom site and 12 trials for remaining atoms. In the ME-1 method, increasing both variables would lead to increases in acceptance but decreases in the acceptance efficiency. Using 2 trials for the centroid position and 1 trial for molecular rotation

results in an acceptance rate of 0.008% and acceptance efficiency of 0.45 (1/sec). In the case of ME-2, by increasing the number of secondary site trials, the acceptance increases while acceptance efficiency decreases. The maximum acceptance efficiency of 114 (1/sec) is achieved by using 1 trial for molecular rotation, which leads to 1.65% acceptance. The behavior of the ME-3 method is similar to the standard CBMC method, where the maximum acceptance of 3.85% and acceptance efficiency of 26.3 (1/sec) was achieved by using 18 trials for the first atom and 12 trials for the remaining atoms. Comparing acceptance efficiency of ME methods with standard CBMC using the optimum CBMC parameters, ME-2 and ME-3 are 120 and 28 more efficient, respectively. For ME-1, acceptance efficiency decreases by a factor of 2.
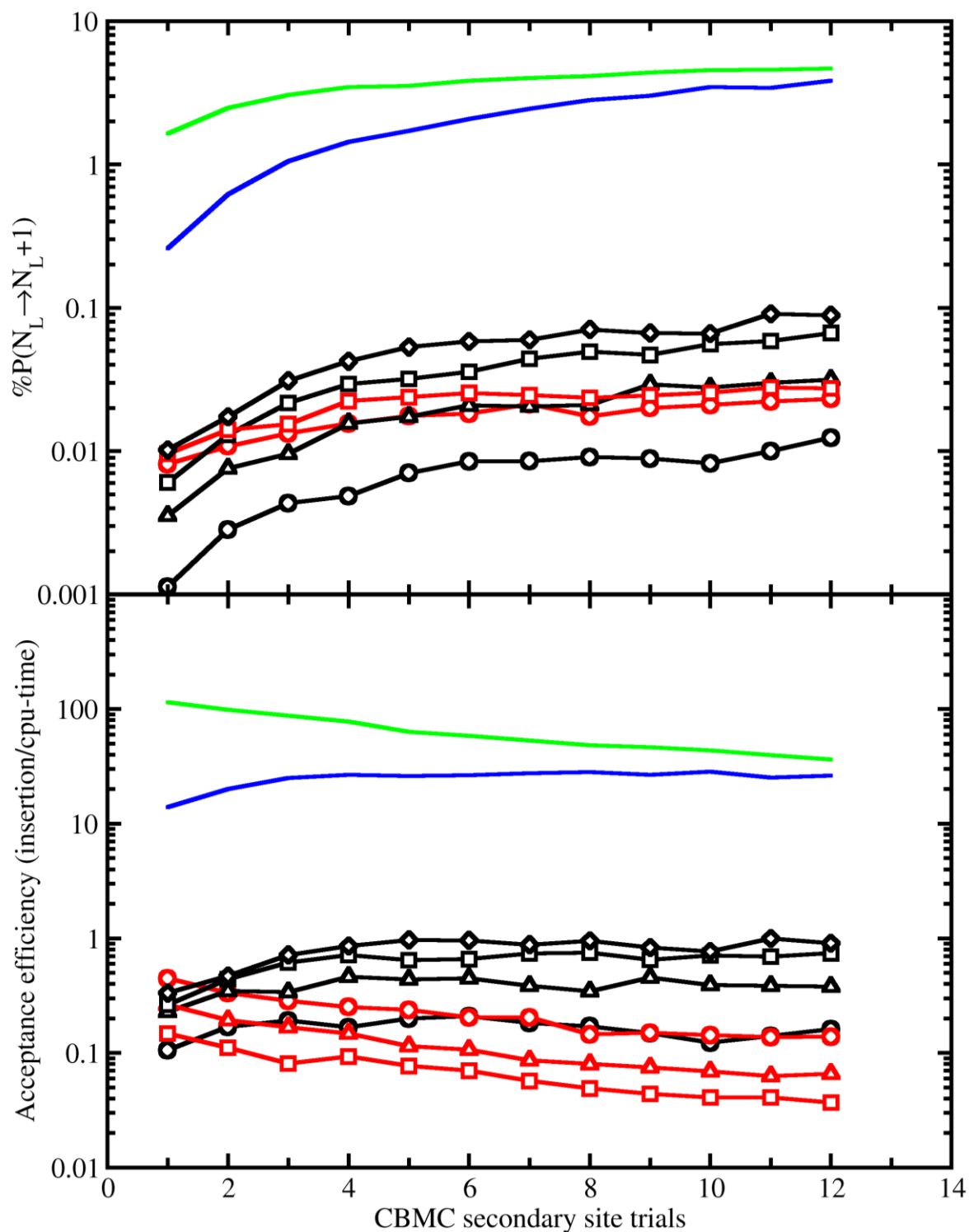
Figure B9: Acceptance and acceptance efficiency in perfluorobutane+n-butane binary mixture at 259.95 K and composition of 0.5. Lines represents acceptance and acceptance efficiency of perfluorobutane insertion in various CBMC trials for the site. Standard CBMC (black), ME-1 (red), ME-2 (green), and ME-3 (blue), 2 trials(circle), 6 trials (squares), 12 trials (triangles), 18 trials (diamonds). MEMC moves were performed with an exchange ratio of one to one.
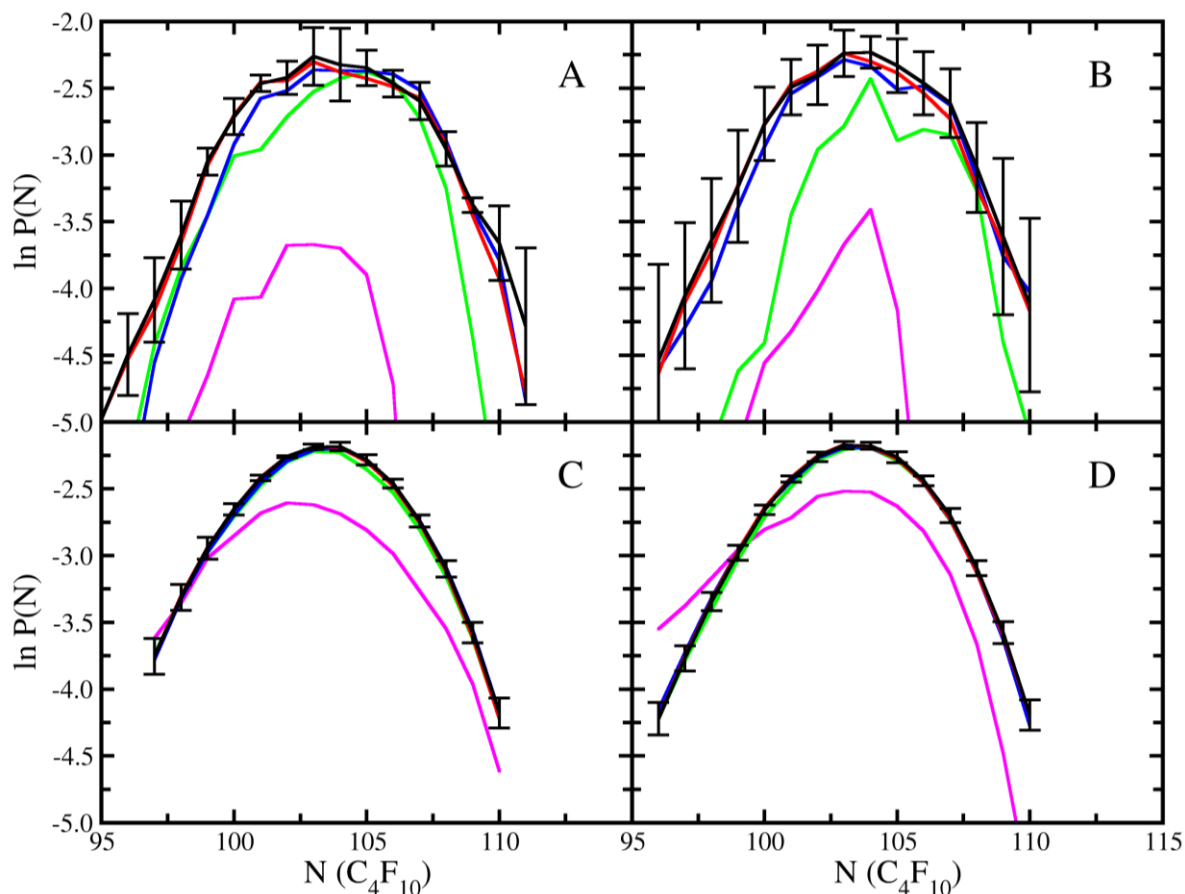
Figure B10: Molecule probability distribution in perfluorobutane+n-butane binary mixture system at $x_{butane} = 0.1$ and 259.95 K. Lines in magenta, green, blue, red, and black represent the probability distribution of perfluorobutane after 1, 5, 10, 15, and 20 million MC steps, respectively. (A) represent probability distribution using standard insertion and deletion with coupled-decoupled CBMC technique, (B), (C), and (D) represent probability distribution using the ME-1, ME-2, and ME-3 method with exchanging ratio of one n-butane with one perfluorobutane, respectively.
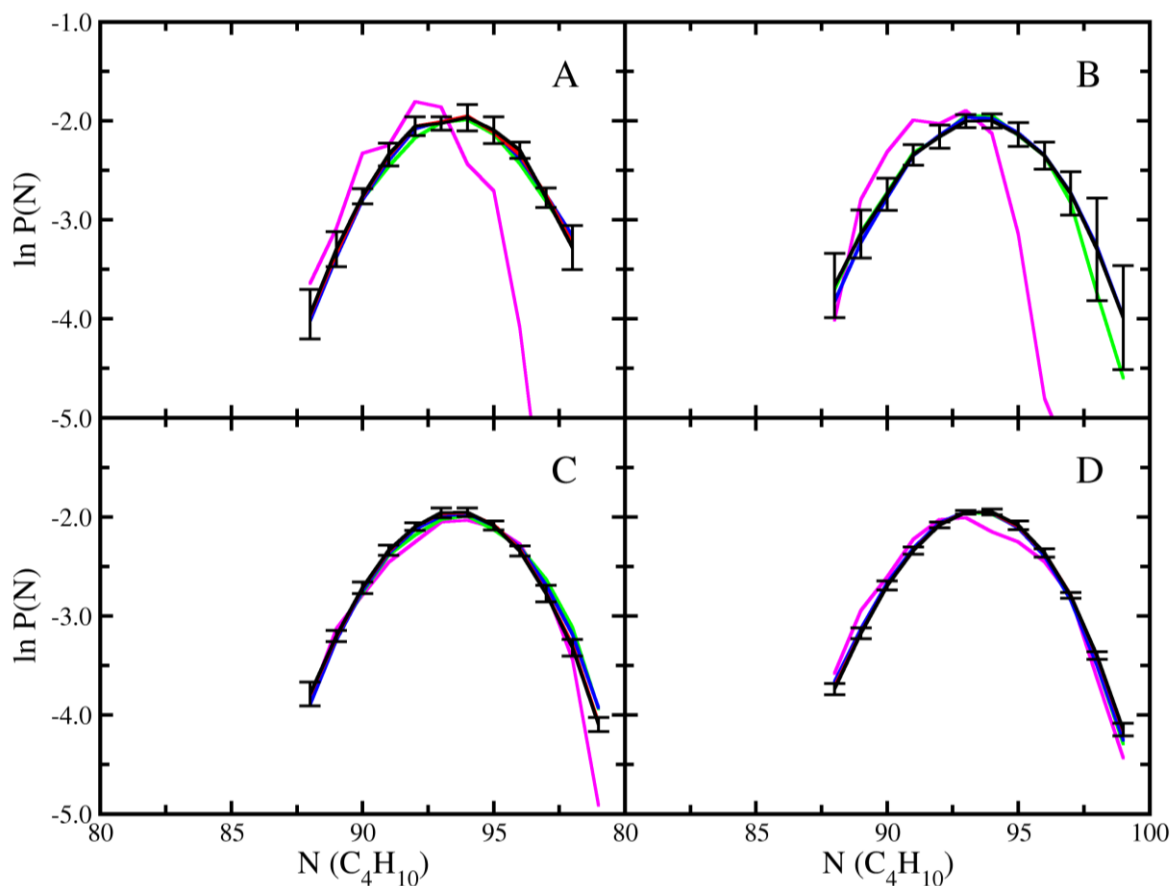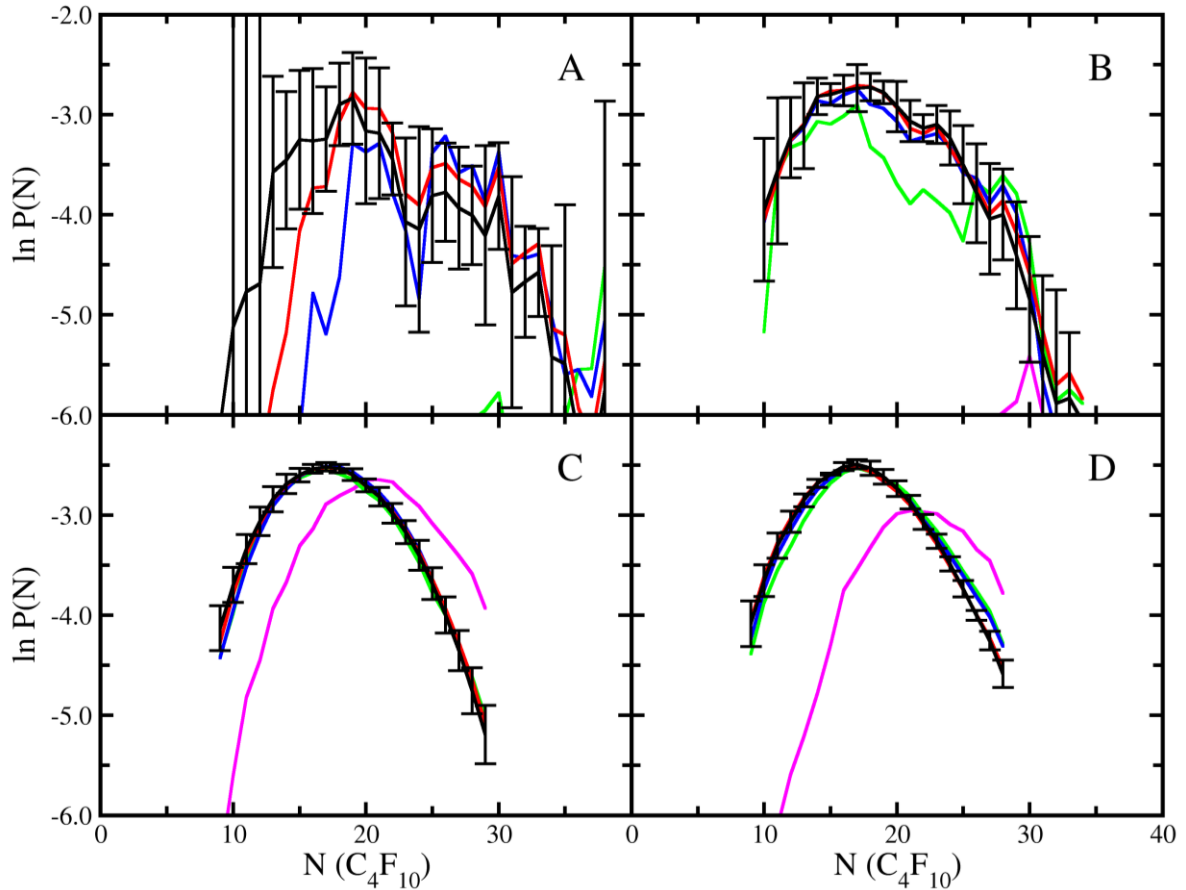
Figure B11: Molecule probability distribution in perfluorobutane+n-butane binary mixture system at $x_{butane} = 0.9$ and 259.95 K. Lines in magenta, green, blue, red, and black represent the probability distribution of perfluorobutane after 1, 5, 10, 15, and 20 million MC steps, respectively. (A) represent probability distribution using standard insertion and deletion with coupled-decoupled CBMC technique, (B), (C), and (D) represent probability distribution using ME-1, ME-2, and ME-3 method with exchanging ratio of one n-butane with one perfluorobutane, respectively.
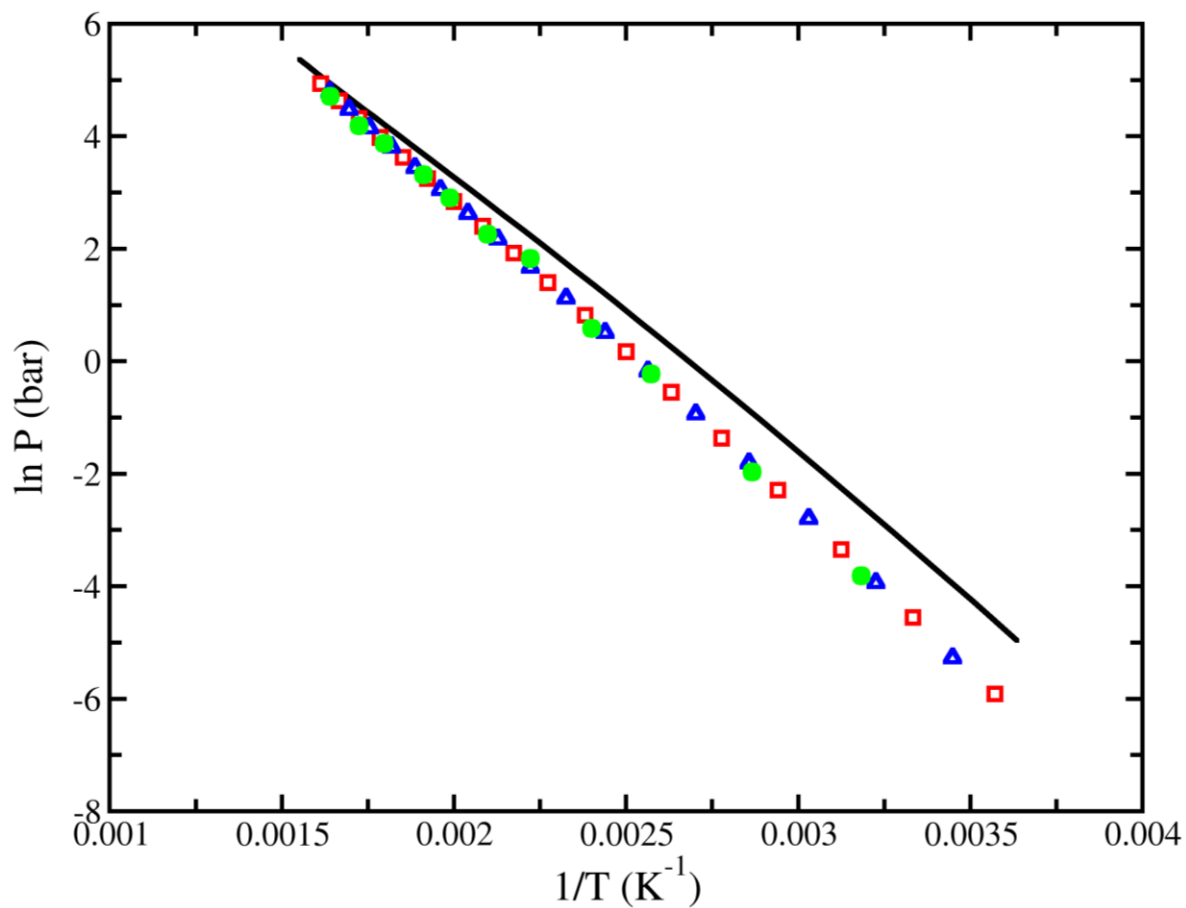
Figure B12: Clausius-Clapeyron plot for SPC/E water predicted from GCMC+histogram reweighting simulations. NIST Chemistry WebBook[158] (solid lines), values obtained by Boulougouris et. al.[189], (green circles) ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles).
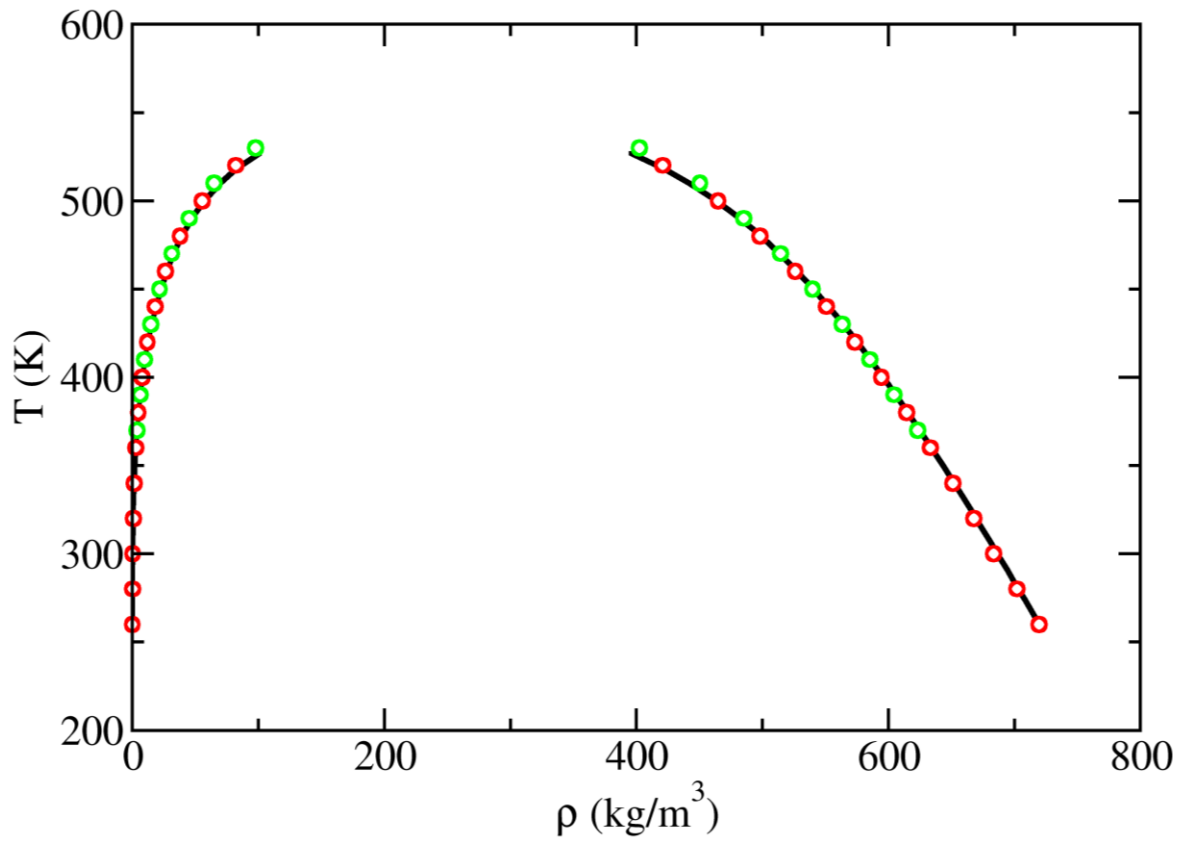
Figure B13: Vapor-liquid coexistence curve for 2,2,4-trimethylpentane predicted from GCMC+histogram reweighting simulations using Mie potentials[124]. Experimental data (solid lines)[151], ME-3 algorithm (red circles), and prior calculations using only configurational-bias Monte Carlo (green circles)[124].
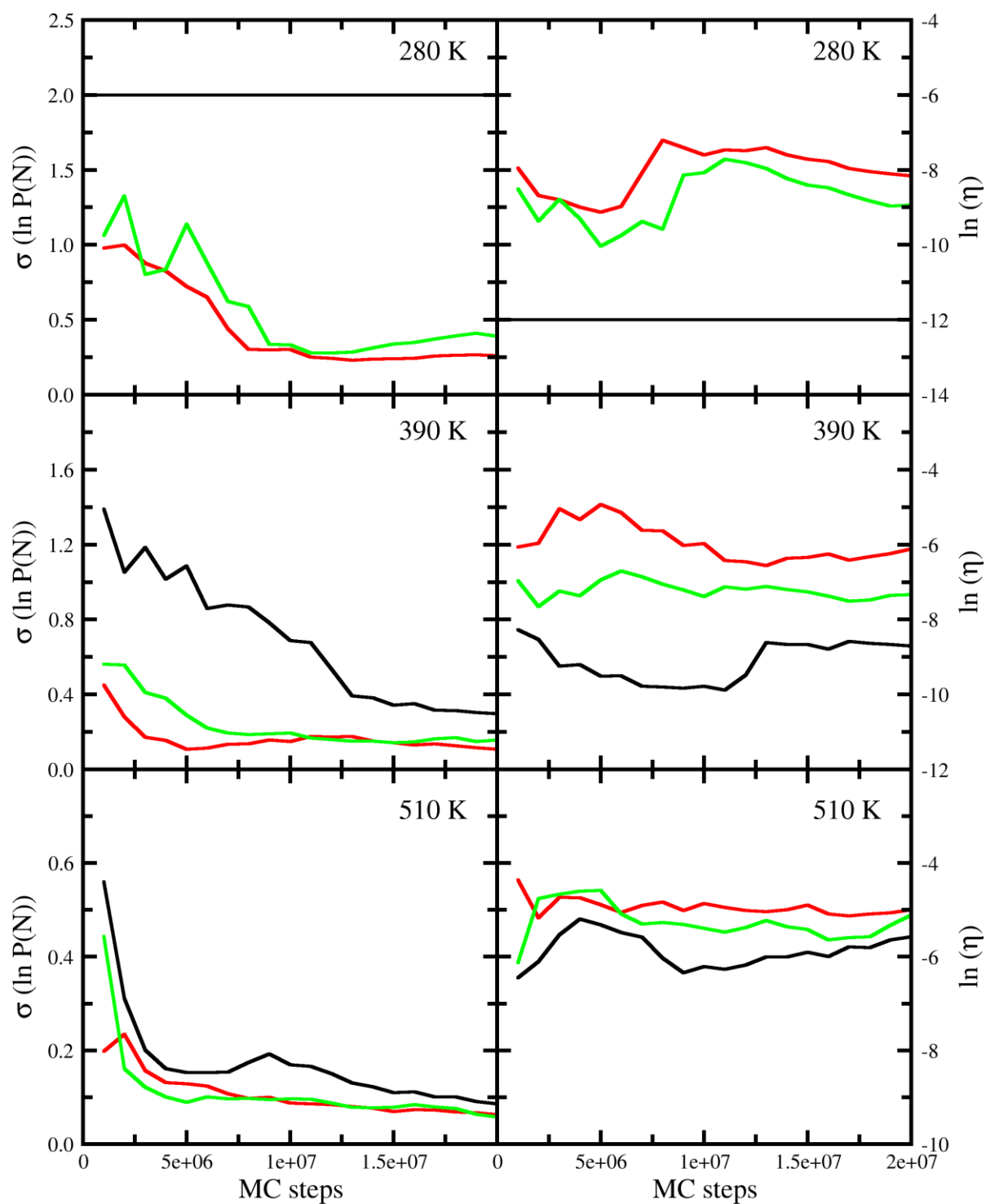
Figure B14: Standard deviation (left panel) and efficiency (right panel) for GCMC simulations for 2,2,4-trimethylpentane in the liquid phase. Configurational-bias insertions (black), ME-2 (red) and ME-3 (green).

APPENDIX C

In this section, additional results with their numerical value for Molecular Exchange Monte Carlo move in GEMC simulation, are provided.
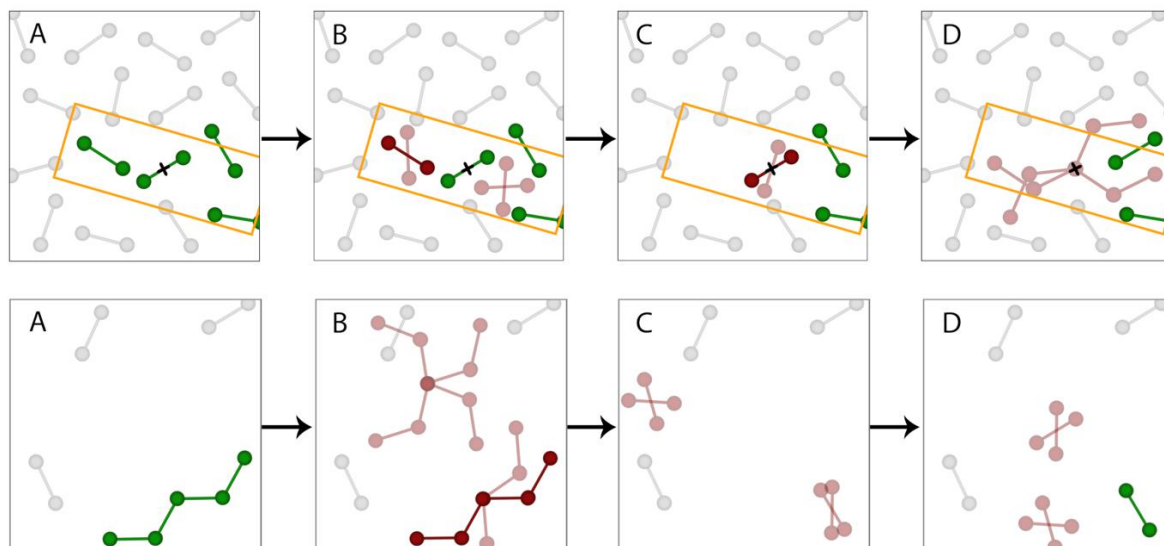
Figure C1: Schematic of the ME-3 algorithm for large molecule transfer from box 2 (gas phase) into box 1 (liquid phase) and transfer of two small molecules from box 1 into box 2. Selected or inserted molecule (green), trial position (light red), and actual position of the molecule (solid red). Top row, represents the exchange of two small molecules with one large molecule in box 1. The sub-volume is defined as the orange box. (A) Defining the sub-volume with a random orientation, where its geometric center is placed at a randomly selected small molecule's centroid, identifying the small molecules within the sub-volume, and randomly pick one small molecule. (B) Generating CBMC trials (3D rotation and centroid location) for the second small molecules and then removing it. (C) Generating CBMC 3D rotational trials for the first small molecule and then removing it. (D) Placing the predefined atom of the large molecule at the geometric center of the sub-volume and growing the large molecule using coupled-decoupled CBMC technique, and inserting it. Bottom row, represents the exchange of one large molecule with two small molecules in box 2. (A) Selecting a random large molecule. (B) Generating coupled-decoupled CBMC trials and then removing it. (C) Generating CBMC trials (3D rotation and centroid location) for the first small molecules and then inserting it. (D) Generating CBMC trials (3D rotation and centroid location) for the second small molecule and then inserting it.
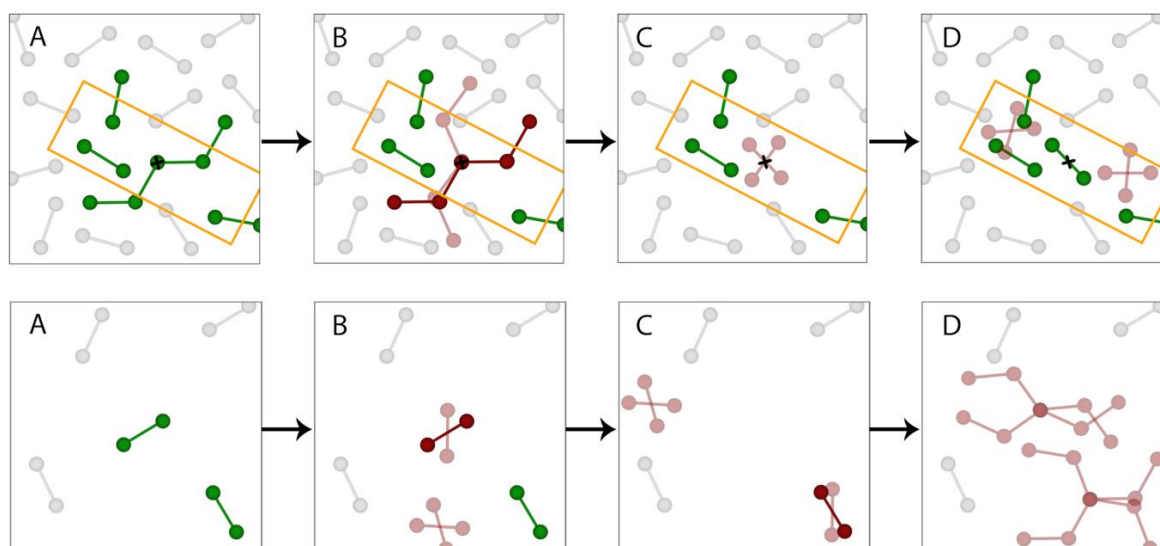
Figure C2: Schematic of the ME-3 algorithm for large molecule transfer from box 1 (liquid phase) into box 2 (gas phase) and transfer of two small molecules from box 2 into box 1. Selected or inserted molecule (green), trial position (light red), and actual position of the molecule (solid red). Top row, represents the exchange of one large molecules with two small molecules in box 1. The sub-volume is defined as the orange box. (A) Defining the sub-volume with a random orientation with geometric center placed at the predefined atom of the large molecule and identifying the small molecules within the sub-volume. (B) Generating coupled-decoupled CBMC trials for the large molecule and then removing it. (C) Placing the centroid of the first small molecule at the geometric center of the sub-volume, generating the CBMC 3D rotational trials, and then inserting it into the sub-volume. (D) Generating CBMC trials (3D rotation and centroid location) for the second small molecule and then inserting it into the sub-volume. Bottom row, represents the exchange of two small molecules with one large molecule in box 2. (A) Selecting two random small molecules. (B) Generating CBMC trials (3D rotation and centroid location) for the first small molecule and then removing it. (C) Generating CBMC trials (3D rotation and centroid location) for the second small molecule and then removing it. (D) Generating coupled-decoupled CBMC trials for the large molecule and then inserting it.
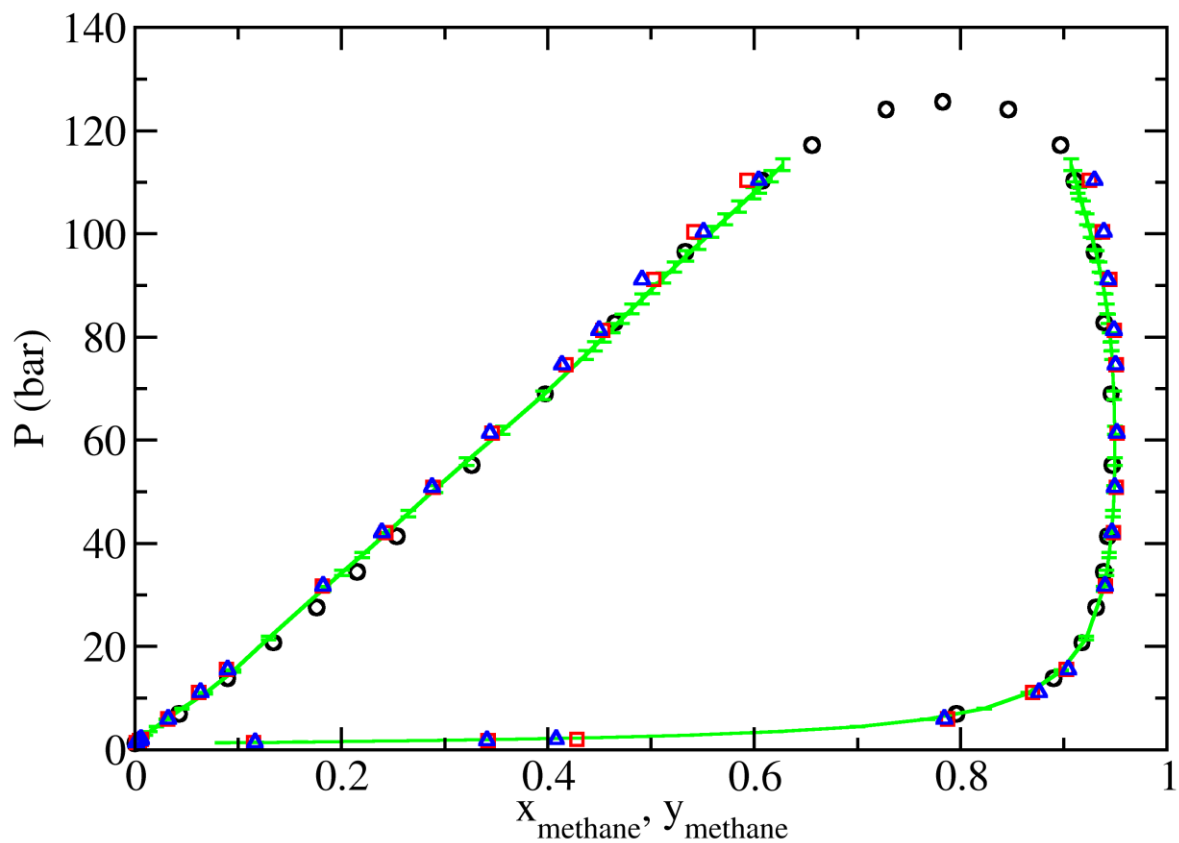
Figure C3: Pressure composition diagram for methane+n-butane at 277 K predicted from NPT-GEMC simulations using Mie potentials[24] using an exchange ratio of one n-butane with two methane molecules. Experimental data (black circles)[181], reference data[83] (green lines), ME-2 algorithm (red squares), and ME-3 algorithm (blue triangles). Uncertainty for methane composition is less than 0.01 and 0.004 in liquid and vapor phase, respectively.

Table C1: Average solute transfer acceptance percentages in GEMC simulations for n-alkane solvation in 1-octanol, n-hexadecane, or +2,2,4-trimethylpentane, using Mie potentials[24, 124]. The coupled-decoupled configurational-bias swap acceptance percentages are presented for the small solute swap. The acceptance percentages for ME-2 and ME-3 are for exchanging a small solute with a large one.

Solvent	Solute (small)	Solute (large)	CD-CBMC	ME-2	ME-3
n-hexadecane	methane	ethane	0.9948	6.2464	12.6108
	ethane	propane	0.3492	6.6767	3.7927
	propane	n-butane	0.0564	3.3728	1.1677
	n-butane	n-pentane	0.0175	2.9190	0.5999
	n-pentane	n-hexane	0.0054	3.6609	0.4923
	n-hexane	n-heptane	0.0015	2.6971	0.2993
	n-heptane	n-octane	0.0002	1.1158	0.0706
	n-octane	-	0.0000	-	-
2,2,4-trimethylpentane	methane	ethane	3.5123	12.6102	24.2642
	ethane	propane	1.7650	13.4010	9.2419
	propane	n-butane	0.4625	7.5406	3.5138
	n-butane	n-pentane	0.1920	6.2619	1.8375
	n-pentane	n-hexane	0.0771	6.6660	1.1497
	n-hexane	n-heptane	0.0231	3.5472	0.4711
	n-heptane	n-octane	0.0054	1.0811	0.1042
	n-octane	-	0.0009	-	-

APPENDIX D

In this section, the bonded and nonbonded potential parameters that has been used in Chapter 6, followed up with additional results, are provided.

Table D1: Bonded parameters for alcohols, fluoroalcohols and fluorotelomer alcohols[187, 203, 204].

Bond type	Bond length (Å)	Angle type	θ_0 (Degrees)	k_θ/k_b (kcal/mol/rad^2)
CH_x-CH_y	1.54	CH_x-CH_2-CH_y	114.0	62.1
CF_x-CF_y	1.54	CH_y-CH_2-O	109.5	50.1
CH_x-CF_y	1.54	CH_x-O-H	108.5	55.0
CH_x-O	1.43	CF_x-CF_2-CF_y	114.0	62.1
CF_x-O	1.43	CF_x-CF_2-O	109.5	62.1
O-H	0.945	CF_x-O-H	108.5	55.0
O-H (water)	1.00	CF_x-CH_2-CH_y	114.0	62.1
		CF_x-CF_2-CH_y	114.0	62.1
		CF_x-CH_2-O	109.5	62.1
		CH_x-CF_y-O	109.5	62.1
		H-O-H	109.47	Fix

Table D2: Dihedral parameters for alcohols, fluoroalcohols and fluorotelomer alcohols[187, 203, 204].

Torsion type	n	c_n/k_b (kcal/mol)	δ_n (Degrees)
CH_x-CH_2-CH_2-CH_y	1	0.705513	0
	2	-0.135507	180
	3	1.572510	0
O-CH_2-CH_2-CH_x	1	0.350977	0
	2	-0.105997	180
	3	1.529998	0
H-O-CH_2-CH_x	1	0.416952	0
	2	-0.057966	180
	3	0.373453	0
CF_x-CF_2-CF_2-CF_y	1	1.588	0
	2	-0.6481	180
	3	1.712	0
	4	-0.6791	180
O-CF_2-CF_2-CF_x	1	-0.0178	0
	2	0.0836	0
	3	1.6976	0
	4	0.0392	0
H-O-CF_2-CF_x	1	0.8392	0
	2	-0.1096	180
	3	0.6556	0
CH_2-CH_2-CF_2-CF_x	1	0.8945	0
	2	-0.5789	180
	3	1.8605	0
	4	-0.1634	180

[*]Optimized dihedral in this work.

Table D2: Continuation of dihedral parameters for alcohols, fluoroalcohols and fluorotelomer alcohols[187, 203, 204].

Torsion type	n	c_n/k_b (kcal/mol)	δ_n (Degrees)
CH_2-CF_2-CF_2-CF_x	1	1.588	0
	2	-0.6481	180
	3	1.712	0
	4	-0.6791	180
*CH_2-CH_2-CH_2-CF_x	1	1.5522	0
	2	-0.8265	180
	3	1.4588	0
	4	-0.1063	180
O-CH_2-CH_2-CF_x	1	1.5951	0
	2	-1.0807	180
	3	1.6495	0
O-CH_2-CF_2-CF_x	1	-0.0421	0
	2	-0.0604	180
	3	2.3476	0
*O-CF_2-CF_2-CH_x	1	-0.2315	0
	2	0.0881	180
	3	2.7794	0
	4	0.1366	0
*O-CF_2-CH_2-CH_x	3	2.0187	0
	4	0.0800	0
H-O-CH_2-CF_x	1	-0.5760	0
	2	0.9738	0
	3	0.8986	0
	4	0.2396	0
*H-O-CF_2-CH_x	1	1.793	0
	2	0.6984	180
	3	0.5409	0

*Optimized dihedral in this work.

Using the specified λ vectors, initial lambda state, free energy parameters, and free energy calculation frequency in the configuration file, GOMC will output the necessary information for free energy analysis. In addition to the simulation parameters, such as temperature and current lambda value ($\lambda_{coul}, \lambda_{LJ}$), GOMC will output the total energy of the system, derivative of energy with respect to lambda for coulomb and LJ potential (for TI free energy method), energy difference between current λ vector and all other λ vectors (for free energy perturbation method), and PV term (for NPT ensemble). In Table D3, a sample of free energy data outputted by GOMC is provided.

Table D3: Sample of GOMC free energy data for first 10,000 Monte Carlo Steps of perfluoro-octanol in octanol. The temperature and λ state used in simulation is printed in the first line, while the header of the each column is provided in the second line with energy unit of kJ/mol. The column's headers from left to right are, the simulation steps, the total energy of the system, energy derivative with respect to λ for coulomb and LJ, the total energy difference evaluated between current lambda state (λ_2) and all other lambda states ($\lambda_0, \lambda_1, \lambda_2, ..., \lambda_{16}$ in this case), and pressure x volume information for NPT ensemble.

$T = 298$ (K)		$\lambda_2 = (0.0, 0.2)$							
Steps	U_{total}	$\dfrac{dU_{coul}}{d\lambda_{2,coul}}$	$\dfrac{dU_{LJ}}{d\lambda_{2,LJ}}$	$\Delta U_{2\to 0}$	$\Delta U_{2\to 1}$	$\Delta U_{2\to 2}$	...	$\Delta U_{2\to 16}$	PV
1000	-8877.456	-3.231	8.077	10.004	2.624	0.000	...	2266.677	3.181
2000	-8835.898	2.683	110.732	-0.256	-4.425	0.000	...	47079.538	3.164
3000	-8827.969	1.666	245.945	-14.514	-14.142	0.000	...	19780.139	3.188
4000	-8841.234	8.217	61.476	3.994	-1.296	0.000	...	6005.224	3.170
5000	-8830.506	-5.707	127.468	-1.455	-5.509	0.000	...	47112.336	3.173
6000	-8809.501	-14.851	35.057	6.910	0.617	0.000	...	1235.585	3.183
7000	-8824.169	5.971	50.141	5.522	-0.361	0.000	...	18668.941	3.176
8000	-8814.229	2.230	9.339	9.444	2.385	0.000	...	1126.329	3.190
9000	-8826.988	7.380	60.722	4.220	-1.148	0.000	...	13442.642	3.172
10000	-8819.957	-11.471	37.901	6.680	0.422	0.000	...	1383.572	3.171

Table D4: Comparison of solvation free energies for n-alkanes in 1-octanol at 298 K and 1 atm, calculated with the TraPPE force field[187, 203, 204] using TI, MBAR methods, and NPT Gibbs ensemble Monte Carlo (NPT-GEMC) simulations. Number in parenthesis corresponds to statistical uncertainties in the last digit.

	Free energy of solvation (kcal/mol)				
Solute \ Method	MEMC-2[84]	MEMC-3[84]	TI	MBAR	Experiment[229]
n-pentane	-2.42(35)	-2.31(30)	-2.48(5)	-2.42(4)	-2.45
n-hexane	-3.02(35)	-2.94(35)	-3.05(6)	-3.00(5)	-3.01
n-heptane	-3.63(37)	-3.52(41)	-3.79(7)	-3.71(5)	-3.74
n-octane	-4.25(40)	-4.13(39)	-4.36(8)	-4.27(6)	-4.18

Table D5: Predicted free energies of hydration/solvation using TI, BAR, and MBAR method implemented in alchemlyb[257].

Molecule	ΔG_{C16}(kcal/mol)			$\Delta G_{1-octanol}$(kcal/mol)			ΔG_{water}(kcal/mol)		
	TI	BAR	MBAR	TI	BAR	MBAR	TI	BAR	MBAR
$CH_3(CH_2)_7OH$ (H8)	-5.14(6)	-5.14(6)	-5.15(5)	-8.8(2)	-8.7(2)	-8.6(2)	-3.1(2)	-2.9(2)	-2.9(2)
$CH_3(CH_2)_6CF_2OH$ (F1H7)	-4.41(7)	-4.35(7)	-4.38(6)	-6.3(2)	-6.1(2)	-6.1(2)	-1.8(3)	-1.7(3)	-1.4(2)
$CH_3(CH_2)_5(CF_2)_2OH$ (F2H6)	-4.29(8)	-4.26(8)	-4.34(6)	-6.0(2)	-5.8(2)	-5.7(2)	-1.3(2)	-1.2(2)	-1.3(2)
$CF_3(CF_2)_5(CH2)_2OH$ (H2F6)	-4.20(8)	-4.16(8)	-4.16(7)	-6.7(3)	-6.8(3)	-7.1(2)	-2.3(3)	-2.0(3)	-1.7(2)
$CF_3(CF_2)_6CH_2OH$ (H1F7)	-4.06(8)	-4.06(8)	-4.10(7)	-6.0(2)	-5.9(2)	-6.0(2)	-1.7(3)	-1.6(3)	-1.6(2)
$CF_3(CF_2)_7OH$ (F8)	-3.38(8)	-3.34(8)	-3.32(7)	-5.3(2)	-5.2(2)	-5.2(2)	0.3(5)	0.0(3)	0.0(3)
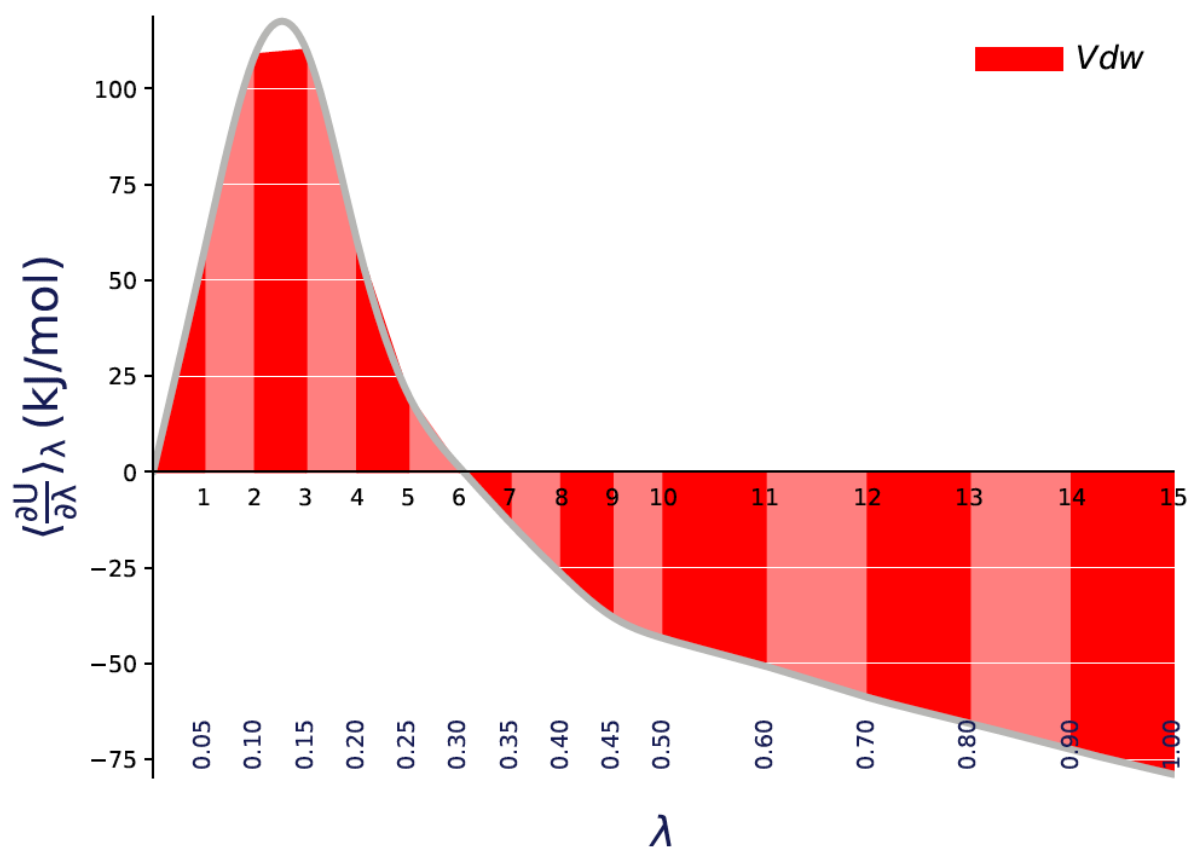
Figure D1: The $\langle \frac{dU}{d\lambda} \rangle$ versus λ plot for F2H6 solvation free energy in hexadecane, with filled areas indicating free energy estimates from the trapezoid rule, and silver curve indicating interpolation via cubic spline.
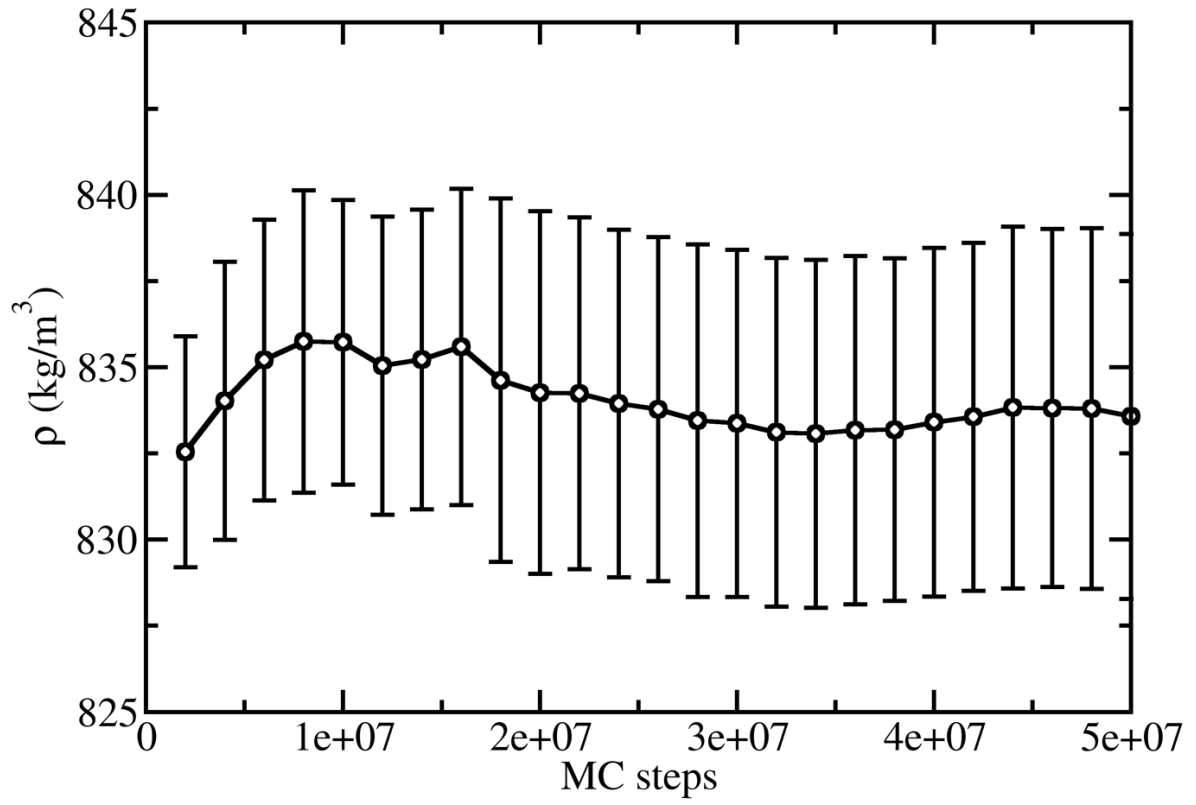
Figure D2: Accumulative average density for perfluorooctanol in 1-octanol.

REFERENCES

[1] Warton, B., Alexander, R. and Kagi, R. I. Identification of some single branched alkanes in crude oils. *Org. Geochem.*, 27, 7-8 (Nov-Dec 1997), 465-476.

[2] Bendoraitis, J. G., Hepner, L. S. and Brown, B. L. Isoprenoid Hydrocarbons in Petroleum - Isolation of 2,6,10,14-Tetramethylpentadecane by High Temperature Gas-Liquid Chromatography. *Anal. Chem.*, 34, 1 (1962), 49-53.

[3] Yu, J. and Eser, S. Kinetics of supercritical-phase thermal decomposition of C-10-C-14 normal alkanes and their mixtures. *Ind. Eng. Chem. Res.*, 36, 3 (Mar 1997), 585-591.

[4] Tsang, W. Thermal Decomposition of Hexamethylethane 2,2,3-Trimethylbutane and Neopentane in a Single-Pulse Shock Tube. *J. Chem. Phys.*, 44, 11 (1966), 4283-4295.

[5] Glotzer, S. C. and Solomon, M. J. Anisotropy of building blocks and their assembly into complex structures. *Nature Mater.*, 6, 8 (Aug 2007), 557-562.

[6] van Anders, G., Klotsa, D., Karas, A. S., Dodd, P. M. and Glotzer, S. C. Digital Alchemy for Materials Design: Colloids and Beyond. *Acs Nano*, 9, 10 (Oct 2015), 9542-9553.

[7] Zhang, Z. L. and Glotzer, S. C. Self-assembly of patchy particles. *Nano Lett.*, 4, 8 (Aug 2004), 1407-1413.

[8] LeBard, D. N., Levine, B. G., Mertmann, P., Barr, S. A., Jusufi, A., Sanders, S., Klein, M. L. and Panagiotopoulos, A. Z. Self-assembly of coarse-grained ionic surfactants accelerated by graphics processing units. *Soft Matter*, 8, 8 (2012), 2385-2397.

[9] Zhang, Y., Schulten, K., Gruebele, M., Bansal, P. S., Wilson, D. and Daly, N. L. Disulfide Bridges: Bringing Together Frustrated Structure in a Bioactive Peptide. *Biophys. J.*, 110, 8 (Apr 26 2016), 1744-1752.

[10] Perilla, J. R. and Schulten, K. Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations. *Nat Commun.*, 8 (Jul 19 2017), 15959.

[11] Klauda, J. B., Kucerka, N., Brooks, B. R., Pastor, R. W. and Nagle, J. F. Simulation-based methods for interpreting X-ray data from lipid bilayers. *Biophys. J.*, 90, 8 (Apr 2006), 2796-2807.

[12] Jorgensen, W. L. The many roles of computation in drug discovery. *Science*, 303, 5665 (Mar 19 2004), 1813-1818.

[13] Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.*, 42, 6 (Jun 2009), 724-733.

[14] Zhong, S., Chen, X., Zhu, X., Dziegielewska, B., Bachman, K. E., Ellenberger, T., Ballin, J. D., Wilson, G. M., Tomkinson, A. E. and MacKerell, A. D. Identification and validation of human DNA ligase inhibitors using computer-aided drug design. *J. Med. Chem.*, 51, 15 (Aug 14 2008), 4553-4562.

[15] Wilmer, C. E. and Snurr, R. Q. Towards rapid computational screening of metal-organic frameworks for carbon dioxide capture: Calculation of framework charges via charge equilibration. *Chem. Eng. J.*, 171, 3 (Jul 15 2011), 775-781.

[16] Kim, J., Lin, L. C., Martin, R. L., Swisher, J. A., Haranczyk, M. and Smit, B. Large-scale computational screening of zeolites for ethane/ethene separation. *Langmuir*, 28, 32 (Aug 14 2012), 11914-11919.

[17] Banerjee, D., Simon, C. M., Plonka, A. M., Motkuri, R. K., Liu, J., Chen, X. Y., Smit, B., Parise, J. B., Haranczyk, M. and Thallapally, P. K. Metal-organic framework with optimally selective xenon adsorption and separation. *Nature Communications*, 7 (Jun 2016).

[18] Wilmer, C. E., Leaf, M., Lee, C. Y., Farha, O. K., Hauser, B. G., Hupp, J. T. and Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nature Chem.*, 4, 2 (Feb 2012), 83-89.

[19] Martin, R. L., Lin, L. C., Jariwala, K., Smit, B. and Haranczyk, M. Mail-Order Metal-Organic Frameworks (MOFs): Designing Isoreticular MOF-5 Analogues Comprising

Commercially Available Organic Molecules. *J. Phys. Chem. C*, 117, 23 (Jun 13 2013), 12159-12167.

[20] Keskin, S., Liu, J., Rankin, R. B., Johnson, J. K. and Sholl, D. S. Progress, Opportunities, and Challenges for Applying Atomically Detailed Modeling to Molecular Adsorption and Transport in Metal-Organic Framework Materials. *Ind. Eng. Chem. Res.*, 48, 5 (Mar 4 2009), 2355-2371.

[21] Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D. and Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B*, 114, 23 (Jun 17 2010), 7830-7843.

[22] Savelyev, A. and MacKerell, A. D. Balancing the Interactions of Ions, Water, and DNA in the Drude Polarizable Force Field. *J. Phys. Chem. B*, 118, 24 (Jun 19 2014), 6742-6757.

[23] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. and MacKerell, A. D. CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.*, 31, 4 (Mar 2010), 671-690.

[24] Potoff, J. J. and Bernard-Brunel, D. A. Mie Potentials for Phase Equilibria Calculations: Application to Alkanes and Perfluoroalkanes. *J. Phys. Chem. B*, 113, 44 (Nov 5 2009), 14725-14731.

[25] Jambeck, J. P. M. and Lyubartsev, A. P. Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids. *J. Phys. Chem. B*, 116, 10 (Mar 15 2012), 3164-3179.

[26] Siu, S. W. I., Pluhackova, K. and Bockmann, R. A. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. *J. Chem. Theory Comput.*, 8, 4 (Apr 2012), 1459-1470.

[27] Dickson, C. J., Rosso, L., Betz, R. M., Walker, R. C. and Gould, I. R. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter*, 8, 37 (2012), 9617-9627.

[28] Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. and Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.*, 25, 9 (Jul 15 2004), 1157-1174.

[29] Laury, M. L., Wang, L. P., Pande, V. S., Head-Gordon, T. and Ponder, J. W. Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model. *J. Phys. Chem. B*, 119, 29 (Jul 23 2015), 9423-9437.

[30] Darve, E. and Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.*, 115, 20 (Nov 22 2001), 9169-9183.

[31] Chodera, J. D., Swope, W. C., Noe, F., Prinz, J. H., Shirts, M. R. and Pande, V. S. Dynamical reweighting: Improved estimates of dynamical properties from simulations at multiple temperatures. *J. Chem. Phys.*, 134, 24 (Jun 28 2011).

[32] Panagiotopoulos, A. Z. Direct Determination of Phase Coexistence Properties of Fluids by Monte-Carlo Simulation in a New Ensemble. *Mol. Phys.*, 61, 4 (Jul 1987), 813-826.

[33] Balsera, M., Stepaniants, S., Izrailev, S., Oono, Y. and Schulten, K. Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys. J.*, 73, 3 (Sep 1997), 1281-1287.

[34] Chen, B. and Siepmann, J. I. Monte Carlo algorithms for simulating systems with adiabatic separation of electronic and nuclear degrees of freedom. *Theor. Chem. Acc.*, 103, 2 (Dec 1999), 87-104.

[35] Lamoureux, G. and Roux, B. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *J. Chem. Phys.*, 119, 6 (Aug 8 2003), 3025-3039.

[36] Jiang, W., Hardy, D. J., Phillips, J. C., MacKerell, A. D., Schulten, K. and Roux, B. High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *J. Phys. Chem. Lett.*, 2, 2 (Jan 20 2011), 87-92.

[37] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. and Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, 26, 16 (Dec 2005), 1781-1802.

[38] Brown, W. M., Wang, P., Plimpton, S. J. and Tharrington, A. N. Implementing molecular dynamics on hybrid high performance computers - short range forces. *Comput. Phys. Commun.*, 182, 4 (Apr 2011), 898-911.

[39] Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S. and Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.*, 9, 9 (Sep 2013), 3878-3888.

[40] Nguyen, T. D., Phillips, C. L., Anderson, J. A. and Glotzer, S. C. Rigid body constraints realized in massively-parallel molecular dynamics on graphics processing units. *Comput. Phys. Commun.*, 182, 11 (Nov 2011), 2307-2313.

[41] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B. and Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Softwarex*, 1–2 (9// 2015), 19-25.

[42] Brooks, B. R., Brooks, C. L., Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. and Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, 30, 10 (Jul 30 2009), 1545-1614.

[43] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.*, 26, 16 (Dec 2005), 1668-1688.

[44] Hynninen, A. P. and Crowley, M. F. New Faster CHARMM Molecular Dynamics Engine. *J. Comput. Chem.*, 35, 5 (Feb 15 2014), 406-413.

[45] Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J. Comput. Phys.*, 117, 1 (Mar 1 1995), 1-19.

[46] Anderson, J. A., Lorenz, C. D. and Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comput. Phys.*, 227, 10 (May 1 2008), 5342-5359.

[47] Martin, M. G. MCCCS Towhee: a tool for Monte Carlo molecular simulation. *Mol. Simulat.*, 39, 14-15 (Dec 1 2013), 1184-1194.

[48] Gowers, R. J., Farmahini, A. H., Friedrich, D. and Sarkisov, L. Automated analysis and benchmarking of GCMC simulation programs in application to gas adsorption. *Mol. Simulat.*, 44, 4 (2018), 309-321.

[49] Anderson, J. A., Jankowski, E., Grubb, T. L., Engel, M. and Glotzer, S. C. Massively parallel Monte Carlo for many-particle simulations on GPUs. *J. Comput. Phys.*, 254 (Dec 1 2013), 27-38.

[50] Schultz, A. J. and Kofke, D. A. Etomica: An Object-Oriented Framework for Molecular Simulation. *J. Comput. Chem.*, 36, 8 (Mar 30 2015), 573-583.

[51] Hatch, H. W., Mahynski, N. A. and Shen, V. K. FEASST: Free Energy and Advanced Sampling Simulation Toolkit. *J. Res. Natl. Inst. Stan.*, 123 (Mar 1 2018).

[52] Deublein, S., Eckl, B., Stoll, J., Lishchuk, S. V., Guevara-Carrion, G., Glass, C. W., Merker, T., Bernreuther, M., Hasse, H. and Vrabec, J. ms2: A molecular simulation tool for thermodynamic properties. *Comput. Phys. Commun.*, 182, 11 (Nov 2011), 2350-2367.

[53] Glass, C. W., Reiser, S., Rutkai, G., Deublein, S., Koster, A., Guevara-Carrion, G., Wafai, A., Horsch, M., Bernreuther, M., Windmann, T., Hasse, H. and Vrabec, J. ms2: A molecular simulation tool for thermodynamic properties, new version release. *Comput. Phys. Commun.*, 185, 12 (Dec 2014), 3302-3306.

[54] Rutkai, G., Koster, A., Guevara-Carrion, G., Janzen, T., Schappals, M., Glass, C. W., Bernreuther, M., Wafai, A., Stephan, S., Kohns, M., Reiser, S., Deublein, S., Horsch, M., Hasse, H. and Vrabec, J. ms2: A molecular simulation tool for thermodynamic properties, release 3.0. *Comput. Phys. Commun.*, 221 (Dec 2017), 343-351.

[55] Dubbeldam, D., Calero, S., Ellis, D. E. and Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simulat.*, 42, 2 (Jan 22 2016), 81-101.

[56] Shah, J. K., Marin-Rimoldi, E., Mullen, R. G., Keene, B. P., Khan, S., Paluch, A. S., Rai, N., Romanielo, L. L., Rosch, T. W., Yoo, B. and Maginn, E. J. Cassandra: An open source Monte Carlo package for molecular simulation. *J. Comput. Chem.*, 38, 19 (Jul 15 2017), 1727-1739.

[57] Nejahi, Y., Barhaghi, M. S., Mick, J., Jackman, B., Rushaidat, K., Li, Y. Z., Schwiebert, L. and Potoff, J. GOMC: GPU Optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids. *Softwarex*, 9 (Jan-Jun 2019), 20-27.

[58] Pässler, P., Hefner, W., Buckl, K., Meinass, H., Meiswinkel, A., Wernicke, H.-J., Ebersberg, G., Müller, R., Bässler, J., Behringer, H. and Mayer, D. *Acetylene*. Wiley-VCH Verlag GmbH & Co. KGaA, City, 2000.

[59] Walker, S., Landovitz, R., Ding, W. D., Ellestad, G. A. and Kahne, D. Cleavage behavior of calicheamicin gamma 1 and calicheamicin T. *Proc. Natl. Acad. Sci.*, 89, 10 (1992), 4608-4612.

[60] Stapleton, M. R., Tildesley, D. J., Panagiotopoulos, A. Z. and Quirke, N. Phase Equilibria of Quadrupolar Fluids by Simulation in the Gibbs Ensemble. *Mol. Simulat.*, 2, 3 (1989), 147-162.

[61] Vrabec, J., Stoll, J. and Hasse, H. A set of molecular models for symmetric quadrupolar fluids. *J. Phys. Chem. B*, 105, 48 (Dec 6 2001), 12126-12133.

[62] Garrison, S. L. and Sandler, S. I. An accurate acetylene intermolecular potential for phase behavior predictions from quantum chemistry. *J. Phys. Chem. B*, 108, 49 (Dec 9 2004), 18972-18979.

[63] Unlu, O., Gray, N. H., Gerek, Z. N. and Elliott, J. R. Transferable step potentials for the straight-chain alkanes, alkenes, alkynes, ethers, and alcohols. *Ind. Eng. Chem. Res.*, 43, 7 (Mar 31 2004), 1788-1793.

[64] Barhaghi, M. S., Mick, J. R. and Potoff, J. J. Optimised Mie potentials for phase equilibria: application to alkynes. *Mol. Phys.*, 115, 9-12 (2017), 1378-1388.

[65] Snurr, R. Q., Bell, A. T. and Theodorou, D. N. Prediction of Adsorption of Aromatic-Hydrocarbons in Silicalite from Grand-Canonical Monte-Carlo Simulations with Biased Insertions. *J. Phys. Chem.*, 97, 51 (Dec 23 1993), 13742-13752.

[66] Smit, B. Grand-Canonical Monte-Carlo Simulations of Chain Molecules - Adsorption-Isotherms of Alkanes in Zeolites. *Mol. Phys.*, 85, 1 (May 1995), 153-172.

[67] Martin, M. G. and Siepmann, J. I. Novel configurational-bias Monte Carlo method for branched molecules. Transferable potentials for phase equilibria. 2. United-atom description of branched alkanes. *J. Phys. Chem. B*, 103, 21 (May 27 1999), 4508-4517.

[68] Stapleton, M. R. and Panagiotopoulos, A. Z. Application of Excluded Volume Map Sampling to Phase-Equilibrium Calculations in the Gibbs Ensemble. *J. Chem. Phys.*, 92, 2 (Jan 15 1990), 1285-1293.

[69] Ortiz, V., Maury-Evertsz, J. R. and Lopez, G. E. Parallel tempering-cavity-bias algorithm in the Gibbs ensemble. *Chem. Phys. Lett.*, 368, 3-4 (Jan 17 2003), 452-457.

[70] Mezei, M. A Cavity-Biased (T,V,Mu) Monte-Carlo Method for the Computer-Simulation of Fluids. *Mol. Phys.*, 40, 4 (1980), 901-906.

[71] Shi, W. and Maginn, E. J. Continuous fractional component Monte Carlo: An adaptive biasing method for open system atomistic simulations. *J. Chem. Theory Comput.*, 3, 4 (Jul-Aug 2007), 1451-1463.

[72] Torres-Knoop, A., Balaji, S. P., Vlugt, T. J. H. and Dubbeldam, D. A Comparison of Advanced Monte Carlo Methods for Open Systems: CFCMC vs CBMC. *J. Chem. Theory Comput.*, 10, 3 (Mar 2014), 942-952.

[73] Escobedo, F. A. and Martinez-Veracoechea, F. J. Optimization of expanded ensemble methods. *J. Chem. Phys.*, 129, 15 (Oct 21 2008).

[74] Escobedo, F. A. and dePablo, J. J. Expanded grand canonical and Gibbs ensemble Monte Carlo simulation of polymers. *J. Chem. Phys.*, 105, 10 (Sep 8 1996), 4391-4394.

[75] Panagiotopoulos, A. Z. Exact Calculations of Fluid-Phase Equilibria by Monte-Carlo Simulation in a New Statistical Ensemble. *Int. J. Thermophys.*, 10, 2 (Mar 1989), 447-457.

[76] Wijmans, C. M., Smit, B. and Groot, R. D. Phase behavior of monomeric mixtures and polymer solutions with soft interaction potentials. *J. Chem. Phys.*, 114, 17 (May 1 2001), 7644-7654.

[77] Martin, M. G. and Siepmann, J. I. Predicting multicomponent phase equilibria and free energies of transfer for alkanes by molecular simulation. *J. Am. Chem. Soc.*, 119, 38 (Sep 24 1997), 8921-8924.

[78] Chen, B., Siepmann, J. I. and Klein, M. L. Vapor-liquid interfacial properties of mutually saturated water/1-butanol solutions. *J. Am. Chem. Soc.*, 124, 41 (Oct 16 2002), 12232-12237.

[79] Wick, C. D., Siepmann, J. I. and Theodorou, D. N. Microscopic origins for the favorable solvation of carbonate ether copolymers in CO2. *J. Am. Chem. Soc.*, 127, 35 (Sep 7 2005), 12338-12342.

[80] Kofke, D. A. and Glandt, E. D. Monte-Carlo Simulation of Multicomponent Equilibria in a Semigrand Canonical Ensemble. *Mol. Phys.*, 64, 6 (Aug 20 1988), 1105-1131.

[81] Vink, R. L. C. and Horbach, J. Grand canonical Monte Carlo simulation of a model colloid-polymer mixture: Coexistence line, critical behavior, and interfacial tension. *J. Chem. Phys.*, 121, 7 (Aug 15 2004), 3253-3258.

[82] Kindt, J. T. Grand canonical Monte Carlo using solvent repacking: Application to phase behavior of hard disk mixtures. *J. Chem. Phys.*, 143, 12 (Sep 28 2015).

[83] Barhaghi, M. S., Torabi, K., Nejahi, Y., Schwiebert, L. and Potoff, J. J. Molecular exchange Monte Carlo: A generalized method for identity exchanges in grand canonical Monte Carlo simulations. *J. Chem. Phys.*, 149, 7 (Aug 21 2018), 072318.

[84] Barhaghi, M. S. and Potoff, J. J. Prediction of phase equilibria and Gibbs free energies of transfer using molecular exchange Monte Carlo in the Gibbs ensemble. *Fluid Phase Equilibr.*, 486 (May 1 2019), 106-118.

[85] Shinoda, K., Hato, M. and Hayashi, T. Physicochemical Properties of Aqueous-Solutions of Fluorinated Surfactants. *J. Phys. Chem.*, 76, 6 (1972), 909-&.

[86] Bernett, M. K. and Zisman, W. A. Wetting of Low-Energy Solids by Aqueous Solutions of Highly Fluorinated Acids and Salts. *J. Phys. Chem.*, 63, 11 (1959), 1911-1916.

[87] Scholberg, H. M., Guenthner, R. A. and Coon, R. I. Surface Chemistry of Fluorocarbons and Their Derivatives. *J. Phys. Chem.*, 57, 9 (1954), 923-925.

[88] Thomas, P. The use of fluoropolymers for non-stick cooking utensils. *Surf. Coat. Int.*, 81, 12 (Dec 1998), 604-609.

[89] Trier, X., Granby, K. and Christensen, J. H. Polyfluorinated surfactants (PFS) in paper and board coatings for food packaging. *Environ. Sci. Pollut. R.*, 18, 7 (Aug 2011), 1108-1120.

[90] Moody, C. A. and Field, J. A. Perfluorinated surfactants and the environmental implications of their use in fire-fighting foams. *Environ. Sci. Technol.*, 34, 18 (Sep 15 2000), 3864-3870.

[91] Prevedouros, K., Cousins, I. T., Buck, R. C. and Korzeniowski, S. H. Sources, fate and transport of perfluorocarboxylates. *Environ. Sci. Technol.*, 40, 1 (Jan 1 2006), 32-44.

[92] Liou, J. S. C., Szostek, B., DeRito, C. M. and Madsen, E. L. Investigating the biodegradability of perfluorooctanoic acid. *Chemosphere*, 80, 2 (Jun 2010), 176-183.

[93] Martin, J. W., Smithwick, M. M., Braune, B. M., Hoekstra, P. F., Muir, D. C. G. and Mabury, S. A. Identification of long-chain perfluorinated acids in biota from the Canadian Arctic. *Environ. Sci. Technol.*, 38, 2 (Jan 15 2004), 373-380.

[94] Key, B. D., Howell, R. D. and Criddle, C. S. Fluorinated organics in the biosphere. *Environ. Sci. Technol.*, 31, 9 (Sep 1997), 2445-2454.

[95] Scheringer, M., Trier, X., Cousins, I. T., de Voogt, P., Fletcher, T., Wang, Z. Y. and Webster, T. F. Helsingor Statement on poly- and perfluorinated alkyl substances (PFASs). *Chemosphere*, 114 (Nov 2014), 337-339.

[96] Wania, F. and Mackay, D. A Global Distribution Model for Persistent Organic-Chemicals. *Sci. Total Environ.*, 160-61 (Jan 15 1995), 211-232.

[97] *Estimation Programs Interface Suite™ for Microsoft® Windows*. United States Environmental Protection Agency, Washington, DC, USA, City, 2012.

[98] Leo, A. J. *CLOGP*. City, 1991.

[99] Hilal, S. H., Karickhoff, S. W. and Carreira, L. A. Prediction of the solubility, activity coefficient and liquid/liquid partition coefficient of organic compounds. *QSAR Comb. Sci.*, 23, 9 (Nov 2004), 709-720.

[100] Eckert, F. and Klamt, A. Fast solvent screening via quantum chemistry: COSMO-RS approach. *AIChE J.*, 48, 2 (Feb 2002), 369-385.

[101] Arp, H. P. H., Niederer, C. and Goss, K. U. Predicting the partitioning behavior of various highly fluorinated compounds. *Environ. Sci. Technol.*, 40, 23 (Dec 1 2006), 7298-7304.

[102] Straatsma, T. P. and Mccammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.*, 43 (1992), 407-435.

[103] Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method .2. Polar Gases. *J. Chem. Phys.*, 23, 10 (1955), 1915-1922.

[104] Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method .1. Nonpolar Gases. *J. Chem. Phys.*, 22, 8 (1954), 1420-1426.

[105] Rodriguez-Gomez, D., Darve, E. and Pohorille, A. Assessing the efficiency of free energy calculation methods. *J. Chem. Phys.*, 120, 8 (Feb 22 2004), 3563-3578.

[106] Mobley, D. L., Wymer, K. L., Lim, N. M. and Guthrie, J. P. Blind prediction of solvation free energies from the SAMPL4 challenge. *J. Comput. Aided Mol. Des.*, 28, 3 (Mar 2014), 135-150.

[107] Bannan, C. C., Burley, K. H., Chiu, M., Shirts, M. R., Gilson, M. K. and Mobley, D. L. Blind prediction of cyclohexane-water distribution coefficients from the SAMPL5 challenge. *J. Comput. Aided Mol. Des.*, 30, 11 (Nov 2016), 927-944.

[108] Mobley, D. L., Bayly, C. I., Cooper, M. D. and Dill, K. A. Predictions of Hydration Free Energies from All-Atom Molecular Dynamics Simulations. *J. Phys. Chem. B*, 113, 14 (Apr 9 2009), 4533-4537.

[109] Mobley, D. L., Bayly, C. I., Cooper, M. D., Shirts, M. R. and Dill, K. A. Small Molecule Hydration Free Energies in Explicit Solvent: An Extensive Test of Fixed-Charge Atomistic Simulations. *J. Chem. Theory Comput.*, 5, 2 (Feb 2009), 350-358.

[110] Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K. and Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Current Opinion in Structural Biology*, 21, 2 (APR 2011 2011), 150-160.

[111] Bhatnagar, N., Kamath, G. and Potoff, J. J. Prediction of 1-octanol-water and air-water partition coefficients for nitro-aromatic compounds from molecular dynamics simulations. *Phys. Chem. Chem. Phys.*, 15, 17 (2013), 6467-6474.

[112] Ahmed, A. and Sandler, S. I. Hydration Free Energies of Multifunctional Nitroaromatic Compounds. *J. Chem. Theory Comput.*, 9, 6 (Jun 2013), 2774-2785.

[113] Kamath, G., Bhatnagar, N., Baker, G. A., Baker, S. N. and Potoff, J. J. Computational prediction of ionic liquid 1-octanol/water partition coefficients. *Phys. Chem. Chem. Phys.*, 14, 13 (2012), 4339-4342.

[114] Zygmunt, W. and Potoff, J. J. The effect of fluorination on the physical properties and the free energies of hydration of 1-alcohols. *Fluid Phase Equilibr.*, 407 (Jan 15 2016), 314-321.

[115] Klimovich, P. V. and Mobley, D. L. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *J. Comput. Aided Mol. Des.*, 24, 4 (Apr 2010), 307-316.

[116] Chen, B. and Siepmann, J. I. Partitioning of alkane and alcohol solutes between water and (Dry or wet) 1-Octanol. *J. Am. Chem. Soc.*, 122, 27 (JUL 12 2000 2000), 6464-6467.

[117] Poursaeidesfahani, A., Torres-Knoop, A., Dubbeldam, D. and Vlugt, T. J. H. Direct Free Energy Calculation in the Continuous Fractional Component Gibbs Ensemble. *J. Chem. Theory Comput.*, 12, 4 (Apr 2016), 1481-1490.

[118] Rai, N., Bhatt, D., Siepmann, J. I. and Fried, L. E. Monte Carlo simulations of 1,3,5-triamino-2,4,6-trinitrobenzene (TATB): Pressure and temperature effects for the solid phase and vapor-liquid phase equilibria. *J. Chem. Phys.*, 129, 19 (Nov 21 2008).

[119] Barhaghi, M. S., Luyet, C. and Potoff, J. J. Effect of Fluorination on the Partitioning of the Alcohols. *Mol. Phys.* (2019).

[120] Lorentz, H. A. Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase. *Ann. Phys-Berlin*, 248, 1 (1881), 127-136.

[121] Berthelot, D. Sur le melange des gaz. *C. R. Hebd. Seanc. Acad. Sci. (Paris)*, 126 (1898), 1703-1855.

[122] Ewald, P. P. The calculation of optical and electrostatic grid potential. *Ann. Phys-Berlin*, 64, 3 (Feb 1921), 253-287.

[123] Shi Shun, A. L. and Tykwinski, R. R. Synthesis of naturally occurring polyynes. *Angewandte Chemie International Edition*, 45, 7 (2006), 1034-1057.

[124] Mick, J. R., Barhaghi, M. S., Jackman, B., Schwiebert, L. and Potoff, J. J. Optimized Mie Potentials for Phase Equilibria: Application to Branched Alkanes. *J. Chem. Eng. Data*, 62, 6 (Jun 2017), 1806-1818.

[125] Potoff, J. J. and Kamath, G. Mie Potentials for Phase Equilibria: Application to Alkenes. *J. Chem. Eng. Data*, 59, 10 (2014), 3144-3150.

[126] Mick, J. R., Soroush Barhaghi, M., Jackman, B., Rushaidat, K., Schwiebert, L. and Potoff, J. J. Optimized Mie potentials for phase equilibria: Application to noble gases and their mixtures with n-alkanes. *J Chem Phys*, 143, 11 (Sep 21 2015), 114504.

[127] Mick, J. R., Barhaghi, M. S. and Potoff, J. J. Prediction of Radon-222 Phase Behavior by Monte Carlo Simulation. *J. Chem. Eng. Data*, 61, 4 (Apr 2016), 1625-1631.

[128] Messerly, R. A., Anderson, M. C., Razavi, S. M. and Elliott, J. R. Improvements and limitations of Mie λ-6 potential for prediction of saturated and compressed liquid viscosity. *Fluid Phase Equilibr.*, 483 (2019/03/15/ 2019), 101-115.

[129] Messerly, R. A., Anderson, M. C., Razavi, S. M. and Elliott, J. R. Mie 16–6 force field predicts viscosity with faster-than-exponential pressure dependence for 2,2,4-trimethylhexane. *Fluid Phase Equilibr.*, 495 (2019/09/01/ 2019), 76-85.

[130] Ferrenberg, A. M. and Swendsen, R. H. Optimized Monte-Carlo Data-Analysis. *Phys. Rev. Lett.*, 63, 12 (Sep 18 1989), 1195-1198.

[131] Ferrenberg, A. M. and Swendsen, R. H. New Monte-Carlo Technique for Studying Phase-Transitions. *Phys. Rev. Lett.*, 61, 23 (Dec 5 1988), 2635-2638.

[132] Razavi, S. M., Messerly, R. A. and Elliott, J. R. Coexistence calculation using the isothermal-isochoric integration method. *Fluid Phase Equilibr.*, 501 (2019/12/01/ 2019), 112236.

[133] Messerly, R. A., Razavi, S. M. and Shirts, M. R. Configuration-Sampling-Based Surrogate Models for Rapid Parameterization of Non-Bonded Interactions. *J. Chem. Theory Comput.*, 14, 6 (2018/06/12 2018), 3144-3162.

[134] Messerly, R. A., Soroush Barhaghi, M., Potoff, J. J. and Shirts, M. R. Histogram-Free Reweighting with Grand Canonical Monte Carlo: Post-simulation Optimization of Non-bonded Potentials for Phase Equilibria. *J. Chem. Eng. Data*, 64, 9 (2019/09/12 2019), 3701-3717.

[135] Stiegler, T. and Sadus, R. J. Molecular simulation of fluids with non-identical intermolecular potentials: Thermodynamic properties of 10-5+12-6 Mie potential binary mixtures. *J. Chem. Phys.*, 142, 8 (Feb 28 2015), 084504.

[136] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M.,

Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J. and Fox, D. J. *Gaussian 09, Revision A.02*. Gaussian Inc, City, 2009.

[137] Heath, G., Thomas, L., Sherrard, E. and Sheridan, J. The microwave spectrum and structure of methyl diacetylene. *Discussions of the Faraday Society*, 19 (1955), 38-43.

[138] Ramanathan, N., Sundararajan, K., Vidya, K. and Jemmis, E. D. Non-covalent C-Cl pi interaction in acetylene-carbon tetrachloride adducts: Matrix isolation infrared and ab initio computational studies. *Spectrochim Acta A*, 157 (Mar 15 2016), 69-78.

[139] Jorgensen, W. L., Madura, J. D. and Swenson, C. J. Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.*, 106, 22 (1984), 6638-6646.

[140] Jorgensen, W. L., Maxwell, D. S. and Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118, 45 (1996), 11225-11236.

[141] Potoff, J. J., Errington, J. R. and Panagiotopoulos, A. Z. Molecular simulation of phase equilibria for mixtures of polar and non-polar components. *Mol. Phys.*, 97, 10 (Nov 20 1999), 1073-1083.

[142] Martinez, L., Andrade, R., Birgin, E. G. and Martinez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.*, 30, 13 (Oct 2009), 2157-2164.

[143] Humphrey, W., Dalke, A. and Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph. Model.*, 14, 1 (Feb 1996), 33-38.

[144] Frenkel, D. and Smit, B. *Understanding molecular simulation: from algorithms to applications*. Academic Press, 2002.

[145] Martin, M. G. and Biddy, M. J. Monte Carlo molecular simulation predictions for the heat of vaporization of acetone and butyramide. *Fluid Phase Equilibr.*, 236, 1 (2005), 53-57.

[146] Rowlinson, J. S. and Widom, B. *Molecular theory of capillarity*. Courier Corporation, 2013.

[147] Rowlinson, J. S. and Swinton, F. *Liquids and Liquid Mixtures: Butterworths Monographs in Chemistry*. Butterworth-Heinemann, 2013.

[148] Potoff, J. J. and Panagiotopoulos, A. Z. Critical point and phase behavior of the pure fluid and a Lennard-Jones mixture. *J. Chem. Phys.*, 109, 24 (1998), 10914-10920.

[149] Perry, R. H., Green, D. W. and Maloney, J. O. *Perry's Chemical engineers' handbook*. McGraw-Hill, New York, 1984.

[150] Vohra, S. P., Kang, T. L., Kobe, K. A. and McKetta, J. J. P-V-T Properties of Propyne. *J. Chem. Eng. Data*, 7, 1 (1962/01/01 1962), 150-155.

[151] Smith, B. D. and Srivastava, R. *Thermodynamic Data for Pure Compounds: Part A Hydrocarbons and Ketones*. Elsevier, Amsterdam, 1986.

[152] Grzeskowiak, R., Jeffery, G. H. and Vogel, A. I. Physical Properties and Chemical Constitution .29. Acetylenic Compounds. *J. Chem. Soc.*, Dec (1960), 4719-4722.

[153] Henne, A. L. and Greenlee, K. W. Preparation and Physical Constants of Acetylenic Compounds. *J. Am. Chem. Soc.*, 67, 3 (1945), 484-485.

[154] Danner, R. P. and Daubert, T. E. *Manual for predicting chemical process design data: data prediction manual/Design Institute for Physical Property Data, American Institute of Chemical Engineers*. American Institute of Chemical Engineers, New York, 1983.

[155] Tsonopoulos, C. and Ambrose, D. Vapor−Liquid Critical Properties of Elements and Compounds. 6. Unsaturated Aliphatic Hydrocarbons. *J. Chem. Eng. Data*, 41, 4 (1996/01/01 1996), 645-656.

[156] Owczarek, I. and Blazej, K. Recommended critical temperatures. Part I. Aliphatic hydrocarbons. *J. Phys. Chem. Ref. Data*, 32, 4 (2003), 1411-1427.

[157] Keir, R. I., Lamb, D. W., Ritchie, G. L. and Watson, J. N. Polarizability anisotropies, magnetizability anisotropies and molecular quadrupole moments of acetylene, methylacetylene and dimethylacetylene. *Chem. Phys. Lett.*, 279, 1-2 (1997), 22-28.

[158] Lemmon, E. W., O., M. M. and Friend, D. G. Thermophysical Properties of Fluid Systems. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard, National Institute of Standards and Technology, Gaithersburg MD, 20899, http://webbook.nist.gov* (

[159] Kyser III, E. A. and Manley, D. B. Precision total pressure measurements of propyne with propane, propene, and propadiene. *J. Chem. Eng. Data*, 40, 4 (1995), 756-764.

[160] de Pablo, J. J., Laso, M., Siepmann, J. I. and Suter, U. W. Continuum-Configurational-Bias Monte-Carlo Simulations of Long-Chain Alkanes. *Mol. Phys.*, 80, 1 (Sep 1993), 55-63.

[161] Siepmann, J. I. and Frenkel, D. Configurational Bias Monte-Carlo - a New Sampling Scheme for Flexible Chains. *Mol. Phys.*, 75, 1 (Jan 1992), 59-70.

[162] Errington, J. R. and Panagiotopoulos, A. Z. A new intermolecular potential model for the n-alkane homologous series. *J. Phys. Chem. B*, 103, 30 (Jul 29 1999), 6314-6322.

[163] Floriano, M. A., Caponetti, E. and Panagiotopoulos, A. Z. Micellization in model surfactant systems. *Langmuir*, 15, 9 (Apr 27 1999), 3143-3151.

[164] Errington, J. R. and Panagiotopoulos, A. Z. New intermolecular potential models for benzene and cyclohexane. *J. Chem. Phys.*, 111, 21 (Dec 1 1999), 9731-9738.

[165] Macedonia, M. D. and Maginn, E. J. A biased grand canonical Monte Carlo method for simulating adsorption using all-atom and branched united atom models. *Mol. Phys.*, 96, 9 (May 10 1999), 1375-1390.

[166] Mehta, M. and Kofke, D. A. Implementation of the Gibbs Ensemble Using a Thermodynamic Model for One of the Coexisting Phases. *Mol. Phys.*, 79, 1 (May 1993), 39-52.

[167] Panagiotopoulos, A. Z., Quirke, N., Stapleton, M. and Tildesley, D. J. Phase-Equilibria by Simulation in the Gibbs Ensemble - Alternative Derivation, Generalization and Application to Mixture and Membrane Equilibria. *Mol. Phys.*, 63, 4 (Mar 1988), 527-545.

[168] de Pablo, J. J. and Prausnitz, J. M. Phase-Equilibria for Fluid Mixtures from Monte-Carlo Simulation. *Fluid Phase Equilibr.*, 53 (Dec 1989), 177-189.

[169] Guo, Z. and Kindt, J. T. Gibbs ensemble Monte Carlo with solvent repacking: phase coexistence of size–asymmetrical binary Lennard-Jones mixtures. *Mol. Simulat.*, 44, 4 (2018/03/04 2018), 300-308.

[170] Nejahi, Y., Barhaghi, M. S., Mick, J., Jackman, B., Rushaidat, K., Li, Y. Z., Schwiebert, L. and Potoff, J. *GOMC: GPU Optimized Monte Carlo*. City, 2019.

[171] Bai, P. and Siepmann, J. I. Assessment and Optimization of Configurational-Bias Monte Carlo Particle Swap Strategies for Simulations of Water in the Gibbs Ensemble. *J. Chem. Theory Comput.*, 13, 2 (Feb 2017), 431-440.

[172] Torres-Knoop, A., Burtch, N. C., Poursaeidesfahani, A., Balaji, S. P., Kools, R., Smit, F. X., Walton, K. S., Vlugt, T. J. H. and Dubbeldam, D. Optimization of Particle Transfers in the Gibbs Ensemble for Systems with Strong and Directional Interactions Using CBMC, CFCMC, and CB/CFCMC. *J. Phys. Chem. C*, 120, 17 (May 5 2016), 9148-9159.

[173] Shi, W. and Maginn, E. J. Improvement in Molecule Exchange Efficiency in Gibbs Ensemble Monte Carlo: Development and Implementation of the Continuous Fractional Component Move. *J. Comput. Chem.*, 29, 15 (Nov 30 2008), 2520-2530.

[174] Chu, T. C., Chen, R. J. J., Chappelear, P. S. and Kobayashi, R. Vapor-Liquid-Equilibrium of Methane-Normal-Pentane System at Low-Temperatures and High-Pressures. *J. Chem. Eng. Data*, 21, 1 (1976), 41-44.

[175] Wichterle, I. and Kobayashi, R. Vapor-Liquid Equilibrium of Methane-Ethane System at Low-Temperatures and High-Pressures. *J. Chem. Eng. Data*, 17, 1 (1972), 9-+.

[176] Wichterle, I. and Kobayashi, R. Vapor-Liquid Equilibrium of Methane-Propane System at Low-Temperatures and High-Pressures. *J. Chem. Eng. Data*, 17, 1 (1972), 4-+.

[177] Beaudoin, J. M. and Kohn, J. P. Multiphase and Volumetric Equilibria of Methane-N-Decane Binary System at Temperatures between -36 Degrees and 150 Degrees C. *J. Chem. Eng. Data*, 12, 2 (1967), 189-&.

[178] Poston, R. S. and Mcketta, J. J. Vapor-Liquid Equilibrium in Methane-N-Hexane System. *J. Chem. Eng. Data*, 11, 3 (1966), 362-&.

[179] Chang, H. L., Hurt, L. J. and Kobayash.R Vapor-Liquid Equilibria of Light Hydrocarbons at Low Temperatures and High Pressures - Methane-N-Heptane System. *AIChE J.*, 12, 6 (1966), 1212-&.

[180] Kohn, J. P. and Bradish, W. F. Multiphase and Volumetric Equilibria of the Methane-n-Octane System at Temperatures between -110 and 150 C. *J. Chem. Eng. Data*, 9, 1 (1964), 5-8.

[181] Elliot, D. G., Chen, R. J. J., Chappele, P. S. and Kobayash, R. Vapor-Liquid-Equilibrium of Methane-N-Butane System at Low-Temperatures and High-Pressures. *J. Chem. Eng. Data*, 19, 1 (1974), 71-77.

[182] Chen, B. and Siepmann, J. I. Transferable potentials for phase equilibria. 3. Explicit-hydrogen description of normal alkanes. *J. Phys. Chem. B*, 103, 25 (Jun 24 1999), 5370-5379.

[183] McCabe, C., Galindo, A., Gil-Villegas, A. and Jackson, G. Predicting the high-pressure phase equilibria of binary mixtures of perfluoro-n-alkanes plus n-alkanes using the SAFT-VR approach. *J. Phys. Chem. B*, 102, 41 (Oct 8 1998), 8060-8069.

[184] Aparicio, S. Phase equilibria in perfluoroalkane plus alkane binary systems from PC-SAFT equation of state. *J. Supercrit. Fluids*, 46, 1 (Aug 2008), 10-20.

[185] Haley, J. D. and McCabe, C. Predicting the phase behavior of fluorinated organic molecules using the GC-SAFT-VR equation of state. *Fluid Phase Equilibr.*, 440 (May 25 2017), 111-121.

[186] Simons, J. H. and Mausteller, J. W. The Properties of Normal-Butforane and Its Mixtures with Normal-Butane. *J. Chem. Phys.*, 20, 10 (1952), 1516-1519.

[187] Zhang, L. and Siepmann, J. I. Pressure dependence of the vapor-liquid-liquid phase behavior in ternary mixtures consisting of n-alkanes, n-perfluoroalkanes, and carbon dioxide. *J. Phys. Chem. B*, 109, 7 (Feb 24 2005), 2911-2919.

[188] Berendsen, H. J. C., Grigera, J. R. and Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.*, 91, 24 (Nov 19 1987), 6269-6271.

[189] Boulougouris, G. C., Economou, I. G. and Theodorou, D. N. Engineering a molecular model for water phase equilibrium over a wide temperature range. *J. Phys. Chem. B*, 102, 6 (Feb 5 1998), 1029-1035.

[190] Martin, M. G. and Frischknecht, A. L. Using arbitrary trial distributions to improve intramolecular sampling in configurational-bias Monte Carlo. *Mol. Phys.*, 104, 15 (Aug 10 2006), 2439-2456.

[191] Sepehri, A., Loeffler, T. D. and Chen, B. Improving the Efficiency of Configurational-Bias Monte Carlo: A Jacobian-Gaussian Scheme for Generating Bending Angle Trials for Linear and Branched Molecules. *J. Chem. Theory Comput.*, 13, 4 (Apr 2017), 1577-1583.

[192] Laso, M., De Pablo, J. J. and Suter, U. W. Simulation of Phase-Equilibria for Chain Molecules. *J. Chem. Phys.*, 97, 4 (Aug 15 1992), 2817-2819.

[193] Siepmann, J. I., Karaborni, S. and Smit, B. Simulating the Critical-Behavior of Complex Fluids. *Nature*, 365, 6444 (Sep 23 1993), 330-332.

[194] Sepehri, A., Loeffler, T. D. and Chen, B. Improving the Efficiency of Configurational-Bias Monte Carlo: Extension of the Jacobian-Gaussian Scheme to Interior Sections of Cyclic and Polymeric Molecules. *J. Chem. Theory Comput.*, 13, 9 (Sep 2017), 4043-4053.

[195] Sepehri, A., Loeffler, T. D. and Chen, B. Improving the efficiency of configurational-bias Monte Carlo: A density-guided method for generating bending angle trials for linear and branched molecules. *J. Chem. Phys.*, 141, 7 (Aug 21 2014).

[196] Neubauer, B., Boutin, A., Tavitian, B. and Fuchs, A. H. Gibbs ensemble simulations of vapour-liquid phase equilibria of cyclic alkanes. *Mol. Phys.*, 97, 6 (Sep 20 1999), 769-776.

[197] Shah, J. K. and Maginn, E. J. A general and efficient Monte Carlo method for sampling intramolecular degrees of freedom of branched and cyclic molecules. *J. Chem. Phys.*, 135, 13 (Oct 7 2011).

[198] Chen, Z. and Escobedo, F. A. A configurational-bias approach for the simulation of inner sections of linear and cyclic molecules. *J. Chem. Phys.*, 113, 24 (Dec 22 2000), 11382-11392.

[199] Wick, C. D. and Siepmann, J. I. Self-adapting fixed-end-point configurational-bias Monte Carlo method for the regrowth of interior segments of chain molecules with strong intramolecular interactions. *Macromolecules*, 33, 19 (Sep 19 2000), 7207-7218.

[200] Escobedo, F. A. and Depablo, J. J. Monte-Carlo Simulation of the Chemical-Potential of Polymers in an Expanded Ensemble. *J. Chem. Phys.*, 103, 7 (Aug 15 1995), 2703-2710.

[201] Escobedo, F. A. and Martinez-Veracoechea, F. J. Optimized expanded ensembles for simulations involving molecular insertions and deletions. I. Closed systems. *J. Chem. Phys.*, 127, 17 (Nov 7 2007), 174103.

[202] Escobedo, F. A. Optimized expanded ensembles for simulations involving molecular insertions and deletions. II. Open systems. *J. Chem. Phys.*, 127, 17 (Nov 7 2007), 174104.

[203] Martin, M. G. and Siepmann, J. I. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B*, 102, 14 (Apr 2 1998), 2569-2577.

[204] Chen, B., Potoff, J. J. and Siepmann, J. I. Monte Carlo calculations for alcohols and their mixtures with alkanes. Transferable potentials for phase equilibria. 5. United-atom description of primary, secondary, and tertiary alcohols. *J. Phys. Chem. B*, 105, 15 (Apr 19 2001), 3093-3104.

[205] Baumgartner, A. and Binder, K. Monte-Carlo Studies on the Freely Jointed Polymer-Chain with Excluded Volume Interaction. *J. Chem. Phys.*, 71, 6 (1979), 2541-2545.

[206] Pertsin, A. J., Hahn, J. and Grossmann, H. P. Incorporation of Bond-Length Constraints in Monte-Carlo Simulations of Cyclic and Linear-Molecules - Conformational Sampling for Cyclic Alkanes as Test Systems. *J. Comput. Chem.*, 15, 10 (Oct 1994), 1121-1126.

[207] Berthelot, D. Hebd. *Seances Acad. Sci*, 126 (1898), 1703.

[208] Hansch, C. and Fujita, T. Rho-Sigma-Pi Analysis . Method for Correlation of Biological Activity + Chemical Structure. *J. Am. Chem. Soc.*, 86, 8 (1964), 1616-1626.

[209] Leo, A., Hansch, C. and Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.*, 71, 6 (1971), 525-616.

[210] Huddleston, J., Veide, A., Kohler, K., Flanagan, J., Enfors, S. O. and Lyddiatt, A. The Molecular-Basis of Partitioning in Aqueous 2-Phase Systems. *Trends Biotechnol.*, 9, 11 (Nov 1991), 381-388.

[211] Meylan, W., Howard, P., Boethling, R., Aronson, D., Printup, H. and Gouchie, S. Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient. *Environmental Toxicology and Chemistry*, 18, 4 (APR 1999 1999), 664-672.

[212] Braekevelt, E., Tittlemier, S. and Tomy, G. Direct measurement of octanol-water partition coefficients of some environmentally relevant brominated diphenyl ether congeners. *Chemosphere*, 51, 7 (MAY 2003 2003), 563-567.

[213] Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3, 5 (May 1935), 300-313.

[214] Kollman, P. Free-Energy Calculations - Applications to Chemical and Biochemical Phenomena. *Chem. Rev.*, 93, 7 (Nov 1993), 2395-2417.

[215] Shirts, M. R., Pitera, J. W., Swope, W. C. and Pande, V. S. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J. Chem. Phys.*, 119, 11 (Sep 15 2003), 5740-5761.

[216] Aqvist, J. Ion Water Interaction Potentials Derived from Free-Energy Perturbation Simulations. *J. Phys. Chem.*, 94, 21 (Oct 18 1990), 8021-8024.

[217] Debolt, S. E. and Kollman, P. A. Investigation of Structure, Dynamics, and Solvation in 1-Octanol and Its Water-Saturated Solution - Molecular-Dynamics and Free-Energy Perturbation Studies. *J. Am. Chem. Soc.*, 117, 19 (May 17 1995), 5316-5340.

[218] Henin, J. and Chipot, C. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.*, 121, 7 (Aug 15 2004), 2904-2914.

[219] Henin, J., Fiorin, G., Chipot, C. and Klein, M. L. Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. *J. Chem. Theory Comput.*, 6, 1 (Jan 2010), 35-47.

[220] Garrido, N. M., Queimada, A. J., Jorge, M., Macedo, E. A. and Economou, I. G. 1-Octanol/Water Partition Coefficients of n-Alkanes from Molecular Simulations of Absolute Solvation Free Energies. *J. Chem. Theory Comput.*, 5, 9 (Sep 2009), 2436-2446.

[221] Liu, P., Dehez, F., Cai, W. S. and Chipot, C. A Toolkit for the Analysis of Free-Energy Perturbation Calculations. *J. Chem. Theory Comput.*, 8, 8 (Aug 2012), 2606-2616.

[222] Ben-Naim, A. Standard Thermodynamics of Transfer-Uses and Misuses. *J. Phys. Chem.*, 82, 7 (1978 1978), 792-803.

[223] Chen, B. and Siepmann, J. Microscopic structure and solvation in dry and wet octanol. *J. Phys. Chem. B*, 110, 8 (MAR 2 2006 2006), 3555-3563.

[224] Poursaeidesfahani, A., Rahbari, A., Torres-Knoop, A., Dubbeldam, D. and Vlugt, T. J. H. Computation of thermodynamic properties in the continuous fractional component Monte Carlo Gibbs ensemble. *Mol. Simulat.*, 43, 3 (2017), 189-195.

[225] Su, Y., Wania, F., Harner, T. and Lei, Y. Deposition of polybrominated diphenyl ethers, polychlorinated biphenyls, and polycyclic aromatic hydrocarbons to a boreal deciduous forest. *Environ. Sci. Technol.*, 41, 2 (JAN 15 2007 2007), 534-540.

[226] Hippelein, M. and McLachlan, M. Soil/air partitioning of semivolatile organic compounds. 1. Method development and influence of physical-chemical properties. *Environ. Sci. Technol.*, 32, 2 (JAN 15 1998 1998), 310-316.

[227] Goss, K., Buschmann, J. and Schwarzenbach, R. Adsorption of organic vapors to air-dry soils: Model predictions and experimental validation. *Environ. Sci. Technol.*, 38, 13 (JUL 1 2004 2004), 3667-3673.

[228] Harner, T. and Bidleman, T. Octanol-air partition coefficient for describing particle/gas partitioning of aromatic compounds in urban air. *Environ. Sci. Technol.*, 32, 10 (MAY 15 1998 1998), 1494-1502.

[229] Li, J. B., Zhu, T. H., Hawkins, G. D., Winget, P., Liotard, D. A., Cramer, C. J. and Truhlar, D. G. Extension of the platform of applicability of the SM5.42R universal solvation model. *Theor. Chem. Acc.*, 103, 1 (Nov 1999), 9-63.

[230] Bhatnagar, N., Kamath, G., Chelst, I. and Potoff, J. J. Direct calculation of 1-octanol-water partition coefficients from adaptive biasing force molecular dynamics simulations. *J. Chem. Phys.*, 137, 1 (Jul 7 2012).

[231] Bronner, G., Fenner, K. and Goss, K. U. Hexadecane/air partitioning coefficients of multifunctional compounds: Experimental data and modeling. *Fluid Phase Equilibr.*, 299, 2 (Dec 25 2010), 207-215.

[232] Stenzel, A., Endo, S. and Goss, K. U. Measurements and predictions of hexadecane/air partition coefficients for 387 environmentally relevant compounds. *J. Chromatogr. A.*, 1220 (Jan 13 2012), 132-142.

[233] Jorge, M., Garrido, N. M., Simoes, C. J. V., Silva, C. G. and Brito, R. M. M. Predicting Hydrophobic Solvation by Molecular Simulation: 1. Testing United-Atom Alkane Models. *J. Comput. Chem.*, 38, 6 (Mar 2017), 346-358.

[234] Garrido, N. M., Jorge, M., Queimada, A. J., Macedo, E. A. and Economou, I. G. Using molecular simulation to predict solute solvation and partition coefficients in solvents of different polarity. *Phys. Chem. Chem. Phys.*, 13, 20 (2011), 9155-9164.

[235] Kostov, G., Boschet, F. and Ameduri, B. Original fluorinated surfactants potentially non-bioaccumulable. *J. Fluorine Chem.*, 130, 12 (Dec 2009), 1192-1199.

[236] Drame, A., de Givenchy, E. T., Dieng, S. Y., Amigoni, S., Oumar, M., Diouf, A., Darmanin, T. and Guittard, F. One F-Octyl versus Two F-Butyl Chains in Surfactant Aggregation Behavior. *Langmuir*, 29, 48 (Dec 3 2013), 14815-14822.

[237] D'Agostino, L. A. and Mabury, S. A. Identification of Novel Fluorinated Surfactants in Aqueous Film Forming Foams and Commercial Surfactant Concentrates. *Environ. Sci. Technol.*, 48, 1 (Jan 7 2014), 121-129.

[238] Place, B. J. and Field, J. A. Identification of Novel Fluorochemicals in Aqueous Film-Forming Foams Used by the US Military. *Environ. Sci. Technol.*, 46, 13 (Jul 3 2012), 7120-7127.

[239] Stock, N. L., Ellis, D. A., Deleebeeck, L., Muir, D. C. G. and Mabury, S. A. Vapor pressures of the fluorinated telomer alcohols - Limitations of estimation methods. *Environ. Sci. Technol.*, 38, 6 (Mar 15 2004), 1693-1699.

[240] Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F. and Hermans, J. *Intermolecular Forces*. Reidel, Dordrecht, 1981.

[241] Valleau, J. P. and Card, D. N. Monte-Carlo Estimation of Free-Energy by Multistage Sampling. *J. Chem. Phys.*, 57, 12 (1972), 5457-&.

[242] Wu, D. and Kofke, D. A. Phase-space overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J. Chem. Phys.*, 123, 8 (Aug 22 2005).

[243] Wu, D. and Kofke, D. A. Phase-space overlap measures. I. Fail-safe bias detection in free energies calculated by molecular simulation. *J. Chem. Phys.*, 123, 5 (Aug 1 2005).

[244] Pohorille, A., Jarzynski, C. and Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B*, 114, 32 (Aug 19 2010), 10235-10253.

[245] Beveridge, D. L. and Dicapua, F. M. Free-Energy Via Molecular Simulation - Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Bio.*, 18 (1989), 431-492.

[246] Simonson, T. Free-Energy of Particle Insertion - an Exact Analysis of the Origin Singularity for Simple Liquids. *Mol. Phys.*, 80, 2 (Oct 10 1993), 441-447.

[247] Beutler, T. C., Mark, A. E., Vanschaik, R. C., Gerber, P. R. and Vangunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free-Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.*, 222, 6 (Jun 3 1994), 529-539.

[248] Zacharias, M., Straatsma, T. P. and Mccammon, J. A. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.*, 100, 12 (Jun 15 1994), 9025-9031.

[249] Steinbrecher, T., Joung, I. and Case, D. A. Soft-Core Potentials in Thermodynamic Integration: Comparing One- and Two-Step Transformations. *J. Comput. Chem.*, 32, 15 (Nov 30 2011), 3253-3263.

[250] Naden, L. N. and Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 2. Inserting and Deleting Particles with Coulombic Interactions. *J. Chem. Theory Comput.*, 11, 6 (Jun 2015), 2536-2549.

[251] Naden, L. N., Pham, T. T. and Shirts, M. R. Linear Basis Function Approach to Efficient Alchemical Free Energy Calculations. 1. Removal of Uncharged Atomic Sites. *J. Chem. Theory Comput.*, 10, 3 (Mar 2014), 1128-1149.

[252] Anwar, J. and Heyes, D. M. Robust and accurate method for free-energy calculation of charged molecular systems. *J. Chem. Phys.*, 122, 22 (Jun 8 2005).

[253] Pitera, J. W. and van Gunsteren, W. F. One-step perturbation methods for solvation free energies of polar solutes. *J. Phys. Chem. B*, 105, 45 (Nov 15 2001), 11264-11274.

[254] Paliwal, H. and Shirts, M. R. A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. *J. Chem. Theory Comput.*, 7, 12 (Dec 2011), 4115-4134.

[255] Bennett, C. H. Efficient Estimation of Free-Energy Differences from Monte-Carlo Data. *J. Comput. Phys.*, 22, 2 (1976), 245-268.

[256] Shirts, M. R. and Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, 129, 12 (Sep 28 2008).

[257] Dotson D., K. I., Beckstein O. *alchemistry/alchemlyb: Release 0.1.0.* City, 2017.

[258] Klimovich, P. V., Shirts, M. R. and Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput. Aided Mol. Des.*, 29, 5 (May 2015), 397-411.

[259] Straatsma, T. P., Berendsen, H. J. C. and Postma, J. P. M. Free-Energy of Hydrophobic Hydration - a Molecular-Dynamics Study of Noble-Gases in Water. *J. Chem. Phys.*, 85, 11 (Dec 1 1986), 6720-6727.

[260] Flyvbjerg, H. and Petersen, H. G. Error-Estimates on Averages of Correlated Data. *J. Chem. Phys.*, 91, 1 (Jul 1 1989), 461-466.

[261] Chodera, J. D., Swope, W. C., Pitera, J. W., Seok, C. and Dill, K. A. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. *J. Chem. Theory Comput.*, 3, 1 (Jan-Feb 2007), 26-41.

[262] Bannan, C. C., Calabro, G., Kyu, D. Y. and Mobley, D. L. Calculating Partition Coefficients of Small Molecules in Octanol/Water and Cyclohexane/Water. *J. Chem. Theory Comput.*, 12, 8 (Aug 2016), 4015-4024.

[263] Fincham, D. Optimization of the Ewald Sum for Large Systems. *Mol. Simulat.*, 13, 1 (1994), 1-9.

[264] Barker, J. A. and Watts, R. O. Structure of water; A Monte Carlo calculation. *Chem. Phys. Lett.*, 3, 3 (Mar 1969), 144-145.

[265] Levine, B. G., Stone, J. E. and Kohlmeyer, A. Fast analysis of molecular dynamics trajectories with graphics processing units-Radial distribution function histogramming. *J. Comput. Phys.*, 230, 9 (May 1 2011), 3556-3569.

[266] Kelly, C. P., Cramer, C. J. and Truhlar, D. G. SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *J. Chem. Theory Comput.*, 1, 6 (Nov-Dec 2005), 1133-1152.

[267] Curutchet, C., Orozco, M. and Luque, F. J. Solvation in octanol: Parametrization of the continuum MST model. *J. Comput. Chem.*, 22, 11 (Aug 2001), 1180-1193.

[268] Goss, K. U., Bronner, G., Harner, T., Monika, H. and Schmidt, T. C. The partition behavior of fluorotelomer alcohols and olefins. *Environ. Sci. Technol.*, 40, 11 (Jun 1 2006), 3572-3577.

[269] Lei, Y. D., Wania, F., Mathers, D., Mabury, S.A. Determination of Vapor Pressures, Octanol-Air, and Water-Air Partition Coefficients for Polyfluorinated Sulfonamide, Sulfonamidoethanols, and Telomer Alcohols. *J. Chem. Eng. Data*, 49 (2004), 1013-1022.

[270] Wu, Y. X. and Chang, V. W. C. The Effect of Surface Adsorption and Molecular Geometry on the Determination of Henry's Law Constants for Fluorotelomer Alcohols. *J. Chem. Eng. Data*, 56, 8 (Aug 2011), 3442-3448.

[271] Wescott, J. T., Fisher, L. R. and Hanna, S. Use of thermodynamic integration to calculate the hydration free energies of n-alkanes. *J. Chem. Phys.*, 116, 6 (Feb 8 2002), 2361-2369.

[272] Dalvi, V. H. and Rossky, P. J. Molecular origins of fluorocarbon hydrophobicity. *Proc. Natl. Acad. Sci. U.S.A.*, 107, 31 (Aug 3 2010), 13603-13607.

[273] Franks, N. P. and Lieb, W. R. Where do general anaesthetics act? *Nature*, 274, 5669 (1978/07/01 1978), 339-342.

[274] Finkelstein, A. Water and Nonelectrolyte Permeability of Lipid Bilayer Membranes. *J. Gen. Physiol.*, 68, 2 (1976), 127-135.

[275] Cox, S. R. and Williams, D. E. Representation of the Molecular Electrostatic Potential by a Net Atomic Charge Model. *J. Comput. Chem.*, 2, 3 (1981), 304-323.

[276] Breneman, C. M. and Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials - the Need for High Sampling Density in Formamide Conformational-Analysis. *J. Comput. Chem.*, 11, 3 (Apr 1990), 361-373.

ABSTRACT

FORCE FIELD OPTIMIZATION, ADVANCED SAMPLING, AND FREE ENERGY METHODS WITH GPU-OPTIMIZED MONTE CARLO (GOMC) SOFTWARE

by

MOHAMMAD SOROUSH BARHAGHI

December 2019

Advisor: Dr. Jeffrey Potoff

Major: Doctor of Philosophy

Degree: Chemical Engineering

In this work, to address the sampling problem for systems at high densities and low temperatures, a generalized identity exchange algorithm is developed for grand canonical Monte Carlo simulations. The algorithm, referred to as Molecular Exchange Monte Carlo (MEMC), is implemented in the GPU-Optimized Monte Carlo (GOMC) software and may be applied to multicomponent systems of arbitrary molecular topology, and provides significant enhancements in the sampling of phase space over a wide range of compositions and temperatures. Three different approaches are presented for the insertion/deletion of the large molecules, and the pros and cons of each method are discussed. Next, the MEMC method is extended to Gibbs ensemble Monte Carlo (GEMC). The utility of the MEMC method is demonstrated through the calculation of the free energies of transfer of n-alkanes from vapor into liquid 1-octanol, n-hexadecane, and 2,2,4-trimethylpentane, using isobaric-isothermal GEMC simulations.

Alternatively, for system with strong inter-molecular interaction (*e.g.* hydrogen bonds), it's more efficient to calculate the free energies of transfer, using standard thermodynamic integration (TI) and free energy perturbation (FEP) methods. The TI and FEP free energy

calculation methods are implemented in GOMC and utility of these methods are demonstrated by calculating the hydration and solvation free energies of fluorinated 1-octanol, to understand the role of fluorination on the interactions and partitioning of alcohols in aqueous and organic environments.

Additionally, using GOMC, a transferable united-atom (UA) force field, based on Mie potentials, is optimized for alkynes to accurately reproduce experimental phase equilibrium properties. The performance of the optimized Mie potential parameters is assessed for 1-alkynes and 2-alkynes using grand canonical histogram-reweighting Monte Carlo simulations. For each compound, vapor-liquid coexistence curves, vapor pressures, heats of vaporization, critical properties, and normal boiling points are predicted and compared to experiment.

AUTOBIOGRAPHICAL STATEMENT

MOHAMMAD SOROUSH BARHAGHI
5050 Anthony Wayne Dr., Detroit, MI, 48201

Education:

Ph.D.   Chemical Engineering and Materials Science
                    Wayne State University, Detroit, US, 2014 – 2019
                    Advisor: Dr. Jeffrey Potoff

M.S.   Computer Science
                    Wayne State University, Detroit, US, 2017 – 2019
                    Advisor: Dr. Loren Schwiebert

B.S.   Chemical Engineering
                    University of Tehran, Tehran, Iran, 2008 – 2012
                    Advisor: Dr. Omid Tavakoli

Publication:

1. Mohammad Soroush Barhaghi, Chloe Luyet, and Jeffrey J. Potoff, "Effect of Fluorination on the Partitioning of the Alcohols," *J. Mol. Phys.*, (2019).

2. Richard A. Messerly, Mohammad Soroush Barhaghi, Jeffrey J. Potoff, and Michael R. Shirts, "Histogram-free reweighting with grand canonical Monte Carlo: Post-simulation optimization of non-bonded potentials for phase equilibria," *J. Chem. Eng. Data*, (2019) 0021-9568.

3. Younes Nejahi, Mohammad Soroush Barhaghi, Jason Mick, Brock Jackman, Kamel Rushaidat, Yuanzhe Li, Loren Schwiebert, and Jeffrey J. Potoff, "GOMC: GPU Optimized Monte Carlo for the simulation of phase equilibria and physical properties of complex fluids," *J. SoftwareX*, 9 (2019) 20-27.

4. Mohammad Soroush Barhaghi and Jeffrey J. Potoff, "Prediction of phase equilibria and Gibbs free energies of transfer using molecular exchange Monte Carlo in the Gibbs ensemble," *J. Fluid Phase Equilibria*, 486 (2019) 106-118.

5. Mohammad Soroush Barhaghi, Korosh Torabi, Younes Nejahi, Loren Schwiebert, Jeffrey J. Potoff, "Molecular exchange Monte Carlo: A generalized method for identity exchanges in grand canonical Monte Carlo simulations," *J. Chem. Phys.*, 149 (2018) 072318.

6. Mohammad Soroush Barhaghi, Jason R. Mick, and Jeffrey J. Potoff, "Optimized Mie potentials for phase equilibria: application to alkynes," *J. Mol. Phys.*, 115 (2017) 1378-1388.

7. Jason R. Mick, Mohammad Soroush Barhaghi, Brock Jackman, Loren Schwiebert, and Jeffrey J. Potoff, "Optimized Mie Potentials for Phase Equilibria: Application to Branched Alkanes," *J. Chem. Eng. Data*, 62 (2017) 1806-1818.

8. Jason R. Mick, Mohammad Soroush Barhaghi, and Jeffrey J. Potoff, "Prediction of Radon-222 Phase Behavior by Monte Carlo Simulation, " *J. Chem. Eng. Data*, 61 (2016) 1625-1631.

9. Jason R. Mick, Mohammad Soroush Barhaghi, Brock Jackman, Kamel Rushaidat, Loren Schwiebert, and Jeffrey J. Potoff, "Optimized Mie potentials for phase equilibria: Application to noble gases and their mixtures with n-alkanes," *J. Chem. Phys.*, 143 (2015) 114504.