# Development of coarse-grained force fields from a molecular based equation of state for thermodynamic and structural properties of complex fluids

by

Olga Lobanova

A thesis submitted for the Degree of Doctor of Philosophy
and for the Diploma of Imperial College

Department of Chemical Engineering,
Imperial College London
London SW7 2AZ
United Kingdom

August 5, 2014

# Declaration of Originality

I hereby declare that all the material in this dissertation is the result of my own work and that all has been referenced appropriately.

Olga Lobanova

August 2014

# Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

# Abstract

In spite of the vast array of modelling techniques and force fields available, the study of the phase behaviour, structure, microstructure, and dynamics of mixtures remains a challenging task. A systematic coarse-graining (CG) methodology is employed in this thesis involving the parameterisation of force fields using a top-down approach, by effectively describing a large number of target macroscopic thermodynamic states with a rigorous molecular-based equation of state. A recent incarnation of the Statistical Associating Fluid Theory (SAFT-$\gamma$) is used. The underlying force field is based on the Mie intermolecular potential, which is a generalised form of the Lennard-Jones potential with a variable and versatile form of the repulsive and attractive interactions. The coarse-grained force fields developed in this manner are used directly in Molecular Dynamics simulations in order to explore the dynamical, structural, and interfacial properties, which can not be directly accessed by the equation of state (unless a suitable treatment of the inhomogenous properties of the system is made [1]).

The goal of any coarse-graining procedure is to derive simple, but accurate, robust, and transferable force fields. By aiming for the simplicity, the coarse-grained models developed in our work are typically based on the three-to-one mapping, i.e., one bead containing approximately three heavy atoms, or one-to-one mapping for the small spherical molecules, with the polar, directional, and long-ranged interactions between the beads treated implicitly using the effective spherically-symmetric Mie potentials.

The SAFT-$\gamma$ Mie coarse-graining methodology is exemplified for a number of fluid systems of different complexities, including pure component systems, such as: the homologous series of $n-$alkanes, $n-$perfluoroalkanes, semifluorinated alkanes, ethers and water; binary and ternary mixtures, comprising the carbon dioxide, $n-$alkanes, and water; and finally the aqueous mixtures of alkyl polyoxyethylene glycol non-ionic surfactants.

An accurate representation of the vapour-liquid properties with both, the equation of state and molecular simulation, is obtained for the molecules of different size and chemical nature. Describing the properties of water is, however, a much more difficult task. The

CG model suffers from issues associated with the transferability and representability of the various properties for different thermodynamic conditions, as a consequence of the aggressive averaging of the strongly directional and polar forces into an effective spherically symmetrical potential. It has been shown that an isotropic single-site CG model based on a spherically symmetrical potential cannot capture all of the thermodynamic properties of water simultaneously (the issue of representability). Two different CG models of water are proposed: the first is designed to accurately reproduce the saturation liquid density and vapour pressure, and the second to capture the saturation liquid density and surface tension with high precision. Both models benefit from an accurate parameterisation of temperature dependence following the target properties over the entire temperature range of the fluid. An additional model is developed based on the two-to-one mapping, enabling more efficient large scale simulations in, for example, biomolecular systems.

The models of the binary mixtures are developed by using the corresponding pure component models with an additional adjustable parameter to account for the unlike interactions; the latter are obtained by considering appropriate properties of the mixtures such as the fluid-phase equilibria or the thermodynamic properties of mixing. The unlike interactions are shown to be transferable for a quantitative description of the phase behaviour over a wide range of conditions and for the systems of related components. We are able to obtain an accurate prediction of the azeotropic point, critical loci, tree phase line, global density, and the shape of phase envelopes for studied mixtures. The quality of predictions is found comparable to the results from the atomistic models and other equations of state.

The aqueous mixtures of alkyl polyoxyethylene glycol non-ionic surfactants are a key final goal of the research presented in this thesis. The CG models of the surfactants are developed within the SAFT-$\gamma$ group-contribution framework, where each functional group is derived from an accurate representation of the corresponding chemical moiety. By capturing a delicate interplay of the repulsive and attractive intermolecular interactions and obtaining the right balance between energetic and entropic effects, the various phase morphologies at ambient conditions can be reproduced in agreement with the experimental findings over the entire concentration range. The force fields developed in the current work allow for a prediction of key structural and interfacial properties. The Molecular Dynamics simulations reveal the spontaneous formation of micelles at low surfactant concentrations and a self-assembly into a bilayer at high surfactant concentrations. The aggregation

numbers, the critical micelle concentration, area per molecule, the surface excess properties, and bilayer thickness are found in very good agreement with experimental data. This is very encouraging considering that only macroscopic thermophysical properties are used to develop the underlying force fields that describe the fine interactions between the molecules in the system.

Despite the simplicity, coarse-grained force fields are shown to be robust and transferable; they can be applied to predict the properties which were not used in the original parameterisation procedure, with an accuracy comparable to the more sophisticated and computationally demanding models.

# Acknowledgements

This work would not have been possible without many people who I would like to thank here. Foremost, I am extremely grateful to my both supervisors, Professor George Jackson and Professor Erich Müller, not only for their invaluable scientific guidance, but also for their extraordinarily caring attitude, their commitment, patience, and inspiration. Their support and encouragement helped me greatly through these years and motivated me a lot, even at times when nothing seemed to work.

I would also like to thank my examiners, Professor Amparo Galindo and Professor Alejandro Gil-Villegas for making the viva such an enjoyable experience, as I would never have thought this was possible.

A crucial contribution to this work was made by people, who patiently answered all of my questions and shared their knowledge and experience with me, in particular Simon, Peter, Carmelo, Erik, Andrés, and Carlos. I have been privileged to be a part of the MSE group and I would like to thank all my fellow colleges for making it such a nice and pleasant environment. Especially I would like to thank Zara, Simon, Eirini, Manolis, Panos, Charles, Lara, Nina, Roochi, Hendrik, Xenia, and Jens for all the fun times together and our gym team, Sadia, Christianna, Katerina, and Suela, for motivating each other to stay fit and healthy.

I am particularly thankful to Ruddy, who was there with me and for me all these years, encouraged and cheered me up when I needed it most. Of course, nothing would have been possible without endless love and support of my parents and grandparents, who would always be there for me, worry about me, believe in me, and tell me that I can do it.

# Contents

**Abstract**

**Contents** **i**

**List of Figures** **v**

**List of Tables** **xi**

**1 Introduction** **1**

1.1 Motivation and objective . . . . . 1

1.2 Thesis overview . . . . . 3

**2 Force fields, molecular theory and simulation** **5**

2.1 Intermolecular forces and potentials . . . . . 5

2.1.1 Intermolecular forces . . . . . 5

2.1.2 Pair potentials . . . . . 8

2.2 Statistical mechanics . . . . . 13

2.2.1 Equation of state . . . . . 17

2.3 SAFT-$\gamma$ Mie equation of state . . . . . 21

2.3.1 Short history of SAFT . . . . . 21

2.3.2 SAFT-$\gamma$ Mie model and theory . . . . . 23

2.3.3 Recent developments of the SAFT-$\gamma$ Mie coarse-grained force fields . 27

2.4 Molecular simulation . . . . . 28

2.4.1 Monte Carlo . . . . . 29

2.4.2 Molecular Dynamics . . . . . 29

2.5 Force field parameterisation . . . . . 30

2.6 Coarse-graining . . . 32
2.6.1 Computer simulations at different scales . . . 32
2.6.2 Philosophy of coarse-graining . . . 35
2.6.3 Bottom-up methods . . . 38
2.6.4 Top-down methods . . . 44

**3 Chain molecules** **52**
3.1 Chapter overview . . . 52
3.2 Introduction . . . 52
3.3 Methods and models . . . 54
3.3.1 SAFT-$\gamma$ Mie force field development . . . 54
3.3.2 Intramolecular potentials . . . 54
3.3.3 Simulation details . . . 56
3.3.4 Scaling laws and critical data . . . 57
3.4 Results . . . 58
3.4.1 Alkanes . . . 58
3.4.2 Perfluoroalkanes . . . 61
3.4.3 Ethers . . . 64
3.4.4 Triethylene glycol . . . 65
3.4.5 Heteronuclear models . . . 65
3.5 Conclusions . . . 72

**4 Water** **74**
4.1 Chapter overview . . . 74
4.2 Introduction . . . 74
4.3 Methods . . . 81
4.3.1 Molecular simulation details . . . 81
4.3.2 Interfacial tension by the mechanical and test-area methods . . . 81
4.4 Results and discussion . . . 83
4.4.1 The effect of the cutoff of the potential on the thermodynamic properties . . . 83
4.4.2 Transferability and representability issues . . . 87

4.4.3 Model 1 with the liquid density and vapour pressure as target properties . . . 94
4.4.4 Model 2 with the saturated liquid density and vapour-liquid interfacial tension as target properties . . . 98
4.4.5 Different levels of coarse-graining of water . . . 100
4.4.6 Comparison of the Mie (8-6) CG models of water with existing models 104
4.5 Conclusions . . . 109

**5 Binary and ternary mixtures 111**
5.1 Chapter overview . . . 111
5.2 Phase behaviour of binary mixtures . . . 111
5.3 Introduction to the systems . . . 114
5.4 Methods and models . . . 116
5.4.1 SAFT-$\gamma$ force field . . . 116
5.4.2 Simulation details . . . 117
5.5 Results and discussion . . . 118
5.5.1 $CO_2$+$H_2O$ mixture . . . 118
5.5.2 $n-$alkane + $H_2O$ mixtures . . . 127
5.5.3 $n-$alkane + $CO_2$ mixtures . . . 133
5.5.4 Ternary $n-$decane +$CO_2$+$H_2O$ mixture . . . 139
5.6 Conclusions . . . 143

**6 Aqueous non-ionic surfactant mixtures 145**
6.1 Chapter overview . . . 145
6.2 Introduction . . . 146
6.3 Methods and models . . . 150
6.3.1 Simulation details . . . 150
6.3.2 SAFT-$\gamma$ molecular CG force field development . . . 150
6.3.3 Unlike interactions between water and alkanes . . . 151
6.3.4 Unlike interactions between water and polar groups . . . 152
6.4 Results and discussion . . . 155
6.4.1 Morphology of the micelles as a function of the surfactant concentration . . . 155

6.4.2 Surfactant chain length effects . . . 159
6.4.3 Equilibrium surface tension and surface excess . . . 165
6.4.4 Micelle kinetics . . . 166
6.4.5 Isotropic-lamellar phase transition . . . 170
6.5 Conclusions . . . 174

**7 Concluding remarks** **175**
7.1 Future work and potential applications . . . 177

**Appendix** **184**

**A SAFT-$\gamma$ Mie equation of state** **184**
A.1 Ideal contribution . . . 184
A.2 Monomer contribution . . . 185
A.2.1 Repulsive reference fluid . . . 186
A.2.2 Mean-attractive energy . . . 186
A.2.3 Fluctuation term . . . 189
A.2.4 The third-order term . . . 190
A.3 Chain contribution . . . 191

**B Integrating the equations of motion** **194**

**C Force Fields developed in this work** **196**

**D Representation of the VLE properties of the chain molecules by means of the SAFT-$\gamma$ Mie EoS** **200**

**E Representation of properties of water by means of the molecular simulations** **205**

**F Representation of properties of binary mixtures comprising water, carbon dioxide, and $n-$alkanes** **207**

**G Aggregation number distributions of the low concentration non-ionic surfactant solutions** **210**

# List of Figures


1.1 Connection between the experiment, theory and molecular simulations . . . 4

2.1 Hard-sphere, square-well, Sutherland, and Lennard-Jones intermolecular potentials . . . 10
2.2 Intermolecular potentials of the Mie family . . . 13
2.3 The excluded volume per pair of hard spheres $b = (2/3)\pi\sigma^3$ . . . 19
2.4 Schematic representation of different perturbation contributions within the SAFT framework . . . 22
2.5 Computer simulations at different levels of time and length scale . . . 32
2.6 Mapping of the all-atom configuration space into the coarse-grained configuration space . . . 36

3.1 Schematic representation of the MD simulation of the vapour-liquid equilibrium properties in the canonical ensemble . . . 57
3.2 Saturation density, vapour pressure and surface tension as a function of temperature for $n$-alkanes ($C_6H_{14}$, $C_9H_{20}$, $C_{12}H_{26}$, $C_{15}H_{32}$). . . . 60
3.3 Saturation density, vapour pressure and surface tension as a function of temperature for $n$-perfluroalkanes ($C_6F_{14}$, $C_8F_{18}$, $C_{10}F_{22}$) . . . 63
3.4 Liquid-liquid and vapour-liquid equilibria of $C_6F_{14} + C_6H_{14}$ mixture . . . 66
3.5 Atmospheric liquid density, vapour pressure and surface tension as a function of temperature for semi-fluorinated alkanes ($F_4H_5$, $F_4H_8$, $F_6H_8$, $F_8H_8$ . 69
3.6 CG molecular model for $C_{10}E_4$ surfactant . . . 70
3.7 Liquid-liquid equilibrium of $C_7E_5$ surfactant in $n$-dodecane . . . 71

4.1 CG molecular model for water, based on the one-to-one and two-to-one level of coarse-graining. . . . 80

4.2 The vapour-liquid coexistence properties including the saturated liquid and vapour densities, and vapour pressure for Model 1 of water at $T$=393 K as a function of the cutoff radius . . . . . 85
4.3 Vapour-liquid interfacial tension of Model 2 of water at T=293 K as a function of the cutoff radius . . . . . 86
4.4 The saturated liquid density and vapour pressure of water, predicted with a universal parameter set . . . . . 88
4.5 Phase envelops as a function of van der Waals constant $\alpha$ including the fluid and the solid phases . . . . . 90
4.6 The ratio of the critical and triple point temperatures as a function of the van der Waals constant $\alpha$ corresponding to different Mie fluids . . . . . 90
4.7 The various CG intermolecular potentials used to represent water at 298 K 93
4.8 The temperature dependence of the size and energy parameters $\sigma$ and $\epsilon$ for Mie (8-6) Model 1 of water . . . . . 95
4.9 The temperature-density coexistence envelope in dimensionless units $T^* = k_BT/\epsilon$ and $rho^* = \rho\sigma^3$ for the Mie (8-6) system . . . . . 96
4.10 The saturated liquid density and vapour pressure of water as represented with Mie (8-6) Model 1. . . . . 97
4.11 The enthalpy of vaporisation and vapour-liquid interfacial tension of water as represented with Mie (8-6) Model 1. . . . . 98
4.12 The temperature dependence of the size and energy parameters $\sigma$ and $\epsilon$ parameters of our Model 2 . . . . . 100
4.13 The saturated liquid density and surface tension of water as predicted with Model 2. . . . . 101
4.14 The enthalpy of vaporisation and vapour pressure of water as predicted with Model 2. . . . . 102
4.15 Comparison of the CPU time per time step in typical Molecular Dynamics simulations as a function of number of molecules for Model 2 and the atomistic SPC water model . . . . . 105
4.16 The self-diffusion coefficient as a function the inverse temperature . . . . . 107

5.1 $PT$-projections illustrating the main six types of the critical phase behaviour according to the classification by Scott and van Konynenburg . . . 112
5.2 Isothermal pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at $T$ =383 and 473 K . . . . . 119
5.3 Three-dimensional ($PTx$) phase diagram for carbon dioxide+water system . 120
5.4 $PT$ projection of the phase diagram for the carbon dioxide+water binary mixture. The inset shows the enlarged region close to the critical point of pure $CO_2$. . . . . 121
5.5 Isothermal pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at $T = 298$ K . . . . . 122
5.6 Total density profile of carbon dioxide +water mixture at the three phase point at $T = 298.15$ K and $P = 6.82$ MPa via MD simulation . . . . . 123
5.7 Density profiles at the interphase of the carbon dioxide+water mixture at $T = 383$ and 423 K at pressures $P$ =8.18, 31.8, 62.30, and 102.10 MPa . . . 124
5.8 Pressure-density ($P\rho$) diagram for carbon-dioxide+water mixture at isotherms corresponding to $T$ =298, 304, 383, 423, and 473 K . . . . . 124
5.9 Isothermal pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at $T = 423$ and 548 K. Comparison between the experiment, SAFT-$\gamma$ Mie EoS, CG simulations, and the atomistic simulations . . . . . 126
5.10 Isothermal pressure-composition ($Px$) phase diagram for water+ $n-$hexane mixture at $T = 473.2$ K . . . . . 128
5.11 Total density profile of $n-$hexane+water mixture at the three phase point (heteroazeotrope) at $T = 473.2$ K and $P = 3.33$ MPa obtained via MD simulation . . . . . 129
5.12 Three-dimensional ($PTx$) phase diagram for water+ $n-$hexane system . . . 130
5.13 $PT$ projection of the phase diagram for water+ $n-$hexane binary mixture 130
5.14 Density profiles at the interphase of the water+$n-$hexane mixture at $T =$ 473 K at $P$ =2.16, 2.87, and 6.25 MPa . . . . . 131
5.15 Pressure-density ($P\rho$) diagram for water+$n-$hexane mixture at isotherms corresponding to $T$ =473 K . . . . . 132
5.16 Isothermal pressure-composition ($Px$) phase diagram for water+ $n-$dodecane mixture at $T = 603$ K . . . . . 133

5.17 Three-dimensional ($PTx$) phase diagram for water+ $n-$dodecane system . 133
5.18 $PT$ projection of the phase diagram for water + $n-$dodecane binary mixture 134
5.19 Pressure-composition ($Px$) phase diagram for mixtures of carbon dioxide with $n-$hexane and $n-$pentadecane at isotherms corresponding to $T$ =313, 353, 393 K and to $T$ =292, 353, and 393 K . . . . . 135
5.20 Three-dimensional ($PTx$) phase diagram for carbon dioxide+ $n-$hexane and carbon dioxide+ $n-$pentadecane system . . . . . 137
5.21 $PT$ projection of the phase diagram for the carbon dioxide+$n-$hexane and carbon dioxide+ $n-$pentadecane binary mixture . . . . . 138
5.22 Density profiles at the interface of the carbon dioxide+ $n-$pentadecane mixture at $T = 292$ K and 353 K and different pressures . . . . . 139
5.23 Pressure-density ($P\rho$) diagram for carbon dioxide+$n-$pentadecane mixture the isotherms $T$ =292, 313, 353, and 393 K . . . . . 140
5.24 Phase composition diagram for the $n-$decane+carbon dioxide+water system at $T = 393.14K$ and $P$ =6.02, 7.58, 9.59, and 11.81 MPa . . . . . 142
5.25 Total density profile of the ternary $n-$decane+carbon dioxide+water mixture at $T = 393.14$ K and $P = 6.02$ MPa obtained via MD simulation . . . 143

6.1 The excess enthalpy of the triethylene glycol-water and $C_8H_{18}O_4$-water mixture at $T = 298.15K$ and $P = 1MPa$ . . . . . 153
6.2 Comparison of the van der Waals energy values of Martini, Klein, and CG models developed in this work. . . . . 154
6.3 Simulated $C_{10}E_4$ micelle formation. The snapshots are at the beginning of the process (0 ns), after 0.6 ns, after 1.7 ns, and after about 14 ns . . . . . 155
6.4 Snapshots of the monodisperse spherical micelles and polydisperse spherical and rod-like micelles at different concentrations . . . . . 157
6.5 Relative shape anisotropy $\kappa^2$ and radius of gyration as functions of the aggregation number of the $C_{10}E_4$ micelles, summarised over different concentrations and averaged over time. . . . . 158
6.6 Density profiles for $OA$, $EO$ and $CM$ beads within a spherical $C_{10}E_4$ micelle 159

6.7 Probability distribution of end-to-end distances for $C_iE_j$ surfactants of different alkyl chain length, represented by $C_iE_4$ and of different ethoxy chain length, represented by $C_{10}E_j$ surfactants. . . . . 160
6.8 Average aggregation numbers as a function of the alkyl and ethoxy groups . 162
6.9 Radius of gyration as a function of alkyl and ethoxy chain length of various $C_iE_j$ calculated by MD simulation. . . . . 163
6.10 Critical micelle concentration as a function of the alkyl chain length within the $C_iE_4$ surfactant family, and as a function of the ethoxy chain length within $C_{10}E_i$ . . . . . 164
6.11 Surface tension as a function of the surface excess concentration for $C_{10}E_4$ . 166
6.12 Evolution in time of the self-assembly of $C_{10}E_4$ surfactants . . . . . 169
6.13 Mobility in the micellar solution of $C_{10}E_4$ in dependence of the chain length 170
6.14 Simulated bilayer formation at the surfactant concentration $C_{10}E_4$= 0.74 at $t$ =1, 20, 50, 150, and 190 ns . . . . . 171
6.15 Orientational order parameter $S_2$ as a function of concentration in an aqueous $C_{10}E_4$ system . . . . . 172
6.16 Beads distribution along the director $g_{||}(r_{||})$ for $C_{10}E_4$-water system at $w_{C_{10}E_4} = 0.74$ . . . . . 173
6.17 Experimental phase diagram for $C_{10}E_4$ and the MD simulation snapshots at different concentrations, indicating the spherical micelles and rod-like micelles in the micellar phase, lamellar phase, and isotropic surfactant rich phase . . . . . 173

D.1 Saturation densities and pressures for homologous series of alkanes non-divisible by three ($C_5H_{12}$-$C_{14}H_{30}$) . . . . . 201
D.2 Saturation densities and pressures for ethers ($C_4H_{10}O_2$ to $C_{10}H_{26}O_5$) . . . . 202
D.3 Saturation density and pressure for triethyleneglycol . . . . . 203
D.4 Liquid densities and vapour pressures of surfactants of different chain lengths 204

F.1 Pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at isotherms corresponding to $T$ =383, 423, 473, 523, 548, 573, and 598 K . . . . . 207

F.2 Pressure-composition ($Px$) phase diagram for mixtures of carbon dioxide with $n-$nonane, $n-$dodecane, and $n-$octadecane at different isotherms . . 208
F.3 Interfacial tension as a function of pressure in carbon dioxide+ $n-$dodecane system at $T = 403$ and 443 K . . . . . 209

G.1 Aggregation number distribution at the surfactant concentrations $x = 3.5 \times 10^{-3}$ for different alkyl and ethoxy chain length . . . . . 211

# List of Tables


3.1 Bead name and atomistic definition . . . . . 54
3.3 CG intramolecular parameters obtained from AA simulations. . . . . 56
3.4 Intermolecular interaction parameters ($\sigma$ and $\epsilon$) and corresponding deviations in saturation liquid density and vapour pressure for homologous series of $n-$alkanes based on the Mie (15-6) potential . . . . . 61
3.5 Intermolecular interaction parameters ($\sigma$ and $\epsilon$) and corresponding deviations in saturation liquid density and vapour pressure for homologous series of $n-$perfluoroalkanes based on Mie (13-6) potential . . . . . 62
3.6 Intermolecular interaction parameters ($\sigma$ and $\epsilon$) and corresponding deviations in the target properties for the homologous series of ethers based on the Mie (19-6) potential. . . . . 64
3.7 Unlike interaction parameters between the CG groups of semi-fluorinated alkanes. . . . . 68
3.8 Unlike interaction parameters between the CG groups of $C_iE_j$ surfactants. 71
3.2 CG representation for chain molecules . . . . . 73

4.1 Mie (8-6) models of water (with a cutoff of $R_c = 20$ Å) at different molecule-to-bead levels of coarse-graining developed to reproduce the saturated liquid density and vapour-liquid interfacial tension at $T = 298$ K with the corresponding melting temperatures. . . . . 103
4.2 Biowater model parameterised to reproduce the saturated liquid density and vapour-liquid interfacial tension at ambient temperatures . . . . . 104
4.3 Parameter sets for the atomistic and coarse-grained water models . . . . . 105
4.4 Comparison of the CG and atomistic water models . . . . . 108

5.1 Barotropic density inversion points for the carbon dioxide +water mixture . 123
5.2 Comparison between the experimental equilibrium phase compositions and those predicted by the SAFT-$\gamma$ Mie EoS and MD simulations of the $n-$decane+carbon dioxide+water system at $T = 393.15$ K and different pressures. . . . . 141

6.1 Comparison of the experimental interfacial tensions at $T = 298.15$ K with those calculated by MD simulation using $k_{ij} = 0.3$ for $n$-hexane, $n$-nonane, $n$-dodecane . . . . . 151
6.2 Number of surfactant and water molecules with corresponding molar and weight fraction in the modelled systems. . . . . 157

A.1 Coefficients $\phi_{m,n}$ . . . . . 191

C.1 Experimental data used in the parameter estimation procedure. . . . . 197
C.2 CG intermolecular interaction parameters obtained with SAFT-$\gamma$ Mie EoS . 198
C.3 Parameter set for Model 1 which reproduces saturation liquid density and vapour pressure . . . . . 199
C.4 Parameter set for Model 2 which reproduces liquid density and surface tension 199

E.1 The vapour-liquid coexistence properties for Model 1 of water at $T$=393 K as a function of the cutoff radius . . . . . 205
E.2 The vapour-liquid coexistence properties for the Mie (8-6) Model 1 of water at different state points obtained via MD simulations . . . . . 206
E.3 The vapour-liquid coexistence properties for the Mie (8-6) Model 2 of water at different state points obtained via MD simulations . . . . . 206

# Chapter 1

# Introduction

## 1.1 Motivation and objective

Understanding phase behaviour is essential for a reliable and optimal design of chemical processes. Since experiments are time and cost intensive, great saving can be achieved by using molecular modelling approaches. Even if modelling can never completely replace experiments, it can provide guidance, thereby reducing the number of experiments that are required. Additionally to providing insight and a fundamental understanding, modelling can be used to determine physical properties where experiments are difficult or unfeasible, as for example when treating toxic substances or for systems at extreme temperatures and pressures [2]. The diversity of compounds encountered in everyday life represents a challenge and an extra level of complexity, therefore reliable and rigorous models that are capable of dealing with the thermophysical properties of multicomponent systems under various conditions are of great importance for industrial applications. Although an equation of state is often a preferable tool in the chemical and petrochemical community due to its efficiency and the corresponding ease of computation, molecular simulation has emerged as invaluable theoretical method over the last few decades [3, 4]. Molecular simulation provides an insight at the molecular level and allows one to study properties that cannot by accessed via an equation of state, such as interfacial and transport properties or the properties of confined systems.

Remarkable advances in molecular modelling have been made in recent years, both in terms of more accurate theories of fluids and mixtures and molecular simulations. One of the important reasons for the great progress in fluid theories lies in the possibility to

test the theories by direct comparison with molecular simulation. Improvements in molecular simulations are consequences of the advent of more accurate force fields, more efficient algorithms, and of course the increase in computational power according to Moore's law [5], which suggests that computational performance doubles approximately every two years. The most recent record for the largest molecular simulation is currently four trillion Lennard-Jones particles ($4.125\times10^{12}$) performed on 146016 cores [6]. Nowadays, molecular simulations have advanced so much that one refers to them not only as computer experiments but as computational microscopy. Notwithstanding, this striking increase in computer power, many processes in fluids, soft matter, mesoscale, and biological systems are still beyond the capacities of conventional atomistic simulations. For example, in order to simulate a single bacterial cell with a diameter of 0.5 $\mu$m, a system of $\sim 10^9$ atoms is required, while an eukaryotic cell with a diameter of 10 $\mu$m comprises $\sim 10^{14}$ atoms. The relevant cellular processes embrace timescales from microseconds to seconds [7]. It is therefore crucial to develop reliable simulations techniques that can access these large size and length scales.

One step towards increasing the efficiency of computational modelling is the application of the so-called coarse-graining methodology instead of the commonly used atomistic level of description, which can be two to three orders of magnitude apart in terms of computational effort. However, the development of reliable and robust coarse-grained models remains a challenge. A common "bottom-up" approach to derive coarse-grained force fields is to start with a model at a high level of resolution and to integrate out the unwanted degrees of freedom, by applying an iterative heuristic procedure. Apart from being time consuming, this technique can suffer from the lack of transferability and robustness [8], implying that the force fields might not provide reliable predictions at conditions different from those at which the models were derived. The objective of the work presented in this thesis is to develop a modelling framework, that allows for the development of coarse-grained force fields, which are robust and transferable and allow for accurate predictions of various thermophysical properties over a wide range of conditions. For this purpose, we employ a molecular-based equation of state (SAFT-$\gamma$ Mie) which provides a direct link between the microscopic and macroscopic properties. Our "top-down" coarse-graining technique is illustrated for several systems that are of significant industrial relevance, such as mixtures comprising water, carbon dioxide, hydrocarbons, and non-ionic surfactants.

Water is perhaps the most important liquid in our world and is the most common solvent in industrial and biological systems. Despite being one of the most studied of fluids, there currently is no model of water that can accurately reproduce all of its properties and anomalies in a simultaneous manner. The difficulty in water modelling arises from highly polar nature of the molecule and the specific hydrogen bonding interactions that are active in the system.

Mixtures of water, carbon dioxide, and $n-$alkanes have received particular interest in recent years due to their central role in the petrochemical industry. The range of applications spans enhanced oil recovery (EOR), $CO_2$ capture and storage, natural gas transport and processing, to mention but a few. The most prevalent modelling tool in reservoir engineering remains the cubic equations of state such as the Peng-Robinson [9] or Soave-Redlich-Kwong [10] equations. However, an accurate description of petrochemical mixtures is challenging due to disparity of interactions and molecular sizes between various components so that the use of the cubic equation of state is often not the most appropriate choice for this purpose.

Surfactant systems embrace a wide range of applications from textile manufacturing, emulsification, and drug delivery to electronic printing, biotechnology, and microelectronics [11]. An ideal tool to describe the complex structures and mesophase behaviour of surfactant systems is by means of direct molecular simulation. Conventional atomistic simulations encounter the limitations of the time and length scale if one tries to describe the slow processes such as self-assembly in diluted systems.

The challenge arising in this context is: first, to treat a wide variety of complex fluids within a single framework; and second, to establish a direct link that connects experiment, theory, and simulation (see Figure 1.1). By extracting force field parameters directly from macroscopic experimental data, the modelling of fluid can be performed by the most suitable and efficient technique, be it theory or simulation depending on properties of interest, thereby offering a tool for multi-scale modelling.

## 1.2 Thesis overview

An overview over the relevant background is provided in Chapter 2, embracing the mathematical functions of intermolecular potentials between the atoms and molecules and the

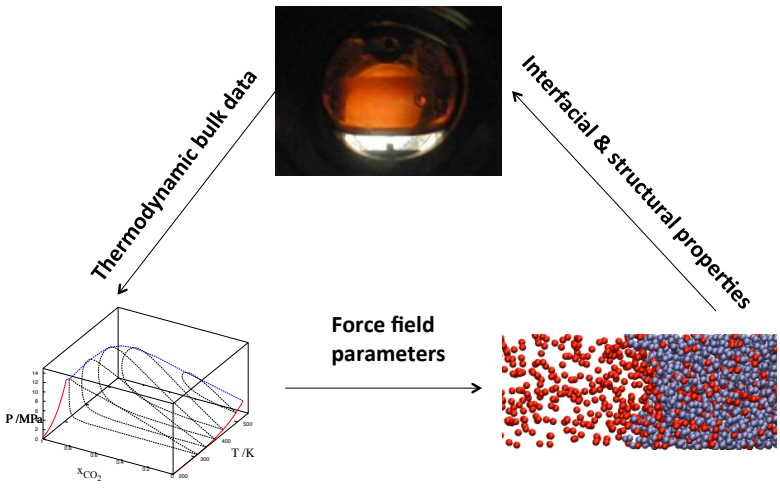



**Figure 1.1:** Connection between the experiment, theory and molecular simulations. Example of the carbon dioxide+$n-$alkane system, when treated by all three techniques.

modelling techniques that allow for a calculation of the macroscopic properties from the intermolecular potentials, by using either an equation of state or direct molecular simulation. The advantages of coarse-grained simulations over the atomistic simulations are then highlighted, along with the currently available coarse-graining approaches. Since the focus of this work is to derive coarse-grained force fields by using the molecular-based SAFT-$\gamma$ Mie equation of state, its current development is explained in detail together with a number of the most recent applications. The following chapters contain an illustration of this technique for various fluid systems: the development of suitable force fields for chain molecules, such as $n-$alkanes, $n-$perfluoroalkanes and $n-$ethers is presented in Chapter 3. Several models for water are proposed in Chapter 4, where we discuss the challenges of the modelling technique and the resulting virtues and defects of the models. After the development of the models for pure components, we turn our attention to the behaviour of mixtures: in Chapter 5 we study the binary and ternary mixtures, comprising water, carbon dioxide, and $n$-alkanes. Finally, in Chapter 6, the most complex of the systems studied are presented: aqueous non-ionic surfactant solutions. We investigate the structural, interfacial, and dynamical properties of surfactants from the alkyl polyoxyethylene glycol family with different architectures.

# Chapter 2

# Force fields, molecular theory and simulation

## 2.1 Intermolecular forces and potentials

### 2.1.1 Intermolecular forces

The macroscopically observed thermophysical properties of a system are determined by the intermolecular forces between its constituent molecules [12]. Already the ancient Greeks defined two fundamental forces: love, which brings things together, and hate, which moves things apart. This idea postulated the main concept of the chemical theory for two thousand years. In the seventeenth century, Newton suggested that the forces must be attractive in their nature and vary as an inverse power of the distance. The period between the Newton's *Principia* and the nineteenth century was marked by the contradiction and controversy between the scientists. At the beginning of the nineteenth century, Young and Laplace were convinced that one needs to take into account not only the attractive but also the repulsive forces in order to explain the cohesion and capillarity in fluids. Half a century later a concept of molecules repelling each other at short distances and attracting each other at intermediate distances was widely accepted by the pioneers of the molecular theories: Clausius, Maxwell, van der Waals, and Boltzmann [13]. The concept of repulsion at the short range was utilised to characterise the solid and fluid structures, whereas the concept of attraction was exploited to explain, for example, the capacity of gases to condense into liquids. The mathematical description of the dispersion

forces remained, however, highly debated and retained an empirical character until the beginning of the twentieth century. Only with the birth of quantum theory in 1920s and the description of the electronic structure of atoms, the origin of forces could be described explicitly. The repulsive forces were attributed to the exclusion of electron clouds according to the Pauli principle, while the attractive forces were ascribed to the correlations between electron clouds surrounding atoms.

The force and the interaction potential function are related by

$$F = -du(r)/dr. \tag{2.1}$$

A positive slope of the potential function represents an attractive force and a negative slope indicates the repulsive force. It is of interest to be able to represent the forces (or the intermolecular potentials) in a closed mathematical form, as then it becomes plausible to infer and model the macroscopic behaviour of a given fluid. It has become common practice to categorise the forces according to their general nature. Some of the commonly used potential functions of intermolecular forces are summarised briefly in this section. For an abridged description the reader is referred to the standard texts available on the subject [13–15] .

Some forces are purely electrostatic in origin and include interactions among charges, ions, permanent dipoles, quadrupole and higher multipoles. The simplest to describe quantitatively is the potential energy $u(r)$ between two point electric charges which is given by the Coulomb interaction

$$u(r) = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r}\frac{1}{r}, \tag{2.2}$$

where $\varepsilon_0$ is the dielectric permitivity of vacuum ($8.854 \times 10^{-12} C^2$ J/m), $\varepsilon_r$ is the dielectric constant, $q_i$ is the electric charge, given as a product of the ionic valence $z_i$ and the unit of charge $e$ ($1.60218 \times 10^{-19}$ C), ($q_i = z_i e$), $r$ is the centre-to-centre distance between the charges. Between two dipoles, the interactions depend not only on the distance but also on the orientation [16]:

$$u(r) = -\mu_i\mu_j\frac{2\cos\theta_i\cos\theta_j - \sin\theta_i\theta_j\cos\phi_{ij}}{4\pi\varepsilon_0}\frac{1}{r^3}, \tag{2.3}$$

where $\mu$ is the dipole moment of a particle, $\theta_i$ is a polar orientation of the dipole and $\phi_{ij}$ is the azimuthal angle between two dipoles. If the interaction energy is smaller than $k_B T$,

dipoles can rotate freely so that the dipole-dipole interactions can be averaged over orientations. The angle averaging can be derived from classical electrodynamics and classical statistical physics by implementing the Boltzmann factor:

$$\exp\left(-\frac{u(r)}{k_BT}\right) = \frac{\int \exp\left(-\frac{u(r,\theta,\phi)}{k_BT}\right)\sin\theta \mathrm{d}\theta\mathrm{d}\phi}{\int \sin\theta\mathrm{d}\theta\mathrm{d}\phi} = \left\langle \exp\left(-\frac{u(r,\theta,\phi)}{k_BT}\right)\right\rangle, \tag{2.4}$$

where $k_B$ is the Boltzmann constant ($1.38 \times 10^{-23}$ J/K) and $T$ is the temperature. The Boltzmann-averaged (angle-averaged) interactions between the dipoles are referred to as the Keesom interactions [13] and are given by

$$u(r) = -\frac{2}{3}\frac{\mu_i^2\mu_j^2}{k_BT(4\pi\varepsilon_0)^2}\frac{1}{r^6}. \tag{2.5}$$

Similarly, an angle-averaged quadrupole-quadrupole interactions are described by

$$u(r) = -\frac{7}{40}\frac{Q_i^2Q_j^2}{k_BT(4\pi\varepsilon_0)^2}\frac{1}{r^{10}}, \tag{2.6}$$

where $Q_i$ is the quadrupole moment of particle $i$. Electrostatic forces include polarisation or induction forces, which arise from the polarisation or tendency to distort an electric charge distribution in an electric field of one molecule caused by the charge distribution on another molecule.

The induction by a permanent dipole is given by Debye relation

$$u(r) = -\frac{\alpha_{0i}\mu_j^2 + \alpha_{0,j}\mu_i^2}{(4\pi\varepsilon_0)^2}\frac{1}{r^6}, \tag{2.7}$$

where $\alpha_0$ is the electronic polarisability.

The molecules that do not have a permanent dipole or quadrupole, still possess an instantaneous (induced) dipole, which is caused by a rapid fluctuation in the electronic charge distribution around the nucleus. These interactions are known variously as London forces, dispersion forces, charge-fluctuation forces, or induced-dipole-induced dipole forces. Whereas all other forces depend on the specifics of the charge distribution, dispersion forces are found to act between all molecules. London interactions can be represented as [17]

$$u(r) = -\frac{3}{2}\frac{\alpha_{0i}\alpha_{0,j}}{(4\pi\varepsilon_0)^2}\frac{I_iI_j}{I_i+I_j}\frac{1}{r^6}, \tag{2.8}$$

where $I$ is the first ionisation potential and refers to the work which is required to remove one electron from an uncharged molecule.

For molecules with a dipole moment of less than about 1 D, the dispersion force is dominant. Even for strongly polar molecules, dispersion forces are not negligible. Compared to the dispersion interaction, induction forces are much weaker, both for polar and non-polar substances.

As a consequence of very anisotropic charge distributions in molecules, in addition to the aforementioned physical forces, there are forces which are referred to as quasi-chemical interactions [17], one of the most important of which is the hydrogen-bonding interaction. Hydrogen bonds are strongly directional attractive forces and they are typically formed between hydrogen and a strong electronegative atoms, such as fluor F, oxygen O or nitrogen N. Molecules that contain hydrogen bonds exhibit a strong tendency to associate with other molecules that contain electronegative atoms (F, O, N). Hydrogen bonds are responsible for a number of interesting and distinct physical phenomena, as for instance highly increased boiling and melting points or formation of chemical complexes, such as dimers for organic acids, hexamers for hydrogen fluoride, and three-dimensional networks for water [13]. Water with its numerous anomalies, such as the density maximum at 4° C, the high dielectric constant, the high surface tension and the hydrophobic effect with other components, is a particular good example of a system with highly pronounced hydrogen bonding interactions.

Understanding the nature of intermolecular forces is necessary to interpret a system's phase behaviour and therefore is a key to the choice of an appropriate thermodynamic models. Modelling intermolecular forces is challenging: for systems that contain moieties of different polarities, the various forces have to be taken into account. For this purpose, effective intermolecular potentials which combine several interactions into a single overall interaction are useful in molecular modelling.

### 2.1.2 Pair potentials

The full potential energy $\mathcal{V}$ of a system can be expressed as a summation of the energies of atoms due to an external field ($u(\mathbf{r_i})$), the pairwise interaction between molecules

($u(\mathbf{r_i}, \mathbf{r_j})$), the three-body ($u(\mathbf{r_i}, \mathbf{r_j}, \mathbf{r_k})$), and many-body contributions:

$$\mathcal{V} = \sum_i u(\mathbf{r_i}) + \sum_i \sum_{j>1} u(\mathbf{r_i}, \mathbf{r_j}) + \sum_i \sum_{j>1} \sum_{k>j>i} u(\mathbf{r_i}, \mathbf{r_j}, \mathbf{r_k}) + ... \tag{2.9}$$

The pairwise interaction provides the largest contribution to the full potential energy and depends only on separation between two particles, exhibiting repulsion at short and attraction at longer distances for the most forms of the intermolecular potentials. The three-body term involves the triplets of molecules and is important at liquid densities. The four- and higher-body terms are expected to be small in comparison with the two- and three-body contributions [18]. The many-body terms are computationally highly expensive. Therefore it is common to incorporate the average three- and many-body interactions into an effective pair potential by assuming pair-wise additivity:

$$\mathcal{V} \simeq \sum_i \sum_{j>1} u^{eff}(r_{i,j}). \tag{2.10}$$

A number of different mathematical functions are commonly used to describe pair potentials, but they all share some common features. Each interaction site in the pair potential represents an atom or group of atoms. The pair potential is based on molecule specific parameters, such as the ranges of the repulsive and attractive interactions, the size of the bead and the energy parameter which characterises the strength of the attractive interaction. The potential is usually centre-symmetric, thereby providing a description of spheres or molecules formed from spherical segments. At short distances, the potential is invariably repulsive and tends to infinity, whereas at larger separations, the potential is attractive and tends to zero so that the energy remains finite. The total potential energy is usually assumed as a sum of both, the repulsive and the attractive, contributions. Some of the commonly used pair potentials are illustrated in Figure 2.1 and briefly described below.

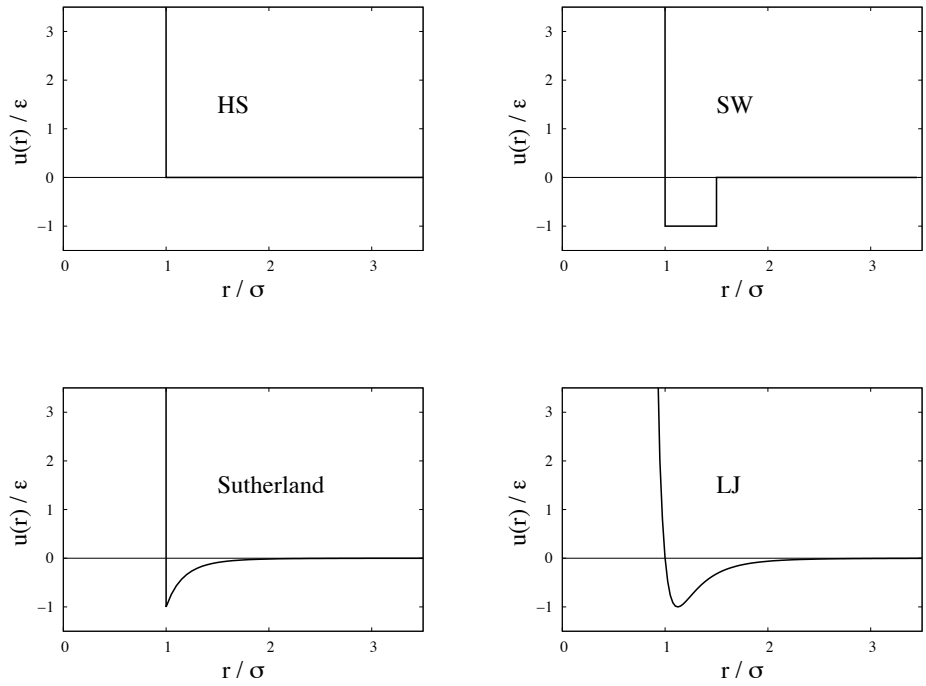



**Figure 2.1:** Commonly used intermolecular potentials: hard-sphere (HS) (top left), square-well (SW) (top right), Sutherland (bottom left), and Lennard-Jones (LJ) (bottom right).

Ideal gas model is the simplest trivial case, where one assumes no interaction between molecules at any distance $u(r) = 0$.

The hard-sphere potential (HS) is one of the mostly studied potentials. The interaction sites are represented as impenetrable spheres that cannot overlap. Based on only one interaction parameter $\sigma$, the potential is characterised by an infinite repulsion for distances $r < \sigma$ and disregards the attractive forces at distances $r \geq \sigma$:

$$u(r) = \begin{cases} \infty & r < \sigma \\ 0 & r \geq \sigma. \end{cases} \tag{2.11}$$

The square-well potential (SW) incorporates a hard-sphere potential with an additional attractive part, described by a well of depth $\epsilon$ and width $\lambda$. Based on three adjustable parameters, the SW potential has a certain versatility and therefore is a better representation

for the real fluids compared to the HS potential:

$$u(r) = \begin{cases} \infty & r < \sigma \\ -\epsilon & \sigma \leq r < \lambda\sigma \\ 0 & r \geq \lambda\sigma. \end{cases} \tag{2.12}$$

In the Sutherland and the Yukawa potentials a hard-sphere repulsion is incorporated with the attractive force. Whereas the SW potential is discontinuous and has a finite attraction range, in the Sutherland and Yukawa potentials the attraction force is continuous and converges to zero at large separations. Namely, in the Sutherland potential, the attraction is described by the London dispersion interaction

$$u(r) = \begin{cases} \infty & r < \sigma \\ -\epsilon \left(\frac{\sigma}{r}\right)^6 & r \geq \sigma, \end{cases} \tag{2.13}$$

whereas in the Yukawa potential, the attraction force is of the form of an exponentially damped Coulombic term

$$u(r) = \begin{cases} \infty & r < \sigma \\ -\epsilon \frac{\exp\left[-\lambda\left(\frac{r}{\sigma}-1\right)\right]}{r/\sigma} & r \geq \sigma. \end{cases} \tag{2.14}$$

Due to the discontinuity, these potentials are, however, difficult to implement in the conventional Molecular Dynamics simulations. Therefore, continuous potentials, such as modified Buckingham, Lennard-Jones, and Mie potentials, appear advantageous for this type of applications.

Modified Buckingham potential, also referred to as the exp-6 potential, is represented by a London type of attraction and a Slater-like exponential function for the repulsive term. The potential function contains three adjustable parameters: the minimum potential energy $\epsilon$, the intermolecular separation $r_{min}$, and the steepness of the repulsive wall $\gamma$. The functional form is given by

$$u(r) = \frac{\epsilon}{1-(6/\gamma)} \left( \frac{6}{\gamma} \exp\left[\gamma\left(1 - \frac{r}{r_{min}}\right)\right] - \left(\frac{r_{min}}{r}\right)^6 \right) \tag{2.15}$$

The Lennard-Jones potential (LJ), which was first proposed by Lennard-Jones [19, 20], is a continuous interaction based on two adjustable parameters $\sigma$ and $\epsilon$, which is now

invariable expressed in the (12-6) form:

$$u(r) = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^{6} \right]. \tag{2.16}$$

The Mie potential, firstly proposed by Mie in 1903 [21], provides a greater flexibility compared to the conventional LJ (12-6) potential afforded by the variable repulsive and attractive exponents. After some algebraic rearrangements, the Mie potential can be expressed as

$$u(r) = \epsilon \mathcal{C} \left[ \left( \frac{\sigma}{r} \right)^{\lambda_r} - \left( \frac{\sigma}{r} \right)^{\lambda_a} \right], \tag{2.17}$$

where

$$\mathcal{C}(\lambda_a, \lambda_r) = \left( \frac{\lambda_r}{\lambda_r - \lambda_a} \right) \left( \frac{\lambda_r}{\lambda_a} \right)^{\frac{\lambda_a}{\lambda_r - \lambda_a}}. \tag{2.18}$$

The functional form of the Mie potential (Figure 2.2) is defined in a way that the minimum is located at $-\epsilon$ [$\epsilon = -u_{min}(r)$] and the intersection with the $x$-axis, where the potential switches from positive to negative, is located at $\sigma$ [$u(\sigma) = 0$]. The exponents controlling the softness of the repulsion $\lambda_r$ and the range of the attraction $\lambda_a$ are not fixed to 12 and 6, like in the standard Lennard-Jones version, but can be adjusted to better represent the target physical properties of the system. Using the versatile exponents has been shown to provide a significant improvement of the vapour pressure and the second-derivative thermodynamic properties of real fluids, such as speed of sound, heat capacity, and compressibility [22–24].

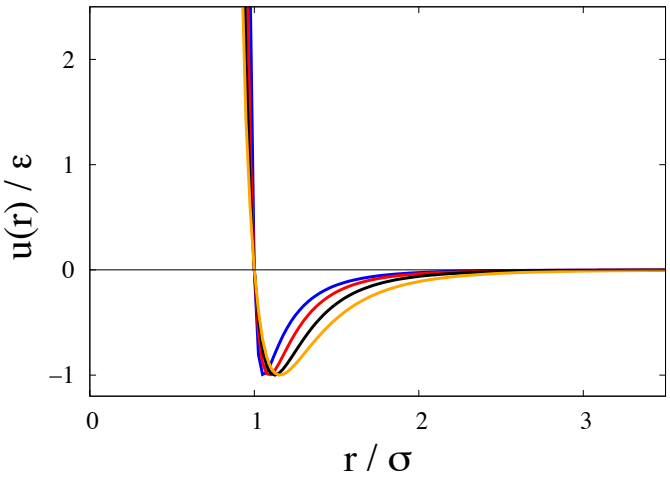



**Figure 2.2:** Intermolecular potentials of the Mie family with $\lambda_a = 6$ and $\lambda_r = 40$ (blue), $\lambda_r = 20$ (red), $\lambda_r = 12$ (black), and $\lambda_r = 8$ (yellow).

## 2.2 Statistical mechanics

The knowledge of intermolecular interaction potential is crucial in order to determine the thermodynamic properties. A link between the microscopic details, such as positions and momenta of the particles, and macroscopic properties of the system, e.g. the pressure, chemical potential etc., is provided by statistical mechanics. The thermodynamic state of the system is exactly defined by a set of variables appropriate for a given thermodynamic ensemble, such as for instance the number of particles $N$, volume $V$, and temperature $T$ in the case of the canonical ensemble. It is clear that the measured the macroscopic properties, one does not require the position and momenta of every single particle, which would define the instantaneous microscopic state. The experimentally observable macroscopic properties are time averages over a number of microscopic states evolved in time. Although the instantaneous value of the experimental observable will deviate from the average value, these fluctuations are negligibly small compared to the average in the thermodynamic limit due to a very large number of particles involved and the time resolution of the measurement. In conventional statistical mechanics, it is common to replace time averages with ensemble averages. This hypothesis, currently referred to as the ergodic hypothesis, states that the time and the ensemble averages of two systems with the same

state variables, e.g., $N$, $V$, and $T$ are identical in infinite time and thermodynamic limits. As an ensemble, one can imagine a collection of subsystems in phase space, where each subsystem is a replica of the macroscopic system of interest. The most commonly used ensemble is probably the canonical ensemble, also known as $NVT$ ensemble due to its characteristic state variables: number of particles, volume, and temperature. The thermodynamic properties can be derived from the corresponding distributions of probability functions. In the canonical ensemble, the probability to find a subsystem in a particular energy state $E_i$ is given by Boltzmann distribution

$$P(E_i) = \frac{\exp(-\beta E_i)}{\sum_i \exp(-\beta E_i)}, \tag{2.19}$$

where $\beta = \frac{1}{k_B T}$. The denominator is a sum over all states $i$ of the system and is referred to as the canonical partition function:

$$Q_{NVT} = \sum_i \exp(-\beta E_i). \tag{2.20}$$

The mean thermodynamic properties, e.g., the average internal energy $U$, can be calculated as a sum of the products of energy of each subsystem and the corresponding probability:

$$U =< E >= \sum_i E_i P_i = \frac{\sum_i E_i \exp(-\beta E_i)}{Q}. \tag{2.21}$$

From definition of $Q$ in Eq. 2.20, its derivative is given as

$$\frac{\partial Q}{\partial \beta} = \sum_i E_i \exp(-\beta E_i) \tag{2.22}$$

Thus, by inserting Eq. 2.22 in Eq. 2.21 we obtain :

$$U = -\frac{1}{Q}\left(\frac{\partial Q}{\partial \beta}\right) = -\left(\frac{\partial \ln Q}{\partial \beta}\right). \tag{2.23}$$

The partition function $Q_{NVT}$ can be constructed in terms of the phase space continuum and the classical Hamiltonian $\mathcal{H}$, which corresponds to total energy of the system and can be obtained a sum of the potential $\mathcal{V}$ and the kinetic energy $\mathcal{K}$:

$$\mathcal{H}(\mathbf{r^N}, \mathbf{p^N}) = \mathcal{V}(\mathbf{r^N}) + \mathcal{K}(\mathbf{p^N}), \tag{2.24}$$

where $\mathbf{r^N}$ is the generalised position of $N$ particles and $\mathbf{p^N}$ is their generalised momenta. The potential energy can be approximated a sum over all effective pair potentials (Eq. 2.10) and the kinetic energy (for spherical particles) takes the form [18]

$$\mathcal{K} = \sum_{i=1}^{N} p_i^2/2m_i, \tag{2.25}$$

where $m_i$ is the molecular mass, and $p_i^2$ is the squared momentum of molecule $i$ in $x$, $y$, and $z$ dimensions.

The direct probability in Eq. 2.19 can be substituted in the classical approximation as the phase space probability density function $f^N(\mathbf{r^N}, \mathbf{p^N})$, which is expressed as

$$f^N(\mathbf{r^N}, \mathbf{p^N}) = \frac{1}{N!}\frac{1}{h^{3N}}\left(\frac{\exp(-\beta\mathcal{H}(\mathbf{r^N}, \mathbf{p^N}))}{Q}\right). \tag{2.26}$$

The term $h^{3N}$ results from the Heisenberg uncertainty principle, stating that the position and the momentum of the particle cannot precisely be determined simultaneously. The term $1/N!$ accounts for the indistinguishability of the particles and is introduced in order to avoid over-counting of the states.

Instead of a summation over the energy levels as in Eq. 2.20, the canonical partition function can be expressed as an integral of the Boltzmann factor of $\mathcal{H}$ over the phase space $d\mathbf{r}^N d\mathbf{p}^N$:

$$Q = \frac{1}{N!h^{3N}} \int .. \int \exp(-\beta\mathcal{H}(\mathbf{r^N}, \mathbf{p^N}))d\mathbf{r}^N d\mathbf{p}^N \tag{2.27}$$

The energy can be decomposed into the kinetic (**p**-dependent) and the potential (**r**-dependent) energy contributions, so that the partition function can be expressed as a product of two integrals of the ideal (kinetic) and residual (potential) parts, which can be considered independently:

$$Q = \frac{1}{N!h^{3N}} \int \exp(-\beta\mathcal{K}(\mathbf{p}^N))d\mathbf{p}^N \int \exp(-\beta\mathcal{V}(\mathbf{r}^N))d\mathbf{r}^N = Q^{id}Q^{ex} \tag{2.28}$$

For a monoatomic system, the kinetic energy is given by

$$Q^{id} = \frac{V^N}{N!\Lambda^{3N}}, \tag{2.29}$$

where $\Lambda = (h^2/2\pi m k_B T)^{1/2}$ is the thermal de Broglie wavelength.

The residual contribution $Q^{res}$ in Eq. 2.28 can be expressed as

$$Q^{res} = \frac{Z}{N!\Lambda^{3N}}, \tag{2.30}$$

with $Z$ being a configurational integral

$$Z = \int \exp(-\beta \mathcal{V}(\mathbf{r}^N)) d\mathbf{r}^N. \tag{2.31}$$

The link between the microscopic and macroscopic properties is given through the relation between the canonical partition function $Q$ and the Helmholtz free energy $A$. According to the fundamental thermodynamic relation $A = U - TS$, so that

$$\frac{A}{T} = \frac{U}{T} - S. \tag{2.32}$$

Substituting $U$ by Eq. 2.23 we obtain

$$\frac{\partial}{\partial T}\left(\frac{A}{T}\right) = \frac{-U}{T^2} = \frac{1}{T^2}\left(\frac{\partial \ln Q}{\partial \beta}\right) = -k_B\left(\frac{\partial \ln Q}{\partial T}\right) \tag{2.33}$$

indicating that

$$A = -k_B T \ln Q. \tag{2.34}$$

Once $A$ is known, all other macroscopic properties can be derived by fundamental thermodynamic relations. For example, the pressure can be obtained as the appropriate partial derivative of the Helmholtz free energy with respect to the volume:

$$P = -\left(\frac{\partial A}{\partial V}\right)_{N,T} = k_B T\left(\frac{\partial \ln Q}{\partial V}\right)_{N,T} = k_B T\left(\frac{\partial \ln Z}{\partial V}\right)_{N,T}. \tag{2.35}$$

The partition function and the corresponding macroscopic properties can be developed not only in $NVT$ but also in all other ensembles. For instance, in the isothermal-isobaric ensemble $NPT$, the probability density is proportional to

$$\exp(-\beta\left(\mathcal{H}(\mathbf{r^N}, \mathbf{p^N}) + PV\right)), \tag{2.36}$$

Hence, the appropriate partition function is

$$Q_{NPT} = \frac{1}{N!h^{3N}}\frac{1}{V_0}\int dV \int \exp\left(-\beta(\mathcal{H}(\mathbf{r^N}, \mathbf{p^N})) + PV\right) d\mathbf{r}^N d\mathbf{p}^N, \tag{2.37}$$

where $V_0$ is a basic unit volume which is chosen to yield a dimensionless $Q_{NPT}$. The configurational integral in $NPT$ ensemble is given as

$$Z_{NPT} = \int dV \exp(-\beta PV) \int d\mathbf{r}^N \exp\left(-\beta \mathcal{V}(\mathbf{r}^N)\right) \tag{2.38}$$

For the spherical particles, the corresponding link with a macroscopic thermodynamic property is through the Gibbs free energy $G$ obtained as:

$$G = -k_B T \ln Q_{NPT}. \tag{2.39}$$

The calculation of the partition function which includes the configuration integral $Z$ is, however, not straightforward. Apart from very few simple systems such as the ideal gas or the Einstein crystal, there is no exact solution of the configurational integral and some approximations have to be made in order to evaluate it. The main challenge represents the evaluation of expression for the potential energy $\mathcal{V}(\mathbf{r}^N)$ as it requires a knowledge of the particle coordinates in the $3N$-dimensional space $d\mathbf{r}^N$. The average positions of the particles can be described by means of the radial distribution function (RDF) $g(r)$ of the fluid. The average of the intermolecular potential energy $\mathcal{V}(\mathbf{r}^N)$, or the residual energy $U^{res}$, can be expressed in terms of the "energy equation" [25]:

$$< \mathcal{V}(\mathbf{r}^N) >= U^{res} = 2\pi\rho N \int_0^\infty g(r_{12}) u(r_{12}) r_{12}^2 dr_{12}, \tag{2.40}$$

which indicates the dependency of the intermolecular potential energy on the RDF and the intermolecular potential.

The equation of state and molecular simulations are the two main techniques that allow one to apply the basis of statistical mechanics to provide a link between the explicit intermolecular force fields and the macroscopic thermodynamic properties.

### 2.2.1 Equation of state

An equation of state (EoS) is a closed form mathematical relation which links the macroscopic thermodynamic variables of a system. In its more common form it explicitly relates pressure $P$ , temperature $T$ , volume $V$, and composition. The simplest equation of state, which neglects intermolecular interaction, is the ideal gas equation first formulated by Clapeyron in 1834 [26]. It is often introduced in its common form

$$PV = nRT, \tag{2.41}$$

where $n$ is the amount of substance and $R$ is the ideal gas constant.

The Helmholtz free energy of ideal gas can be expressed as [27]

$$\frac{A_{ideal}}{Nk_BT} = \ln \frac{N\Lambda^3}{V} - 1. \tag{2.42}$$

The ideal gas equation of state can be derived exactly from statistical mechanics by invoking a fluid made of point particles with no intermolecular forces and which occupy no volume. The consequence of these conditions is that the ideal gas model shows no cohesive forces, i.e. it only describes a gas phase with no phase change. An extension of the ideal gas EoS to take into account intermolecular forces in a sensible way are the various forms of the virial EoS. However, the virial EoS faces the problems of convergence at high densities and is therefore not suitable to describe the liquid phase or the vapour-liquid equilibrium.

A major improvement, which later set the basis for the fluid theories and molecular-based equations of state, was the expression proposed by van der Waals in his Ph.D thesis at the University of Leiden in 1873 [28]. It was the first attempt to describe the phase behaviour exhibited by the real fluids, by ascribing a finite volume and interactions between the particles. The particles were represented as hard spheres with an additional attractive tail. It was rigorously shown by Kac and co-workers [29–31] in 1963 that for the van der Waals fluid described by the hard core repulsion and a weak very long-ranged attraction ($\lambda \to \infty$, $\epsilon \to 0$), the partition function in the thermodynamic limit is exact and finite. The volume occupied by a particle, or more precisely the volume $b$ excluded by one particle to another (Figure 2.3) was defined as an impenetrable volume per sphere for a pair of spheres separated by distance $\sigma$, which results in $b = (2/3)\pi\sigma^3$, four times the molecular hard-core volume.

One may derive the van der Waals EoS using modern statistical mechanical treatments. The free energy can be decomposed into the repulsive and the attractive contributions. For a van der Waals fluid, the Helmholtz free energy of the repulsive contribution corresponds to that of a hard sphere with excluded volume $b$ given by

$$\frac{A_{HS}}{Nk_BT} = \ln \frac{N\Lambda^3}{V - Nb} - 1 = \ln \frac{N\Lambda^3}{V} - 1 - \ln \frac{V - Nb}{V} = \frac{A_{ideal}}{Nk_BT} - \ln \frac{V - Nb}{V}. \tag{2.43}$$

The attractive contribution can be obtained from Eq. 2.40. By employing the mean-field

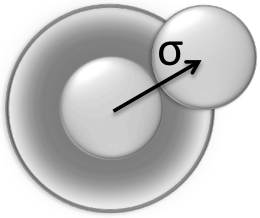



**Figure 2.3:** The excluded volume per pair of hard spheres. Two hard spheres are separated by $\sigma$, and the volume excluded by the central particle to the second particle (grey shadow) is $(4/3)\pi\sigma^3$. Per molecule, the excluded volume $b = (2/3)\pi\sigma^3$.

approximation for the radial distribution function of the reference hard-sphere $g_{HS}(r) = 1$, we obtain

$$A = A^{HS} + 2\pi\frac{N^2}{V}\int_{\sigma}^{\infty} u(r)r^2 dr. \tag{2.44}$$

After introducing the van der Waals integrated energy

$$a = -2\pi\int_{\sigma}^{\infty} u(r)r^2 dr, \tag{2.45}$$

Eq. 2.44 takes the form

$$A = Nk_BT\left(\ln\frac{N\Lambda^3}{V} - 1 - \ln\frac{V - Nb}{V}\right) - \frac{N^2}{V}a. \tag{2.46}$$

By differentiating Eq. 2.46 with respect to the volume, we obtain the well-known form of the van der Waals EoS

$$p_{vdW} = \frac{Nk_BT}{V - Nb} - \frac{N^2}{V^2}a, \tag{2.47}$$

or in molar units

$$p_{vdW} = \frac{RT}{v - b_m} - \frac{a_m}{v^2}. \tag{2.48}$$

The van der Waals EoS is the first equation of state that qualitatively reproduces the phase behaviour of a real fluid. It is able to describe both vapour and liquid phases and the existence of the critical point. Additionally, the van der Waals EoS can be used to predict different types of mixture behaviour classified by Scott and van Konynenburg [32, 33]. Moreover, the van der Waals EoS can be recast in a universal reduced form by scaling the

thermodynamic variables ($T$, $P$, $V$) with respect to the critical values. This forms the basis of the corresponding state principle allowing to describe all fluids in a generalised manner [16]. Quantitatively, however, the van der Waals EoS largely overestimates the pressure and the compressibility factor compared to the real fluids. While the van der Waals critical compressibility factor is $Z_c = 3/8 = 0.375$ at the critical point [1], most of the common fluids have a value $Z_c \approx 0.291$.

Due to its mathematical simplicity, van der Waals EoS is very convenient for engineering use and is commonly implemented for a range of industrial applications. A fundamental problem with vdW EoS is the treatment of the repulsive term, which is largely overestimated due to the non-additivity of the excluded volume. At high density, the excluded volumes overlap with each other, thereby resulting in a smaller excluded volume than the van der Waals term $Nb$. An improvement can presumably be achieved, if the repulsive term is given as a function of density, and in fact, this was pointed out by van der Waals himself in his thesis. A very accurate description of a hard-sphere fluid was proposed by Carnahan and Starling in 1969 [34], where pressure is given as

$$p_{\mathrm{CS}} = \frac{RT}{v} \times \frac{1 + y + y^2 - y^3}{(1-y)^3}, \tag{2.49}$$

with $y = b/(4v)$ leading to CS-vdW EoS [35]

$$p_{\mathrm{CS-vdW}} = \frac{RT}{v} \times \frac{1 + y + y^2 - y^3}{(1-y)^3} - \frac{a_m}{v^2}. \tag{2.50}$$

However, by including the repulsive term as proposed by Carnahan and Starling, the analytical cubic form of the equation has to be abandoned. Instead, a lot of effort was devoted to modifications of the attractive term in order to keep the simple analytical form of the equation. In 1949, Redlich and Kwong suggested to modify the attractive term by incorporating a simple temperature dependence. This observation spawned a family of equations, known a cubics EoS, which have flooded the literature and have become a staple of engineering thermodynamics. The most well-known modifications of van der Waals EoS are probably the Redlich-Kwong (RK) [36], and the Soave modification of the

[1] Critical compressibility factor is obtained from $Z_c = p_c v_c / RT_c$; the critical values for molar volume, pressure and temperature are obtained by solving the differential equations at the critical point: $p > 0$, $(\partial p/\partial v)_T = 0$, and $(\partial^2 p/\partial v^2)_T = 0$ and given as $v_c = 3b$, $p_c = \frac{a}{27b^2}$, and $T_c = \frac{8a}{27Rb}$

Redlich-Kwong (SRK) [10] and the Peng-Robinsion (PR) [9] equations of sate:

$$p_{RK} = \frac{RT}{v-b} - \frac{a}{\sqrt{T}v(v+b)}, \tag{2.51}$$

$$p_{SRK} = \frac{RT}{v-b} - \frac{a\alpha(T)}{v(v+b)}, \tag{2.52}$$

$$p_{PR} = \frac{RT}{v-b} - \frac{a\alpha(T)}{v^2+2bv-b^2}, \tag{2.53}$$

where $\alpha(T)$ is a function of temperature and the acentric factor.

## 2.3 SAFT-$\gamma$ Mie equation of state

### 2.3.1 Short history of SAFT

The conventional cubic equations of state, such as the seminal relation of van der Waals [28] and its more modern counterparts, such as PR [9], RK [36], and SRK [10] relations, are based on a hard-sphere repulsive term and mean-field attractive term. This leads to several limitations: first, the hard-sphere reference makes the equation mostly suitable for simple spherical molecules, and cannot be accurately used to describe chain or associating compounds; second, as the model is not expressed explicitly in terms of intermolecular potential, it cannot be tested directly by molecular simulations. Therefore, improvements can be performed only in a heuristic and empirical manner, which limits the predictive capability of these models. An approach to tackle this is to consider chain fluids as the basis for a perturbation approach. Such is the nature of some of the first of the so-called "molecular" equations such as the perturbed hard chain theory [37]. Other approaches, such as the lattice theories and polymer treatments deserve mention, but will not be expanded upon. For a comprehensive treatment of EoS the reader is referred to Ref. [38]. In 1980s, Wertheim presented series of papers [39–42], with a theory of strong directional attractions, such as hydrogen bonds, treated as a separate perturbation term to the free energy. The Wertheim formalism can be expressed as a first-order thermodynamic perturbation theory (TPT1). Inspired by the Wertheim TPT, the Gubbins's group proposed the Statistical Association Fluid Theory (SAFT) [43] to describe the thermodynamic and

fluid-phase equilibria of the spherical and the chain molecules with one or more hydrogen-bonding association sites. As illustrated in Figure 2.4, the Helmholtz free energy includes separate ideal, monomer, chain, and association contributions:

$$\frac{A}{Nk_BT} = \frac{A^{Ideal}}{Nk_BT} + \frac{A^{Mono}}{Nk_BT} + \frac{A^{Chain}}{Nk_BT} + \frac{A^{Assoc}}{Nk_BT}. \tag{2.54}$$

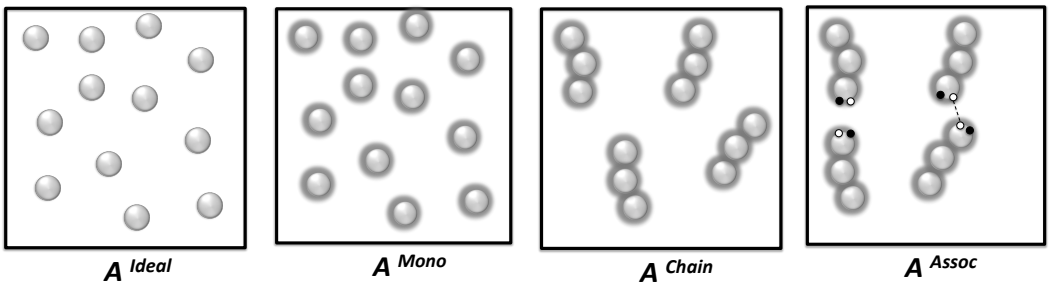



**Figure 2.4:** Schematic representation of different perturbation contributions within the SAFT framework.

The associating contribution $A^{Assoc}$ is based on the Wertheim TPT1. The chain contribution $A^{Chain}$ can be obtained from $A^{Assoc}$ in the limit of complete association, i.e. the association bonds correspond to covalent bonds. For the monomer term $A^{Mono}$, the hard-sphere fluid described by an accurate relation by Carnahan and Starling [34] is chosen as a reference. In order to mimic the real fluid behaviour, a Barker and Henderson - like perturbation approach is used to describe the dispersive attractive term.

Twenty five years after the appearance of the original SAFT EoS [43, 44], many versions of this model have now been developed by various research groups around the world. A detailed description of various versions along with the different applications is a subject of several comprehensive reviews [45–48]. The main differences in the various reformulations arise from the different reference fluids and the underlying intermolecular potentials employed. One exception from the common SAFT framework is the PC SAFT version [49], where instead of the monomer, the chain term is chosen as a reference from which to obtain the dispersion contributions. The perturbed chain term is determined from a Taylor expansion correlated to the VLE data of pure alkanes. In this manner, the direct link between the underlying molecular model and the intermolecular potential is essentially lost.

The original version of SAFT [43, 44], which is referred to as the first generation of

this equation of state, is based on a reference hard-sphere potential for the associative and chain contributions. The models from the second generation of SAFT are based on more realistic attractive intermolecular force fields, such as SW, Yukawa, Sutherland, LJ, and Mie potentials. The SAFT LJ [50, 51] and soft-SAFT [52] versions are based on a reference LJ monomer fluid, the radial distribution function of which was exhaustively parameterised using the computer simulation data. The SAFT-VR EoS [53, 54], referred to as the Variable Range version, was originally developed for the SW, Yukawa, and Sutherland potentials. Up until quite recently the SAFT-VR has been associated with the square-well potential. In the SAFT-VR development, the square-well monomer term was described by means of the second-order Barker and Henderson perturbation theory, and the chain and monomer terms account for the structure of the square-well fluid. The recent work of Lafitte *et al.* [23, 24] extended the SAFT-VR version to the family of the Mie $\lambda_r - \lambda_a$ potentials. Using different steepness (hardness and softness) and range of the potential allows for an improvement in the representation of the dense liquid phase and the second-derivative thermodynamic properties, such as the heat capacity, isothermal compressibility, and speed of sound, while still retaining an accurate representation of the vapour-liquid equilibrium properties. The versatile form of the Mie potential allows one to characterise a wide variety of fluids of very different chemical nature. The description of the perturbation expansion of the thermodynamic properties of the Mie monomer fluid has recently been extended to third order, thereby enabling a more accurate representation of the near-critical region [24]. This latest SAFT-VR Mie version [24] is basis for a third generation SAFT-$\gamma$ EoS [55] ($\gamma$ is the third letter in the greek alphabet), which cast as a group-contribution method to treat heteronuclear molecules formed from segments of differing chemical functionality.

### 2.3.2 SAFT-$\gamma$ Mie model and theory

In our current work, we apply the group-contribution SAFT-$\gamma$ Mie EoS to estimate the force field parameters for the complex real fluids using suitable macroscopic thermodynamic experimental data. The intermolecular models developed with the EoS can subsequently be used for direct molecular simulation in order to predict properties that are not accessible from the equation of state, such as interfacial, structural, and dynamic properties. The strength of a group-contribution (GC) method lies in its predictive capa-

bility: based on a set of interaction parameters for a limited number of chemical groups, new pure components and multi-component mixtures can be described without the need for additional fitting. Group-contribution methods are based on the assumption that the properties of chemical components can be determined as functions of chemical groups the components comprise. This implies that the contribution of each functional group is independent of the molecular structure and that the group behaves in the same manner, independent of the the environment. The framework of the method can be extended to account for the position of the group in the molecule by considering second-order groups [56]. The idea of the group-contribution approach was pioneered by Prausnitz and coworkers in the development of the successful universal quasi-chemical functional group activity coefficient (UNIFAC) method [57]. The limitation of UNIFAC, or any other such activity coefficient model, is that it is based on a lattice model and as a consequence is strictly suitable to describe the liquid phase and neglect pressure effects. By contrast, the advantage of an equation of state is that it can be used to describe the liquid and vapour phases on an equal footing.

In the SAFT-$\gamma$ Mie framework, the free energy of the fluid is expressed explicitly in terms of the parameters characterising the various functional groups. The corresponding heteronuclear description is implemented in both - the monomer and the chain contributions, with the latter being described in terms of an effective molecular contact value of the radial distribution function. The force fields developed with the SAFT-$\gamma$ Mie EoS can be used in a transferable way, i.e. once the group parameters have been estimated from the macroscopic thermodynamic properties of a particular species (or homologous families), they can be used in the representation of other compounds which contain the given group along with additional functional groups. For instance, once the parameters for the homologous series of alkanes have been obtained, they can be used to represent the methyl $CH_3$ and methylene $CH_2$ groups in surfactant molecules from the alkyl polyoxyethylene glycol family. The parameters for groups of different type obtained from the parameterisation of the pure components, can be used to predict properties of mixtures and vice versa.

In our current work the focus is on the description of non-associating compounds formed from tangentially bonded chains. This choice is made to facilitate the use of the force fields developed with the EoS in traditional MD simulation. The treatment of models with additional short-ranged association sites, which are described by the discontinuous

potentials, typically of the square-well form, is computationally demanding and requires the development of specific algorithms to solve the equations of motion for the discontinuous potential. Furthermore, the use of models comprising tangentially bonded segments makes the link between the theory and simulation straightforward; the development of force fields within the SAFT framework for highly fused segments is more problematic. The description of molecules formed from tangentially bonded segments within the SAFT-$\gamma$ formalism [55, 58, 59] simply implies the use of shape factor with a value of unity.
The fluids of interest are thus described by a Helmholtz free energy comprising the usual SAFT contributions

$$\frac{A}{Nk_BT} = \frac{A^{Ideal}}{Nk_BT} + \frac{A^{Mono}}{Nk_BT} + \frac{A^{Chain}}{Nk_BT}. \tag{2.55}$$

Each term of the free energy is described in greater detail in Appendix A.

**Estimation of parameters for the SAFT-$\gamma$ Mie Force Field**

After the Helmholtz free energy has been specified, the other thermodynamic properties can be obtained from the standard relations, e.g. the pressure $p = -(\partial A/\partial V)_{T,N}$, chemical potential of the component $i$ $\mu_i = (\partial A/\partial N_i)_{T,V,N_j}$, and the Gibbs free energy $G = A + pV = \sum_i^{N_C} n_i\mu_i$. The phase equilibrium properties can then be determined in the usual way by ensuring that the temperature, pressure, and chemical potential of each component in each phase are equal and that the overall Gibbs free energy corresponds to a global minimum.

In our work, the liquid saturation densities and vapour pressures are used as target properties in the force field parameter estimation procedure. Other properties, such as high-temperature density isobars and second-derivative properties (such as speed of sound, heat capacity, or compressibility), can also be included. The parameters are estimated by minimising the difference between the experimental and calculated values of the chosen properties over a range of thermodynamic conditions. Due to the classical mean-field nature of the SAFT EoS, which does not properly account for the density fluctuations near the critical region, the consideration of data close to the critical point would lead to unphysical features in the parameters. State points between the triple point and up to 90 % of the critical point of the pure components are therefore included in the parameterisation. The minimisation is carried out by means of the relative least-squares objective function,

which consists of the sum of individual residues:

$$minF_\theta = \frac{w_1}{N_{P_V}} \sum_{i=1}^{N_{P_V}} \left[ \frac{P_{i,V}^{exp}(T_i) - P_{V,i}^{SAFT}(T_i,\theta)}{P_{V,i}^{exp}(T_i)} \right]^2 + \frac{w_2}{N_{\rho_L}} \sum_{j=1}^{N_{\rho_{sat}}} \left[ \frac{\rho_{sat,j}^{exp}(T_j) - \rho_{sat,j}^{SAFT}(T_j,\theta)}{\rho_{sat,j}^{exp}(T_j)} \right]^2, \tag{2.56}$$

where $N_{P_V}$ and $N_{\rho_{sat}}$ are the number data points of vapour pressures $P_V$ and saturated liquid densities $\rho_{sat}$, respectively. $\theta$ is the vector of parameters of the intermolecular potential, which includes the diameter, well-depth, and repulsive exponent ( $\sigma, \epsilon, \lambda_r$), while the attractive exponent is fixed to the London dispersion value ($\lambda_a = 6$). The superscripts $exp$ and $SAFT$ refer to experimental data points and values calculated with the SAFT-$\gamma$ Mie EoS, respectively. Equal weights are usually taken for these properties ($w_1 = w_2$).
A substantial asset of this methodology is that the parameter space can be explored efficiently over a wide range of thermodynamic conditions, thereby providing a near optimal solution that is transferable to various state points. By contrast, commonly used bottom-up approaches often make use of temperature dependent parameters that have to be re-derived at various state points. The parameters obtained with our SAFT-$\gamma$ top-down methodology can be used as a direct input in molecular simulation. Unlike intermolecular interactions can also be obtained by using appropriate experimental data for the mixtures, but often combining rules are employed for some of the parameters as described in the next section.

**Combining rules for the unlike intermolecular interactions**

Simple combining rules can be applied in order to describe the unlike interactions between the pure components $i$ and $j$. The unlike size parameter $\sigma_{ij}$ is obtained by arithmetic mean, originating from the classical Lorentz combining rule:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \tag{2.57}$$

The unlike repulsive exponent $\lambda_{r,ij}$ is obtained from the geometric mean of the integrated van der Waals energy of a Sutherland fluid

$$\alpha_{vdW,ij}^s = 2\pi\epsilon_{ij}\sigma_{ij}^3 \left( \frac{1}{\lambda_{ij} - 3} \right), \tag{2.58}$$

based on the Berthelot combining rule, $\alpha^s_{vdW,ij} = (\alpha^s_{vdW,ii}\alpha^s_{vdW,jj})^{1/2}$, which results in

$$\lambda_{r,ij} - 3 = \sqrt{(\lambda_{r,ii} - 3)(\lambda_{r,jj} - 3)}. \tag{2.59}$$

The unlike energy parameter $\epsilon_{ij}$ can be derived from a Berthelot-like geometric criterion, where the asymmetry in size is taken into account:

$$\epsilon_{ij} = (1 - k_{ij}) \frac{\sqrt{\sigma^3_{ii}\sigma^3_{jj}}}{\sigma^3_{ij}} \sqrt{\epsilon_{ii}\epsilon_{jj}}. \tag{2.60}$$

Knowledge of the energetic unlike interaction parameter $\epsilon_{ij}$ is critical for an accurate representation of properties of mixtures. It it well known [60, 61] that upon introducing an empirical adjustable parameter $k_{ij}$ for the unlike energy interactions, the quality of the representation is often improved significantly, for both, atomisitic [62–64] and coarse-grained models [65]. The unlike energetic parameter $\epsilon_{ij}$ will be estimated directly from experimental data for the thermodynamic properties of mixtures, whereas $\sigma_{ij}$ and $\lambda_{r,ij}$ are prescribed according to the appropriate combining rule (Eqs. 2.57 and 2.59) unless otherwise indicated.

### 2.3.3 Recent developments of the SAFT-$\gamma$ Mie coarse-grained force fields

The methodology of deriving a CG force field using SAFT-$\gamma$ Mie EoS has been successfully applied for a number of systems to describe the fluid-phase thermophysical properties. One of the first examples of the use of this SAFT-based top-down approach is the development of a CG force field for carbon dioxide, which is represented by a single sphere based on the Mie (23-6.66) potential, without any additional electrostatic interactions [22]. Despite the simplicity of the model, it allows one to accurately reproduce the entire phase envelop with the model using a single set of parameters. The prediction of properties not used in the parameterisation procedure such as the enthalpy of vaporisation, interfacial tension, supercritical density, as well as the second-derivative thermodynamic properties (thermal expansion, isobaric heat capacity, isothermal compressibility, speed of sound and Joule-Thomson coefficient) are also found to be in a very good agreement with experimental data.

The SAFT-$\gamma$ Mie coarse-graining methodology has now been applied to study other systems, such as the green house gases carbon tetrafluoride $CF_4$ and sulfur hexafluoride $SF_6$,

both molecules being represented as single spherical Mie beads, and also to chain molecules such as $n$-decane, and $n-$eicosane [66], and ring molecules, such as benzene represented by a number of tangentially bonded Mie segments. By combining the homonuclear models of benzene and $n$-decane, Avandaño *et al.* [67] extended the methodology to obtain a heteronuclear model for $n-$decylbenzene, in the spirit of a group contribution methodology. More examples and applications of the SAFT-$\gamma$ Mie methodology in the development of the CG force fields for molecular simulations are given in the recent review by Müller and Jackson [68]. This approach is employed throughout the rest of the thesis to develop force fields to represent the thermodynamic and structural properties of a broad variety of systems including water, alkanes, perfluoroalkanes, and aqueous mixtures comprising carbon dioxide, hydrocarbons or alkyl polyoxyethylene glycol non-ionic surfactants.

## 2.4 Molecular simulation

Whereas equations of state, by incorporating a number of approximations to the theory, provide an analytical solution for the statistical mechanical description of the system, molecular simulations can be used to solve the statistical mechanical formulations numerically [18]. Additionally, molecular simulations allow for an insight into thermodynamic, structural and dynamic properties at the molecular level. After the development of the first electronic computers in early 1950s, molecular simulations were employed in connection with the studies of the hydrogen bomb. A broader interest in molecular simulations evolved in 1970, with the ever increasing use of simulations as a test of the approximations made in molecular theories [2]. The two main techniques of performing molecular simulation include Monte Carlo and Molecular Dynamics. With Molecular Dynamics one evaluates the properties of the system as time averages, while Monte Carlo is a stochastic method based on ensemble averages. According to ergodic hypothesis, both techniques performed under the same macroscopic conditions (e.g. same $NVT$), provide the same averages in the thermodynamic limit. In case of Molecular Dynamics, the ergodic hypothesis leads to the conclusion that after a sufficiently long time, the time-averaged properties do not depend on the initial configuration of the simulation. The correspondence of this in Monte Carlos simulations is that every accessible point in a configurational space can be reached in a finite number of trial moves from any other point [69]. A very short overview

over both techniques is presented in the following section.

### 2.4.1 Monte Carlo

Mote Carlo (MC) simulation was firstly proposed by Metropolis in 1953 as an importance-sampling algorithm [70, 71]. Instead of the conventional quadrature scheme, where the value of configurational integral (Eq. 2.31) is calculated from a predetermined set of points, where the Boltzmann factor might be negligible, MC is based on the so-called Metropolis sampling [69]. The idea of Metropolis sampling is to construct an importance weighted random walk through the points in the phase space with a non-negligible Boltzmann factor. Metropolis sampling is conventionally conducted in the following steps: first, an initial configuration is chosen in a way that the system is close to a local minimum in the potential energy, thereby ensuring that the Boltzmann factor is non-negligible (for this purpose, a lattice configuration is usually a suitable choice); thereafter, molecules are moved to different positions, following a random walk, and the new energy of the system is measured, the new move is accepted with probability proportional to the Boltzmann factor $\exp(-\beta\Delta\mathcal{U})$; procedure is repeated until no further change in energy and other properties is observed; after the equilibrium is reached, the thermodynamic properties are computed as instantaneous properties and finally averaged over all possible configurations as

$$< A >= \frac{\int A(\mathbf{r}^N)\exp(-\beta\mathcal{U}(\mathbf{r}^N)}{\int \exp(-\beta\mathcal{U}(\mathbf{r}^N))}. \tag{2.61}$$

A natural choice for the Monte Carlo simulation is the canonical ensemble $NVT$, however, other ensembles can be more convenient depending on the application.

### 2.4.2 Molecular Dynamics

The Molecular Dynamics (MD) technique is comparable to a real experiment. In a real experiment, after preparing a sample and analysing it with a measuring instrument (e.g. thermometer, barometer, etc.), the thermodynamic properties are measured over certain time intervals. Similarly, MD starts with the preparation of an initial configuration, and the initial velocities are assigned according to Maxwell-Boltzmann distribution using a random number generator. Thereafter, particles move according to their intermolecular

forces following Newton's equation of motion:

$$\mathbf{F_i} = m_i \mathbf{a_i} = m_i \frac{d^2\mathbf{r}}{dt^2} = -\frac{\partial \mathbf{U}}{\partial \mathbf{r_i}} \tag{2.62}$$

where $m_i$, $\mathbf{a_i}$ and $\mathbf{r_i}$ are the mass, acceleration, and the position of particle $i$, respectively. In this manner, the new positions and velocities are determined through the forces acting on the particles. After the equilibrium is established, the thermodynamic properties are computed as time averages over the equilibration period. The equation of motion is commonly integrated numerically using the algorithm proposed by Verlet in 1967 [72] or one of its modifications, such as the leap-frog. In both algorithms, the new position of a molecule is obtained from its previous position and the acting force. The details to the Verlet and leap-frog algorithms are given in Appendix B.

Both, MC and MD allow for a computation of equilibrium properties. However only MD mimics the dynamics of the system and therefore can provide information about the kinetics and transport properties. The description of the real systems with both simulation techniques are limited by the computational effort required and the accuracy of underlaying force field.

## 2.5 Force field parameterisation

The parameterisation of a force field is a non-trivial and a cumbersome task. Some of the main requirements on a force field include the accuracy, transferability, and robustness. Unfortunately, there is no generic force field that can reproduce all properties with a unique parameter set. The variety of available force fields is particularly due to the target properties that the force field aims to reproduce, the mathematical functions used in the description of the molecular interactions, and the parameters describing those functions. Commonly, the target properties include experimental geometries (e.g. crystal packing or structure factor) and / or thermophysical properties, such as liquid densities and enthalpies of vaporisation. The force field parameters are usually adjusted to reproduce the target properties either by trial-and-error or in a relative least-squares optimisation procedure. Molecular force fields typically comprise several energetic contributions including the intramolecular interactions such as bond-stretching, angle-bending, and torsion, and the intermolecular interactions, such as dispersion and electrostatics. The quality of the results depends entirely on the accuracy of the force field. The performance of an atomistic

force field is particularly sensitive to a limited number of parameters, mostly the intermolecular interactions and the torsional terms. In this work, our main focus lies on the dispersion interactions. The dispersion interactions are normally described by the intermolecular potentials presented in Section 2.1.2. The LJ potential is now in widespread use in molecular modelling, due to its simple mathematical form and an improved representation of real fluids. The LJ potential function forms the basis for a large number of widely used molecular force fields, such as OPLS [73], TraPPE [74], NERD [75] force fields, and the TIP [76] and SPC [77] models of water, to mention but a few. The popular coarse-grained MARTINI force field [78] is also based on the LJ representation. The LJ potential has been implemented to correlate molecular simulation data with theoretical models as in LJ equation of state of Johnson *et al.* [79] and represents a basis for the SAFT-LJ [50, 51] and the soft-SAFT [52] versions of the SAFT equations of state. However, use of fixed values of the exponents (12-6) can cause a substantial restriction, preventing the model to accurately reproduce all of thermophysical properties. Whereas the attractive exponent originates from the London theory, the repulsive exponent was chosen arbitrarily as a square of the attractive exponent, mainly for the sake of convenience. Even for a simple spherical molecule such as argon (Ar), Lennard-Jones himself [20] determined an optimal repulsive exponent as $14\frac{1}{3}$. It is therefore not surprising that Potoff and Bernard-Brunel [80] found that the Lennard-Jones potential fails to predict simultaneously the saturation densities and the vapour pressures in the united-atom representation of the *n*-alkanes. In this manner, the description of complex fluids requires a more versatile potential that can capture the finer features of the intermolecular interactions. In particular, the Mie potential, which is a generalised form of the LJ potential, has already been successfully implemented for several systems. For instance, Potoff and Bernard-Brunel [80] succeeded in developing a united-atom force field for alkanes and perfluoroalkanes which can reproduce the liquid densities and vapour pressures by using the Mie potentials with varying degrees of softness / hardness. Nielsen *et al.* [81] used the Mie (9-6) potential to describe the densities and surface tensions for alkanes. He *et al.* [82] tested several Mie potentials to develop a model of water which captures the liquid density and surface tension. The Klein group is actually one of the first who largely applied the Mie potential to describe a wide variety of complex systems [81–88]. In our current work, we make use of the flexibility and versatility of the Mie potential and implement it in order to develop our coarse-grained

models.

## 2.6 Coarse-graining

### 2.6.1 Computer simulations at different scales

The choice of modelling technique is crucial in order to adequately represent the physical details that are relevant for the phenomenon of interest and is determined with a level of description of matter. As illustrated in Figure 2.5, matter can be represented at different levels of time and length scales, where each subsequent level denotes a loss of physical details. Passing on information from one level to another is associated with making some approximations and integrating out the unnecessary degrees of freedom, while at the same time assuring that the system retains the relevant physics.

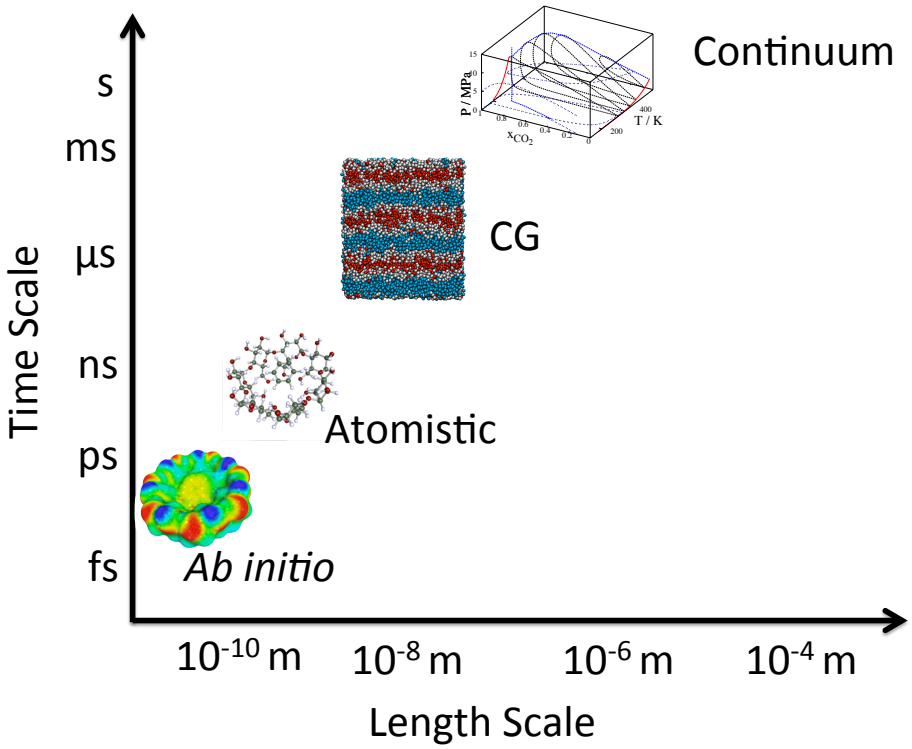



**Figure 2.5:** Computer simulations at different levels of time and length scale.

At the most detailed atomistic level of time and length scale, matter is considered to be made up of fundamental particles, such as electrons and protons, and can be described by the laws of quantum mechanics (QM) by solving the wave equations of Schrödinger for non-relativistic systems and Dirac for relativistic systems. This level of description

is necessary to represent phenomena that involve the motion and the re-arrangement of electrons, as for instance in chemical reaction. The energy of the system is determined solely as a function of nuclear coordinates so that no experimental input is necessary. However, because the many-body problem increases exponentially with the number of particles, this approach is feasible only for very small systems. Larger systems can be treated by applying classical approximations. Hence, at the next level, corresponding to an atomistic representation, the electronic detail is lost and matter is considered to be made up of atoms, which obey the laws of classical statistical mechanics. The atoms interact via pre-defined force fields, usually based on spherically symmetric pair potentials with Coulombic charged sites. This description is adequate for physical processes where the specific electronic distribution is not important, as for instance in phase transitions or transport properties. If the physical process of interest requires the simulation of a large system over longer times, such as self-assembly of amphiphilic systems or protein folding, atomistic description becomes too computationally expensive and therefore the atomistic simulations are not practicable. These type of process can be treated at the less detailed mesoscale level, the level corresponding to coarse-grained (CG) simulations, where matter is considered to be made up of blobs or sites that contain several atoms or small molecules. The CG approach is a central part of our current work and will be described in greater detail in following sections. Finally, at the continuum level, matter is represented as a continuum, and the macroscopic laws, as for example continuity equations, apply. The electronic, atomistic and mesoscale level of detail is now lost. Application of the macroscopic laws requires a lot of experimental input in order to obtain phenomenological coefficients such as diffusion coefficients or chemical reaction rates. The experiments are performed to study one particular problem. This means that the correlation coefficients cannot be directly used to calculate other properties, as it is usually the case of modelling at a higher level of resolution, where the various thermodynamic and structural properties can be assessed in the same simulation run.

We will now focus on the coarse-graining modelling techniques, first by highlighting the advantages compared to the atomistic simulations, which remain the most widely used to date. Modelling of large biological systems and slow biophysical processes, which extend to $\mu$s or ms time scales, can be treated by the CG technique. The main advantage of coarse-graining lies in the immense speed up of the simulation by two to three orders

of magnitude, depending on the chosen level of description. The speed up is due to the incorporation of several atoms into larger beads, reducing the number of interaction sites, which leads to a reduced dimensionality of the system. Additionally, by integrating out fast degrees of freedom, the energy landscape becomes smoother. This allows for implementation of a larger time step, thereby enabling a more effective sampling of phase space and an access to longer times.

Obviously, coarse graining comes at the cost of loosing electronic and atomistic details. Therefore it is crucial to identify the unimportant details and to preserve feature that are essential for the description of the phenomenon on interest. In particular, CG mapping is critical for the accuracy, transferability, robustness, and computational efficiency of the CG model. CG mapping refers to transformation of an atomically detailed structure into a CG representation. Generally, grouping atoms into CG sites is contingent on the chemical structure, so that a CG bead ideally incorporates one or several identical functional groups. In practice, mapping is guided by chemical intuition. An appropriate level of description is crucial for the right balance between the accuracy and the computational efficiency of the model. For instance, Hadley and McCabe [89] have studied different levels of coarse-graining of water molecules by using the K-means algorithm. The authors claim that 4-to-1 mapping is the most reasonable compromise, which is consistent with the common practice to include 3 to 4 heavy atoms into one CG site [78, 85].

The idea of using the effective potentials in order to simplify the representation of matter is not new and originates from the pioneering works by Clausius, Maxwell, and Boltzmann in the field of molecular theories. An early example of coarse-gaining is the seminal work of Levitt and Warshel in 1975 [90], where the authors studied the folding of small proteins by obtaining atomistic potentials from the QM trajectories using MD simulations. Notwithstanding, the parameterisation of interaction potentials and the passing of information from one level of the description of matter to another is not trivial and still remains a challenge. A brief overview on available techniques of coarse-graining and the parameterisation of the pair potential functions by using information at different levels of representation is provided in the next section. A more complete description of the recent coarse-graining techniques and their applications in soft matter are given in some excellent reviews [7, 68, 91–93].

### 2.6.2 Philosophy of coarse-graining

A formal insight into the determination of the effective interactions characterising the CG potentials is given through the statistical mechanical framework. A CG potential can be considered as a many-body potential of mean force (PMF) which is given by

$$U_{CG}(\mathbf{R}) = -k_B T \ln Z(\mathbf{R}), \tag{2.63}$$

where the configurational integral is

$$Z(\mathbf{R}) = C \int \exp(-\beta U_{AA}(\mathbf{r}))\delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}]d\mathbf{r}, \tag{2.64}$$

with $C$ being an arbitrary constant and $\mathbf{M}$ is the "coordinate mapping", illustrated in Figure 2.6. The "coordinate mapping" determines the configuration of the CG model ($\mathbf{R}$) as a linear combination of coordinates of the underlying atomistic model ($\mathbf{r}$) along with constants corresponding to the centre of mass $c_i$:

$$\mathbf{R_i} = \mathbf{M_i}(\mathbf{r}) = \sum_i c_i \mathbf{r_i}. \tag{2.65}$$

The configurational probabilities in the CG ensemble should be equal to those one would observe in the atomistic ensemble:

$$P_{CG}(\mathbf{R}) = p_{AA}(\mathbf{R}). \tag{2.66}$$

The probability of the coarse-grained configuration in the atomistic ensemble is the sum over the probabilities of all atomistic configurations that map into the mapping function (Eq. 2.65)

$$p_{AA}(\mathbf{R}) = \int p_{AA}(\mathbf{R})\delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}]d\mathbf{r}. \tag{2.67}$$

The Dirac delta function selects the configurations in the atomistic space $\mathbf{r}$ and maps it into the CG configuration $\mathbf{R}$. The AA and CG probabilities are obtained from:

$$p_{CG}(\mathbf{R}) = \frac{\exp(-\beta U_{CG}(\mathbf{R}))}{Z_{CG}} \tag{2.68}$$

and

$$p_{AA}(\mathbf{r}) = \frac{\exp(-\beta U_{AA}(\mathbf{r}))}{Z_{AA}}. \tag{2.69}$$

Hence, the CG and the AA probability configurations are equal when

$$\frac{\exp(-\beta U_{CG}(\mathbf{R}))}{Z_{CG}} = \int \frac{\exp(-\beta U_{AA}(\mathbf{r}))}{Z_{AA}} \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}] d\mathbf{r}. \tag{2.70}$$

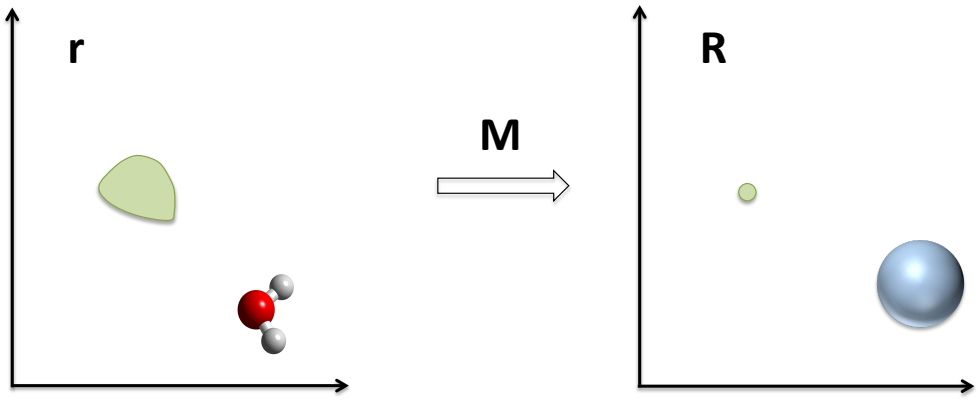



**Figure 2.6:** Mapping of the all-atom configuration space (on the left) into the coarse-grained configuration space (on the right).

In this manner, the PMF is specified in terms of the underlying all-atom (AA) model and the chosen CG mapping:

$$U_{CG}(\mathbf{R}) = -k_B T \ln \left[ C \int \exp(-\beta U_{AA}(\mathbf{r})) \delta[\mathbf{M}(\mathbf{r}) - \mathbf{R}] d\mathbf{r} \right]. \tag{2.71}$$

It can be seen from Eq. 2.71 that the potential of mean force is a function of the thermodynamic state as it depends on the temperature through the Boltzmann factor and the volume (and accordingly density) through the limits of integration. This leads to problems in the transferability of the potential, which imply that the CG force field derived at one state point does not necessarily perform well in a different state point and might need to be re-developed. Another common problem in coarse-graining is known as representability, which implies that one coarse-grained potential cannot accurately reproduce all of the properties of the system.

There is a fundamental difference if the CG potential is parameterised to reproduce structural or thermodynamic properties. By combining Eqs. 2.24 and 2.9, the three-body Hamiltonian can be expressed as a function of the pair and triplet potentials

$$\mathcal{H} = \mathcal{K} + \sum_{i<j} u^{(2)}(\mathbf{r_i}, \mathbf{r_j}) + \sum_{i<j<k} u^{(3)}(\mathbf{r_i}, \mathbf{r_j}, \mathbf{r_k}) \tag{2.72}$$

By choosing the thermodynamic route, the internal energy can be expressed as

$$U(N,V,T) = \frac{1}{2}\rho^2 \int \mathbf{dr_1} \int \mathbf{dr_2} g_{12} u_{12} + \frac{1}{6}\rho^3 \int \mathbf{dr_1} \int \mathbf{dr_2} \int \mathbf{dr_3} g_{123} u_{123} \tag{2.73}$$

or in terms of the effective potential [94–96] as

$$U(N,V,T) = \frac{1}{2}\rho^2 \int \mathbf{dr_1} \int \mathbf{dr_2} g_{12} u_u^{eff}(\mathbf{r_{12}}, \rho), \tag{2.74}$$

where the effective potential is defined as

$$u_u^{eff}(\mathbf{r_{ij}}, \rho) = u_{12}(\mathbf{r_{ij}}) + \delta u_u(\mathbf{r_{ij}}, \rho) \tag{2.75}$$

$$\delta u_u(\mathbf{r_{ij}}, \rho) = \frac{1}{3}\rho \int \mathbf{dr_3} g_{13} g_{23} G_{123} u_{123} \tag{2.76}$$

$$g_{123} = g_{12} g_{13} g_{23} G_{123}. \tag{2.77}$$

Here, $G_{123}$ represents the correction to the Kirkwood superposition approximation [97] for the triplet radial distribution function $g_{123}$.

If the so-called structural route is chosen to parameterise the CG effective potential, the effective pair potential can be expressed as [95, 98, 99]

$$u_g^{eff}(\mathbf{r_{12}}, \rho) = u_{12}(\mathbf{r_{12}}) + \delta u_g(\mathbf{r_{12}}, \rho), \tag{2.78}$$

where

$$\delta u_g(\mathbf{r_{12}}, \rho) = \rho \int \mathbf{dr_3} (1 - \exp(-\beta u_{123})) g_{13} g_{23} \tag{2.79}$$

The potential $u_g^{eff}(r_{12}, \rho)$, which is able to reproduce the given radial distribution function, is unique according to the Henderson uniqueness theorem [100]. Eq. 2.79 differs from Eq. 2.76 by factor 3,

$$\frac{\delta u_u(\mathbf{r_{12}}, \rho)}{\delta u_g(\mathbf{r_{12}}, \rho)} \approx \frac{1}{3}, \tag{2.80}$$

implying that there is no unique way to represent averages over many-body interactions by an effective potential, *i.e.*, the effective potential derived by the structural route is not equal to the potential derived by the thermodynamic route. In other words, an effective potential cannot be used to reproduce the thermodynamic and structural properties simultaneously. Hence, the functional form of the CG potential will differ depending on the target properties and the way the potential has beed derived. Some of the commonly used techniques to derive the CG potential are described in the next section. This effect becomes particularly noticeable for the systems with pronounced many-body interactions, such as dense systems and systems with strong polar and directional interactions.

### 2.6.3 Bottom-up methods

A bottom up CG approach is conducted on the basis of a more detailed model, usually an *ab initio* quantum-mechanical description or all-atom classical model. An obvious advantage of the approach is that the information of the atomistic model itself is required, but, the accuracy of the atomistic model is essential in order to be able to develop a reliable CG force field.

**Structure based methods**

The target of a structure-based approach is to reproduce the correlation function of the underlying detailed atomistic model. This approach relies on the Henderson uniqueness theorem [100], which states that only one pair potential can exactly reproduce a given RDF. In practice, a number of potentials can reproduce the target RDF within acceptable deviation. One can increase the correspondence between the derived coarse-grained model and the target atomistic model by including some of the thermodynamic macroscopic properties in the parameter estimation procedure, such as pressure [101] or activity coefficients [102]. Due to the problem of representability, the incorporation of the thermodynamic properties in the objective function for the structure-derived potential is neither expected to improve the representation of the distribution function, nor to accurately capture thermodynamic properties. The pair potential derived in this way would be a compromise of the simultaneous representations of the structural and thermodynamic properties.

For systems approaching zero-density limit, the intermolecular interactions are mostly influenced by the pair interactions and the many-body effects are far less pronounced than for the dense systems. Therefore, in dilute systems, the CG potential can be obtained from

a direct inversion of the corresponding atomistic correlation function or Direct Boltzmann Inversion (DBI)

$$u_{CG}(r) = -k_B T \ln g_{AA}(r). \tag{2.81}$$

Here, the target $g_{AA}$ is computed for the centre of mass of a corresponding CG bead; $u_{CG}(r)$ is the CG potential of mean force, which at a low density limit can exactly reproduce the structure of the underlying atomistic model. A clear advantage of this procedure is that it is non-iterative. It has been successfully used for derivation of the non-bonded intermolecular potentials in dilute polymer solutions [103] and peptides [104], but mostly to obtain the intramolecular (bonded) potentials [105]. Due to the stiffness of bonded degrees of freedom, the bonded potential can be assumed independent and can therefore be decoupled from the non-bonded potential, thereby leading to a commonly accepted assumption that the total potential is a sum over the bonded and non-bonded interactions. Furthermore, it can be assumed that the contributions of the intramolecular potentials such as bonds, angles and dihedral bonded interactions are independent, so that the bonded potential itself can be decomposed into separate contributions. However, the independence of these degrees of freedom though often appropriate is an approximation and the validity of the assumption depends strongly on the system and the chosen CG scheme. When the degrees of freedom are not independent, the potentials can be refined in the iterative manner until the CG potential converges and reproduces the distribution functions of the target AA model.

If one is interested in the non-bonded potentials of a high-density system, the properties are largely influenced by the many-body interactions so that the potentials obtained from direct Boltzmann inversion can not reproduce the target atomistic correlation function. In order to capture the target correlation function accurately, an iterative procedure can be used, which was proposed by Müller-Plathe and co-workers [106] and is now known as Iterative Boltzmann Inversion (IBI). In this technique, the following steps are undertaken: one starts with an initial guess for the CG potential; usually, this is the potential obtained from DBI, cf. Eq. 2.81; a new simulation is performed based on the computed $u^i_{CG}$ and the corresponding RDF is measured; the potential is then updated based on the ratio of

the new measured $g_{CG}(r)$ and the target $g_{AA}(r)$ according to

$$u_{CG}^{i+1}(r) = u_{CG}^{i}(r) - k_B T \ln \frac{g_{AA}(r)}{g_{CG}(r)}; \tag{2.82}$$

this step is repeated until the full convergence of $u_{CG}$. Faster convergence can be achieved by using the initial guess of improved quality. In this case, the potential can be obtained by solving the inverted Ornstein-Zernike integral equation with the Percus-Yevick closure (OZPY$^{-1}$) [107]. Ornstein-Zernike (OZ) integral equation, which is given as

$$h(r_{12}) = c(r_{12}) + \rho \int c(r_{13}) h(r_{23}) dr_3, \tag{2.83}$$

relates the pair correlation function $h = g - 1$ to the direct correlation function $c$ as a convolution integral. The Percus-Yevick (PY) closure, which is given as

$$c(r_{12}) = (1 - \exp(\beta u_2(r_{12}))\, g(r_{12}), \tag{2.84}$$

is then used to relate $c$ and $g$ to the pair potential $u_2$. By eliminating $c(r_{12})$, the pair potential $u_2$ can be derived in terms of the target $g_{AA}(r)$ using an iterative numerical scheme. Although the Ornstein-Zernike equation is exact, the Percus-Yevick closure (and other such closures) contains some approximations so that the computed $u_2$ does not generate the exact $g_{AA}(r)$ and therefore the effective potential $u_{CG}^{eff}$.

The IBI has been applied to a wide variety of complex fluids including polymers [108], proteins [109], peptides [110], and carbohydrates [111]. Rather than being represented as an analytical function, the effective potential $u_{CG}^{eff}$ is commonly based on flexible basis functions, e.g splines, in order to more accurately reproduce the PMF. The potential function can be described with an interpolating spline or be merely tabulated. The popularity of the method arises from capability of treating the effective pair potential $u_{CG}^{eff}$ at each distance $r$ independently as one only considers the correspondence between the potential and RDF at a given distance. In some cases, however, this approach can lead to the convergence difficulties due to complex couplings between the interactions or to the relative insensitivities of RDF to the change in the CG pair potential.

Similarly, in inverse Monte Carlo (IMC), pioneered by Lyubartsev and Laaksonen [112], an iterative procedure is applied to reproduce the atomistic RDF. In contrast to a local IBI method, the IMC is non-local, i.e. by updating the pair potential at the given distance,

one considers the RDF at all distances. In this manner, the number of the iterative steps can be reduced, however each step is more computationally demanding compared to a simple IBI. The IMC approach has been applied to various complex systems including amino acid solutions [113], lipid layer membranes [114], and even DNA [115].

**Relative entropy**

The IBI and IMC schemes are structure-based approaches and can be considered as particular cases of a more general framework proposed by Shell and co-workers [116, 117], referred to as the relative entropy approach. In this approach, the relative entropy, which can be considered as amount of information lost in the coarse-graining procedure, is obtained by minimising the discrepancy between the atomistic and coarse-grained distribution functions.

$$S_{rel} = \int p_{AA}(\mathbf{r}) \ln \frac{p_{AA}(\mathbf{r})}{p_{CG}(\mathbf{M}(\mathbf{r}))} d\mathbf{r} + S_{map}, \tag{2.85}$$

where $p_{AA}$ and $p_{CG}$ are probability distributions in AA and CG ensembles. The term $S_{map}$ is a measure of the degeneracy of mapping and is given as a logarithm of the average number of AA configurations that can be mapped into a singe CG configuration. $S_{rel}$ is always positive or zero. A value of $S_{rel} = 0$ implies that the CG model perfectly reproduces the target model. A high value of $S_{rel}$ indicates a poor correspondence between the CG and AA models. If $S_{rel}$ is at a minimum, the CG model optimally (albeit not exactly) represents the underlying atomistic model.

In the canonical ensemble, the coarse-grained and the atomistic probability distributions are given by Eqs. 2.68 and 2.69. The relative entropy is a functional of $U_{CG}(\mathbf{R})$ and can be expressed in terms of configurational integral as

$$S_{rel} = \beta \int p_{AA}(\mathbf{r})[U_{CG}(\mathbf{M}(\mathbf{r})) - U_{AA}(\mathbf{r})] d\mathbf{r} + \ln \int e^{-\beta U_{CG}(\mathbf{R})} d\mathbf{R} - \ln \int e^{-\beta U_{AA}(\mathbf{r})} d\mathbf{r} + S_{map}. \tag{2.86}$$

The optimal $U_{CG}$ is the one that functionally minimises the relative entropy

$$0 = \frac{\delta S_{rel}[U_{CG}(\mathbf{R})]}{\delta U_{CG}(\mathbf{R})} = \beta \int p_{AA}((\mathbf{r}))\delta[(\mathbf{M}(\mathbf{r}) - \mathbf{R}')] dr - p_{CG}((\mathbf{R}')) \tag{2.87}$$

so that

$$p_{AA}(\mathbf{R}') = p_{CG}(\mathbf{R}'), \tag{2.88}$$

where $\mathbf{R}'$ is a set of coarse-grained coordinates used to find the variation of the relative entropy with respect to $U_{CG}(R)$. Eq. 2.88 implies that the configurational probability distributions in the AA ensemble are the same as the configurational probability distributions in the CG ensemble when $U_{CG}$ is not constrained to any particular form, e.g. not pairwise additive. Therefore, the unconstrained optimal $U_{CG}$ can rigorously reproduce the free energy landscape in the AA ensemble and is given by Eq. 2.71. In case $U_{CG}$ is constrained to be of a pairwise additive form, $S_{rel}$ will be higher than the optimal minimum value. However, one still aims to find $U_{CG}$ of a given functional form which can reproduce the target probabilities $p_{AA}(\mathbf{R})$ as accurately as possible.

**Force-matching**

The force-matching method was first proposed by Ercolessi and Adams in 1994 [118], who derived classical atomistic potentials from *ab initio* forces and trajectories. Based on a similar idea, Voth and co-workers [119] developed CG potentials by representing the forces and trajectories of classical atomistic simulations. The target of this method, which is referred to as Multi Scale Coarse Graining (FM-MSCG), is to match the average force on CG sites to that of the AA system. CG force field is given in an analytical form, which depends on the set of $M$ parameters $g_1, ..., g_M$. The objective function is minimised with a least-square procedure

$$\min \sum_{\mathrm{l=1}}^{\mathrm{L}} \sum_{\mathrm{i=1}}^{\mathrm{N}} |\mathbf{F}_{\mathrm{il}}^{\mathrm{AA}} - \mathbf{F}_{\mathrm{il}}^{\mathrm{CG}}(\mathrm{g_1}, .. \mathrm{g_M})|^2, \tag{2.89}$$

where $\mathbf{F}_{il}^{AA}$ is a reference AA force and $\mathbf{F}_{il}^{CG}$ is a force predicted using the analytical CG form, which acts on the $i$th atom in the $l$th atomic configuration; $N$ is the number of atoms in the atomic configuration, and $L$ is the total number of atomic configurations used in the regression. The first sum in Eq. 2.89 refers to the entire trajectory of configurations $L$ of the AA ensemble, the second sum is over all CG atoms $N$. Rather than solving one complex non-linear problem, Voth and co-workers proposed averaging the solutions of many small overdetermined linear equations. In contrast to the structure based approaches, which include iterative simulations for the nonlinear optimisation procedure,

the FM- MSCG method is non-iterative and the linear optimisation problem is solved directly. The FM-MSCG method exhibits similarities with the relative entropy method: whereas in the relative entropy approach one minimises the functional of the CG potential, in the FM-MSCG approach one minimises the squared gradient of the CG potential, since the force is defined as a gradient of the many-body PMF ($F_{il}^{CG} = -\frac{\partial U_{CG}(\mathbf{R})}{\partial \mathbf{R_{il}}}$). If $U_{CG}(\mathbf{R})$ is unconstrained to a particular functional form, $U_{CG}(\mathbf{R})$ obtained from the relative entropy method should be the same as $U_{CG}(\mathbf{R})$ obtained from the FM-MSCG method. However, using an effective pair potential, which cannot directly capture the multi-body interactions, leads to a different description with both methods.

Noid and co-workers [120–122] have shown that the FM-MSCG method provides an optimal approximation of the many-body PMF. The functional in Eq. 2.89 determines a unique minimum, so that the exact many-body PMF can be reconstructed with the variational principle within an arbitrary additive constant. Furthermore, by approximating the many-body CG force field by the pair potential, Mullinax and Noid [123] have demonstrated that the relations used to implement the FM-MSCG method are formally equivalent to the well-defined Yvon-Born-Green (YBG) equation for the liquid state [124], which can be expressed as

$$-\frac{\partial g(\mathbf{R_1}, \mathbf{R_2})}{\partial \mathbf{R_1}} - \beta g(\mathbf{R_1}, \mathbf{R_2}) F(\mathbf{R_1}, \mathbf{R_2}) = \beta \rho \int F(\mathbf{R_1}, \mathbf{R_3}) g_3(\mathbf{R_1}, \mathbf{R_2}, \mathbf{R_3}) d\mathbf{R_3} \quad (2.90)$$

The YBG equation is exact and it relates the two- and three-body correlation functions to the pair force field. Therefore, the FM-MSCG method explicitly incorporates the information about the the two- and the three-body correlations, by contrast to the conventional methods that are based just on two-body correlations. The first term in Eq. 2.90 describes the gradient of the pair distribution function. Since the PMF is related to the pair distribution function by Eq. 2.81 and the force is defined as a gradient of the PMF (see above), the YBG equation provides an optimal approximation to the exact many-body CG PMF. In this manner, The FM-MSCG force matching can be directly undertaken with the structural correlation functions without an explicit knowledge of atomistic forces. The FM-MSCG technique has now been successfully applied for a wide range of systems, including lipid bilayers [125], carbohydrates [126, 127], ionic liquids [128], and peptides [129].

### 2.6.4 Top-down methods

In contrast to the bottom-up approaches, where the CG potentials are derived from a model at a higher level of resolution, with top-down approaches the intermolecular parameters are obtained from macroscopic experimental data. Thermodynamic properties are commonly used to parameterise the CG potentials, one exception being the native structure-based models, applied to the study of the proteins.

**Native structure-based models**

The native structure-based models are used to estimate the intermolecular potential parameters from the experimental three dimensional folded protein structures, which are available from protein data bank (PDB). This approach distinguishes between network models, native-centric models (Gõ-models), and knowledge-based models. The network and native-centric models are parameterised to reproduce the structure of a particular protein. The derived potential can be used to study other properties of the given protein, such as building of protein complex formations, conformational transitions, protein docking, and folding. By contrast, in the case of knowledge-based models, the structures of multiple non-homologous proteins are used simultaneously to derive a transferable parameter set which allows one to predict the structure of unknown proteins. The iterative Boltzmann inversion method is commonly applied to reproduce the many-body PMF. This type of optimisation strategy was pioneered by Maiorov and Crippen *et al.* [130], where linear programming methods were used to derive potentials that allow one to reproduce the known PDB structure for each protein from the training set.

**Toy models**

A wide variety of simple toy models have been employed in the physical and biological sciences to simplify the nature of the physical interactions in order to aid the understanding of the behaviour of the system in terms of the basic features of the interactions. A good example is work in the area of CG models for amphiphilic systems by Smit *et al.* [131]. Instead of trying to quantitatively reproduce the structural properties of given systems, the authors aimed to describe physical phenomena qualitatively by using a very simple model. Their aqueous surfactant system consisted of three types of chemical moiety: the head ($h$) and tail ($t$) of the surfactant, and the water particles ($w$). All particles are assumed to interact via a truncated and shifted LJ potential and have the same dis-

tance parameter $\sigma$ and energy parameter $\epsilon$, the only difference being the cutoff radius $R_{ij}^c$. Interactions between $w-w$, $w-h$, and $t-t$ are truncated at $R_{ij}^c = 2.5\sigma$, whereas the cutoff radius between $w-t$, $h-h$ and $t-h$ is $R_{ij}^c$=$2^{1/6}\sigma$, essentially corresponding to soft repulsive interactions. Using this very simple model, it was possible to observe not only self-assembly into micelles with a characteristic size distribution, but also the dynamical processes of a single amphiphile entering a micelle, an amphiphile leaving micelle, the fusion of two micelles and the slow break-up of the micelle. Self-assembly into the aggregates, which is promoted by the anisotropy in the molecular potential, could also be observed by a primitive model represented by a flexible trimer chain interacting via simple square-well and hard-sphere potentials [132]. These so-called "toy" models, while relatively simple compared to real aqueous amphiphiles, provide a physical representation of the free energy of the system as an explicit account is taken of the entropic (free volume, packing, connectivity) and the energetic (dispersion, non-covalent bonding) contributions. Complex mesophase behaviour of amphiphilic molecules can be described with toy models of this type. For instance, Crane *et al.* [133] have examined the global phase behaviour of polyphilic tapered dendrons by using a simple planar arranged CG model. Three micro-segregated thermotropic liquid crystalline phases such as a hexagonal columnar, a supramolecular sphere body-centred cubic and a supramolecular sphere fluid mesogenic phases were detected. The rich liquid crystalline-phase behaviour could be studied by using simple hard-rod models, initiated by the pioneering work by Onsager [134]. The stability of liquid-crystalline phases was investigated by incorporating additional interactions, including the anisotropic dispersion [135], chirality [136], and various positions of the dipole moment [137–139].

**Martini force field**

The most commonly used CG model in the study of biological systems is now probably the Martini CG model [78, 140]. The intramolecular interactions are obtained from the structural data of the atomistic system by using the standard IBI method. The intermolecular interactions are however parameterised within the top-down approach, aiming to reproduce the difference in free energy between the polar and apolar phases. The Martini force field has gained a wide popularity due to its simplicity, high computational efficiency, and transferability which allows for a broad range of possible applications. Computational efficiency is attributed to the relatively high level of coarse-graining : 4-to-1, which implies

four heavy atoms per bead in the chain molecules to represent the solute, and four water molecules per bead to represent the solvent. The simplicity of the model is ascribed to the simple functional form of the intermolecular potential and the relatively small number of parameters. The CG sites are characterised as four main types, depending on the increasing degree of polarity: non-polar (N), apolar (C), polar (P) and charged (Q). Within the main types, one can also distinguish several subtypes depending on the hydrogen-bond capabilities (*d* for donor, *a* for acceptor, *da* for both, and 0 for none) or the degree of polarity (from 1 for a low polarity to 5 for a high polarity). The intermolecular interactions are described with the conventional Lennard-Jones (12-6) potential; the size parameter for most particle types is given a universal value of $\sigma = 0.47$ nm, apart from ring-like molecules, where $\sigma = 0.43$ nm. The range of energy parameters $\epsilon$ lie between $\epsilon = 5.6$ kJ/mol for strong polar interactions and $\epsilon = 2.0$ kJ/mol for weak interactions between polar and apolar components, thereby mimicking the hydrophobic effect. The properties used in the parameterisation procedure include the free energy of hydration, free energy of vaporisation, and partitioning free energy between water and organic phases of different polarities (from hexadecane, through chloroform to octanol) for 18 different types of CG particle. The hydration and vaporisation energies can be reproduced only qualitatively as they are largely overestimated by the model. Matching the partitioning free energy is however crucial for an accurate description of the unlike interactions between CG particles. The Martini force field is reported to reproduce the partitioning free energy $\Delta G_{oil/aq}$, which is obtained from the equilibrium densities of CG particles in both organic $\rho_{oil}$ and aqueous $\rho_{aq}$ liquid phases as

$$\Delta G_{oil/aq} \approx k_B T \ln \left( \frac{\rho_{oil}}{\rho_{aq}} \right), \tag{2.91}$$

within 2 $k_B T$ compared to experimental data.

As mentioned earlier, the Martini force field is developed by assigning the main emphasis to simplicity and transferability. The ideology is similar to that of the toy models, as the aim is for a broader range of applications rather than focusing on an accurate reproduction of structural details at a particular state point for a specific system; as a consequence there is no need to re-parametrise the model each time. Despite this minimalistic approach, one can use the Martini force field to "semi-quantitatively" reproduce a number of structural (e.g. the surface area per lipid group [141], the angle subtended by membrane spanning

helices [142]), dynamical (e.g. the water transmembrane permeation rate [141]), and thermodynamic properties for a wide range of systems, including lipids [141, 143–145], surfactants [146–149], carbohydrates [150], proteins [105, 142, 151–153], various polymers [154–156], DNA [157], fullerene [158], and more.

The model however is known to exhibit some limitations. Because the force field is explicitly parameterised for the liquid phase, the properties of the vapour and solid phases are not expected to be accurate: the water in the vapour state is found to exhibit some unphysical clustering due to the chosen level of coarse-graining, and the crystal packing does not correspond to the correct morphology. The range of the aqueous liquid phase is rather narrow since the melting point of water is overestimated and the boiling point is underestimated [78, 159]. To overcome this premature freezing, antifreeze particles can be added to the model solutions. As a consequences of these deficiencies the temperature range of the Martini force field is limited to the range between 270 and 330 K.

**Models of Klein and co-workers**

Another popular top-down approach has been championed by the group of Klein and successfully applied to develop models for a wide range of systems, including ionic liquids [87], amino acids [88], phospholipids [83, 84], and surfactants [85, 86].

The CG segment intermolecular interactions are described with the versatile Mie potential. The most common choice of exponents include the Mie (9-6), (12-4), and LJ (12-6) potentials. The energy and the size parameters of the various chemical groups are determined by considering liquid densities and surface tensions at ambient temperature. The unlike interaction parameters are usually refined by considering the solvation free energy. Here we will only mention the surfactant model which was used to study the micellisation properties of polyoxyethylene glycol surfactants with one and two hydrophobic tails [85, 86]. The CG surfactants were found to spontaneously organise into micelles, and hexagonal and lamellar lyotropic liquid crystalline structures at the corresponding concentration and thermodynamic conditions for those phases [86]. The dependence of the surface pressure on the surface area for surfactant monolayer at the air-water interface was reproduced in agreement with experiment [86]. The dynamics of budding and fission of micelles from surfactant monolayers and the repartitioning of surfactants at the oil-water interface were observed [160]. The applicability of the model was also extended by introducing a temperature dependence to study temperature effects on the micellisation behaviour such as

critical micelle concentrations (CMC) and aggregate sizes [161]. Surfactants of different chain lengths were studied and a good agreement with experimental data was achieved for the short hydrocarbon chains, while the description for the longer chains was not very satisfactory [161].

**Other promising models**

A wide variety of different CG models has been parameterised using the top-down approach by matching various structural and thermodynamic macroscopic properties. To mention but a few examples, Orsi *et al.* [162] developed the ELBA CG model, where lipids are parameterised to match volume and area per lipid, average segment tail order parameter, spontaneous curvature, and the dipole moment. In the case of aqueous non-ionic surfactants solutions, Vishnyakov *et al.* [163] obtained a CG model for the Dissipative Particle Dynamics simulations by adjusting the parameters to reproduce the infinite dilution activity coefficients.

Notwithstanding, the inconvenience of developing semi-empirical CG models of this type stems from the heuristic and time-consuming parameter refinement, conducted iteratively by adjusting the parameters based on the quality of the output of simulations to the target macroscopic properties [82]. As well as requiring a significant computational effort, the procedure still does not guarantee that the parameter set corresponds to an optimal solution. On the other hand, the use of an analytical molecular-level framework such as a high quality equation of state allows for an effective description of a large number of macroscopic thermodynamic states and exploring a wide parameter space.

**CG models from EoS**

A molecular model can be parameterised by using an equation of state if the EoS is expressed explicitly in terms of well-defined intermolecular potential and Hamiltonian. The idea to develop an intermolecular potential model using a closed-form analytical equation of state was pioneered by Müller and Gubbins in 1995 [164]. They used the statistical associating fluid theory [44], a reliable molecular-based equation of state for associating fluids, as a basis to develop a model of water based on the Stockmayer reference fluid (LJ fluid with dipolar contribution). The residual Helmholtz free energy of the fluid was decomposed into separate contributions due to the reference Lennard-Jones interactions, dipolar contribution, and the contribution due to hydrogen bonding, which was modelled in terms of the association sites according to Wertheim's first-order thermodynamic

perturbation theory (TPT1) [39–42]. The dipolar contribution was described by Padé approxant [165, 166] in order to overcome slow convergence for the large values of the dipole moment that characterises the water molecule :

$$\frac{A^{DD}}{NkT} = \frac{A_2^{DD}/NkT}{1 - A_3^{DD}/A_2^{DD}}. \tag{2.92}$$

The second- ($A_2^{DD}$) and the third-order ($A_3^{DD}$) terms which include the integrals over the two- and three-body correlation functions, respectively, were obtained from molecular simulations. The model was parameterised to reproduce vapour-liquid equilibrium properties. The comparison of the results by the EoS and MC simulation revealed a good agreement for the vapour pressure, energy of vapourisation, liquid density and degree of association in the saturated liquid.

Vrabec and co-workers [167–170] have employed a similar procedure for small dipolar and quadrupolar molecules, by correlating the description of a molecular-based equation of state to a large set of simulated volumetric and phase equilibrium data. In this case the reference fluid is represented as a two-center dimer formed from two fused LJ segments with an embedded central point dipole (2CLJD) or quadrupole (2CLJQ). Both LJ segments are located at a distance $L$ from each other (corresponding to the molecular bond length) and the point diplol $\mu$ or quadrupole $Q$ site is positioned in the geometric center and aligned along the molecular axis. Hence, the pair potential is divided into contributions due to the LJ dispersion forces and the dipolar/quadrupolar forces

$$u(\mathbf{r_{ij}}, \omega_\mathbf{i}, \omega_\mathbf{j}) = u^{2CLJ}(\mathbf{r_{ij}}, \omega_\mathbf{i}, \omega_\mathbf{j}) + u^{DD}(\mathbf{r_{ij}}, \omega_\mathbf{i}, \omega_\mathbf{j}), \tag{2.93}$$

where $\mathbf{r_{ij}}$ is the vector from one molecular centre to the other, and $\omega_\mathbf{i}$ denotes the orientation, which in this case includes azimuthal $\phi_i$ and polar $\theta_i$ angles.

The two-sites LJ term can be expressed as

$$u^{2CLJ}(\mathbf{r_{ij}}, \omega_\mathbf{i}, \omega_\mathbf{j}) = \sum_{\alpha=1}^{2}\sum_{\beta=1}^{2} 4\epsilon \left[ \left( \frac{\sigma}{\mathbf{r}_{\alpha\beta}} \right)^{12} - \left( \frac{\sigma}{\mathbf{r}_{\alpha\beta}} \right)^{6} \right], \tag{2.94}$$

where $r_{\alpha\beta}$ is the distance between two LJ sites of different molecules $i$ and $j$.

The dipolar $u^{DD}$ and / or quadrupolar $u^{QQ}$ contributions are obtained from the corresponding multipolar relations as

$$u^{DD}(\mathbf{r_{ij}}, \omega_\mathbf{i}, \omega_\mathbf{j}) = -\frac{\mu_i \mu_j}{|\mathbf{r_{ij}}|^3} \times (2\cos\theta_i \cos_j - \sin\theta_i \sin\theta_j \cos(\phi_j - \phi_i)) \tag{2.95}$$

and

$$u^{QQ}(\mathbf{r_{ij}}, \omega_\mathbf{i}, \omega_\mathbf{j}) = \frac{3}{4}\frac{Q_i Q_j}{|\mathbf{r_{ij}}|^\mathbf{5}}\psi(\omega_\mathbf{i}, \omega_\mathbf{j}) \tag{2.96}$$

where $\psi$ is a function of angles [171].

The Helmholtz free energy can be expressed within the framework of perturbation theory:

$$\frac{A^{res}}{NkT} = \frac{A^{2CLJ}}{NkT} + \frac{A^{DD/QQ}}{NkT}, \tag{2.97}$$

where $A^{DD/QQ}$ is described by Padé approxant (Eq. 2.92), and the integrals in terms $A_2^{DD}$ and $A_3^{DD}$ are correlations obtained from molecular simulations for the spherical Stockmayer fluid *et al.* [169].

The parameters of the fused 2CLJ model are "translated " from the tangent-sphere model using the empirical correlation determined by Gross [172], which relates the number of tangential segments $m$ to molecular elongation $L^* = L/\sigma$, with $L$ being the bond length. The model is therefore based on four adjustable parameters: $\epsilon$, $\sigma$, $L^*$, and dipole or quadrupole moment $\mu$ or $QQ$. The method has been successfully applied to reproduce the vapour-liquid equilibrium properties of a large number of real pure components, binary and ternary mixtures [170]. The limitations of the model are that only small non-polar, dipolar and quadrupole molecules can be described, where the dipole moment can stretch only over two segments.

The group of Elliot [173–176] employed the top-down framework by combining discontinuous molecular dynamics (DMD) with the high-temperature perturbation theory of Barker and Henderson [177, 178]. In their original work the molecular model was described in terms of multistep square-well potential function (typically based on four square-well potentials with same size parameter and different well-depths and ranges). The method is referred to as Step Potentials for Equilibria And Discontinuous Molecular Dynamics (SPEADMD). The multistep potential is meant to mimic a more realistic continuous Lennard-Jones potential and is an improved representation of realistic interactions compared to a single square-well potential. The computational performance of the discontinuous potential is up to 2 orders of magnitude more efficient compared to the continuous potential. The Helmholtz free energy is formulated as contributions from the reference (repulsive) fluid and the first-order (attractive) and the second-order (fluctuation) perturbations terms. The perturbation contributions are computed from the DMD simulations

of the reference fluid and are consequently incorporated back in the TPT. The accuracy of the method is limited, as the density of the fluid is found to deviate approximately 5 % from the experimental values. Since the method involves only simulation of the reference fluid, the properties of the full potential are not accurately reproduced due to the approximations in the theory. Finally, the implementation of the discontinuous potential in Molecular Dynamics simulation is not trivial and requires the development of specific algorithms to solve the equations of motion for this type of potential.

Recently, the group of Gross [179] has employed the PC SAFT EoS [49] to obtain the force field parameters of $n-$alkanes (from $n$-ethane to $n$-decane) based on LJ segments. However, in the PC SAFT description of the dispersion term, the coefficients of the free energy expansion rather than the intermolecular potential parameters are adjusted to the VLE properties of alkanes. In this manner, the link to the intermolecular potential is not well-defined. Hence, the PC SAFT parameters have to be "translated" empirically into the molecular force field parameters, so that the parameterisation can be performed only in an iterative manner involving a series of simulations. The TraPPE parameters were taken as an initial guess, and based on these parameters, molecular simulation was performed to compute the target properties: liquid densities, vapour pressures and enthalpies of vaporisation. The objective function was then minimised in terms of discrepancy between the molecular simulation and experimental data. If the resulting parameters of the model produced the discrepancy higher than the tolerance, they were "translated" into the PC SAFT parameter set in order to get closer to the molecular simulation results with the EoS. The objective function was then minimised in terms of discrepancy between the target properties calculated by the PC-SAFT EoS and the experimental data in order to obtain the new best guess. The iteration procedure was found to converge within approximately three simulation runs, at least in the case of the $n-$alkanes considered.

By contrast, the SAFT-$\gamma$ Mie formalism describes the Helmholtz free energy of the system explicitly in terms of the intermolecular interaction potential, thereby retaining the direct link between underlying coarse-grained force field and the macroscopic properties of the system [24].

# Chapter 3

# Chain molecules

## 3.1 Chapter overview

The CG molecular models of chain molecules of various chemical nature are presented within the framework of the SAFT-$\gamma$ EoS. Homonuclear tangential models for homologous series of alkanes, perfluoroalkanes, and ethers are developed. Heteronuclear models are proposed for semifluorinated alkanes and non-ionic surfactants by implementing the functional groups obtained from the homonuclear models. The transferability of the CG models is verified by testing molecules of a different architecture and in different environments, such as highly non-ideal mixtures or pure components comprising mutually repellent groups. The subsequently performed molecular simulations are based on the derived set of intermolecular potentials and intramolecular bond and angle potentials derived from the structure-based bottom-up approach. The MD simulation reproduce the experimental saturation densities, vapour pressures, and the surface tensions with a high accuracy.

## 3.2 Introduction

A large number of molecular force fields for common fluids has been proposed in the last years. Of particular mention are the homologous series of alkanes since they consist of a well-defined, well-documented and simple (but not trivial) study case. CHARMM (Chemistry at HARvard Molecular Mechanics) [180] is one of the first force fields to deal explicitly with simple molecules including alkanes. It was developed by Karplus and co-

workers in early eighties and several versions have been proposed since then, including the fluctuating charge description, all-atoms and united atoms versions. Another well-known example is the OPLS (Optimised Potentials for Liquid Simulation) force field, developed by the group of Jorgensen in the mid eighties [73] and it includes all-atom and united-atom versions. Probably the most common force fields for the description of $n-$alkanes by the united-atom representation are the SKS (Siepmann, Karaborni, Smit) [181], TraPPE (Transferable Potentials for Phase Equilibria) [74], and NERD (Nath, Escobedo, de Pablo, Revised) [75] force fields, parameterised to reproduce the vapour-liquid equilibrium properties. All of the named force fields have their strength and weaknesses, which is a subject of a number of studies [182–184]. Whereas the most of the models can accurately reproduce the liquid densities, they fail to describe the vapour pressures. One of the common features of the named force fields is the description of the dispersion forces, based on the Lennard-Jones intermolecular potential. The Lennard-Jones potential has however been shown unable to simultaneously capture the liquid densities and the vapour pressures for the homologous series of $n-$alkanes [80]. A number of studies used alternative potentials, such as Buckingham exponential [183] and the Mie potentials [80, 81, 185] to describe the chain molecules. The Mie potential, which incorporates an additional flexibility for the softness and hardness, has been demonstrated to significantly improve the representation of the VLE equilibrium and the second-derivative properties for a variety of systems [24]. Not only simple non-polar molecules such as $n-$alkanes can be reproduced by the flexible Mie potential, but also highly polar refrigerant gases, ring molecules, etc [66, 67]. In this chapter, we use the SAFT-$\gamma$ Mie EoS to develop the homonuclear and heteronuclear CG models for the chain molecules of different chemical nature. The SAFT-$\gamma$ Mie formalism has already been applied by Avendaño *et al.* [67] to derive the CG models for long $n-$alkanes such as $n-$decane ($n-\mathrm{C_{10}H_{22}}$) and $n-$eicosane ($n-\mathrm{C_{20}H_{42}}$). $n-$decane and $n-$eicosane were represented by the homonuclear chains of three and six tangentially bonded spheres, respectively. The work by Avendaño *et al.* was a proof-of-concept study and a systematic examination of both the quality of the fit to experiments and the scale of the agreement between the theory and the simulations of the underlying potential are missing.

## 3.3 Methods and models

### 3.3.1 SAFT-$\gamma$ Mie force field development

In this chapter, we present CG models for chain molecules of different chemical nature, including $n-$alkanes, $n-$perfluoroalkanes, $n-$ethers, and the mixed components combing several functional groups. As well established [2], the first step to coarse grain a model is to split a molecule into the CG beads. Based on a heuristic approach that one bead contains approximately three heavy atoms with their accompanying hydrogen atoms [85, 149], a 3-to-1 mapping is applied when possible. The studied moieties and the chosen mapping is presented in Table 3.2. The underlying atomistic model is shown only as a guide to the eye, as the parameterisation does not follow the geometry or energetics of any particular atomistic model. The atomistic definition of each type of group along with the chosen name is given in Table 3.1.

The second step in the development of a CG model is to derive the force field parameters. The intermolecular potentials are adjusted to macroscopic properties of the system by means of the the SAFT-$\gamma$ Mie EoS. The parameter matrix involves the estimation of both like interactions and crucially the unlike interactions. The details to the parameterisation of each type of group are given in the Results section.

**Table 3.1:** Bead name and atomistic definition.

| Name | Atomistic definition |
|---|---|
| $CM$ | $-CH_2-CH_2-CH_2-$ |
| $FF$ | $-CF_2-CF_2-$ |
| $CF$ | $-CH_2-CH_2-CF_2-CF_2-$ |
| $EO$ | $-CH_2-O-CH_2-$ |
| $OA$ (two beads) | $OH-CH_2-CH_2-O-CH_2-$ |

### 3.3.2 Intramolecular potentials

The SAFT theory does not have any built-in specification with respect to the internal degrees of freedom between adjacent beads in a chain. By construction, chains are bonded at a distance corresponding to a "contact" value equivalent to $\sigma$ and the bonding of a bead in a chain is independent on the position of the other beads in the same chain. The result of these restrictions of the first-order thermodynamic perturbation theory underlying all

common SAFT treatments, is that the theory is more accurate for straight elongated chains [186]. To improve the quality of the match between theory, simulations and experimental data, we add an internal restriction to the chain molecules to bias towards extended configurations [187]. Thereby, by incorporating the intramolecular potentials, additional structural information is added to the description of the Helmholtz free energy. To do this in a consistent way, we borrow a concept from "bottom-up" CG approaches and match the bond distributions of the CG models to those suggested by conventional all-atom force fields.

AA-MD simulations are carried out to obtain the CG intramolecular potentials of $n-$alkanes [187]. Simulations of $n$-hexane, $n$-nonane and $n$-dodecane are performed at $T = 400$ and 600 K using the NERD force field. The CG intramolecular potentials are constructed from the corresponding distribution functions of atomistic trajectories. The intramolecular interactions are described by harmonic potentials that include bond stretching and bond-angle bending:

$$U_{intra} = \sum^{Bond} k_b(r - r_0)^2 + \sum^{Angle} k_a(\theta - \theta_0)^2, \tag{3.1}$$

where $k_b$ and $k_a$ are the bond and the bending spring constants, respectively, and $r_0$ and $\theta_0$ are the distance and angle at the potential minimum. The equilibrium distance $r_0$ is assumed to be equal to the size parameter $\sigma$ of the given component or the arithmetic mean in the case of cross-bead links. The average values of the spring constants and the bending angles of the named alkanes are obtained from the fit to a Gaussian distribution function and are summarised in Table 3.3. The bond stretching constant is relatively large, hence the bond distance is essentially constant. Although counterintuitive, the same intramolecular potentials are applied to describe all types of chain molecules studied in this work, including perfluoroalkanes, semi-fluorinated alkanes, and non-ionic surfactants as the details of the potential do not seem to be of importance. In principle, any method to characterise the intramolecular degrees of freedom could be employed in this context. The choice made here makes no preference of other approaches, much as an *ab initio* calculations or the use of other all-atoms force fields. The influence on the final CG potential is in any case, minor.

**Table 3.3:** CG intramolecular parameters obtained from AA simulations.

| $k_b$ / (kcal × mol$^{-1}$ × Å$^{-2}$) | $k_a$ / (kcal × mol$^{-1}$ × rad$^{-2}$) | $\theta$/∘ |
|---|---|---|
| 7.583 | 2.650 | 157.6 |

### 3.3.3 Simulation details

The vapour-liquid equilibria of SAFT-$\gamma$ CG models is determined by Molecular Dynamics simulation in the canonical ensemble, with constant number of particles $N$, temperature $T$ and volume $V$. The system temperature is maintained using a Nosé-Hoover thermostat [188, 189] with the coupling constant set to 1.0 ps. The simulations are carried out with the Gromacs package, version 4.5.5 [190]. The global density of the systems is chosen inside the coexistence region (Figure 3.1). Since SAFT EoS accounts for a "full" potential, i.e. infinite cutoff, the cutoff radius in MD simulation is set to 3.0 nm, which is approximately $7 \times \bar{\sigma}$ and therefore is long enough to mimic the theory. A detailed description of the effect of the cutoff radius on the thermodynamic properties is given in Section 4.4.1. A simple truncation of non-bonded interactions is used without any tail-corrections.

The system containing 3000 molecules of various chain length are arranged in an orthorhombic box of a constant volume, where $L_x = L_y < L_z$, with $z$-dimension being approximately three times larger than the $x$ and $y$. The standard periodic boundary conditions [18] are applied. After the equilibrium has been established, the configurations are sampled for further 20 ns. Two phases are formed between two planar interfaces and the saturation liquid and vapour densities are obtained as the corresponding averages from the density profiles. The vapour pressure and the interfacial tensions are calculated from the components of the pressure tensor. Whereas the vapour pressure corresponds to the normal component of the pressure tensor $P^V = P_{zz}$, and the surface tension is obtained from the difference of the normal and tangential components as:

$$\gamma = \frac{1}{2}\left(P_{zz} - \frac{1}{2}(P_{xx} + P_{yy})\right), \tag{3.2}$$

where the pre-factor $\frac{1}{2}$ implies the presence of two interfaces.

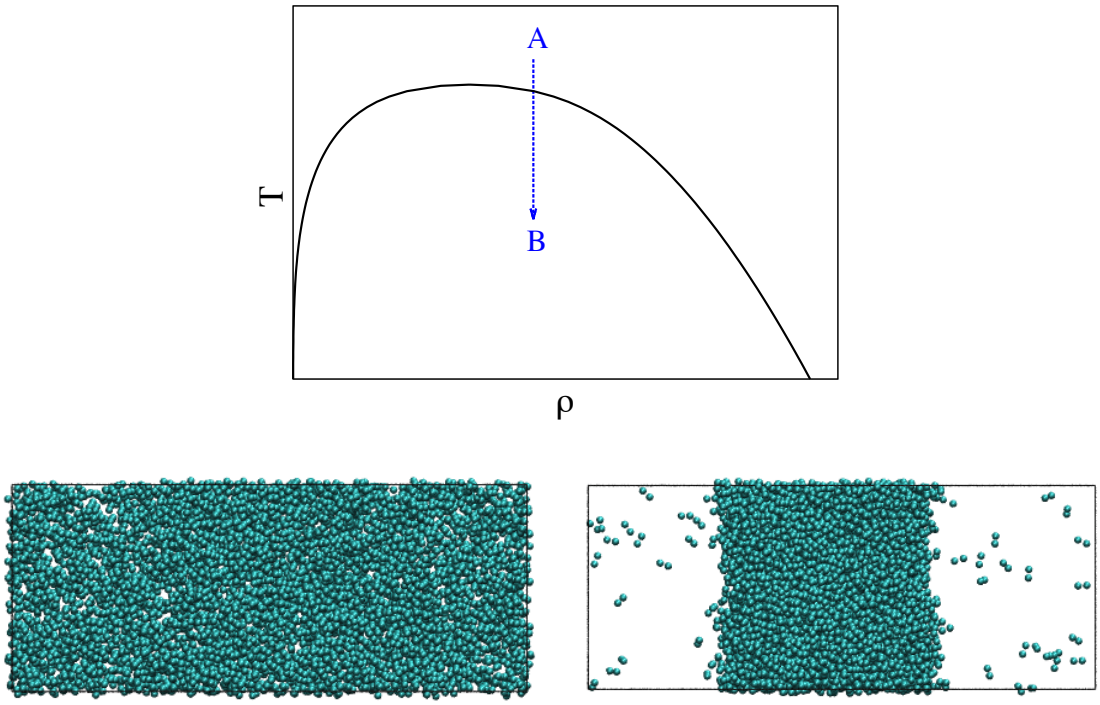



**Figure 3.1:** Schematic representation of the MD simulation of the vapour-liquid equilibrium properties in the canonical ensemble. The initial state (A) is being quenched to the state B, which is inside the coexistence region, so that two phases are formed. On the left: snapshot of the state A, corresponding to the one phase region; on the right: snapshot of the state B, corresponding to the two phase region.

### 3.3.4 Scaling laws and critical data

The critical points data are estimated from the vapour-liquid equilibrium MD simulation results using the scaling laws [60]. At each temperature, the corresponding liquid ($\rho_l$) and vapour ($\rho_v$) densities are related to the critical temperature $T_c$ as

$$\rho_l - \rho_v = B_0|\tau|^{\beta_c}, \tag{3.3}$$

where $\tau = 1 - T/T_c$, $B_0$ is an adjustable constant, and $\beta_c = 0.325$ is the critical exponent fixed at its universal renormalisation-group value. The critical density $\rho_c$ is estimated from the law of rectilinear diameters:

$$\frac{\rho_l + \rho_v}{2} = \rho_c + D|\tau| \tag{3.4}$$

and the critical pressure from the extrapolation of Clausius-Clapeyron equation

$$\ln P_c = C_1 + \frac{C_2}{T_c}, \tag{3.5}$$

where $D$, $C_1$ and $C_2$ are correlation parameters.

## 3.4 Results

### 3.4.1 Alkanes

The series of $n$-alkanes are modelled as homonuclear chains of tangential Mie segments. The main difference to the previous work [67] is the inclusion of an intramolecular potential in order to account for the structural contributions. All beads, i.e. the end-group containing a terminal $CH_3-$ group and the middle group containing exclusively $-CH_2-$ groups, are based on the same parameters. This contributes to the simplicity of the model by reducing the number of required unlike interactions with other groups. The target properties are the saturation density and vapour pressure over the fluid temperature range between the triple point up to 90 % of the critical point with a step of 20 K. The 3-to-1 mapping is applied, so that $n$-alkanes containing $C_3$-unit can be represented: $C_6H_{14}$, $C_9H_{20}$, $C_{12}H_{26}$, $C_{15}H_{32}$, $C_{18}H_{38}$, etc. This mapping is in line with the optimal number of segments $m$ for alkanes suggested by Jackson and co-workers as $m = 1 + (C - 1)/3$ [191, 192], with $C$ being a number of carbon atoms. The alkyl $-C_3H_6-$ bead is referred to as $CM$ bead.
For the chosen mapping, the exponent pair (15-6) is found to provide the best representation of the target properties of the entire alkane family. Hence, leaving these exponents fixed to (15-6), the size $\sigma$ and the energy $\epsilon$ parameters are adjusted individually for each component to reproduce the saturation liquid densities and the vapour pressures. Parameters for the $n-$alkanes along with the corresponding deviations from the experimental data are listed in Table 3.4 (with the $n$-alkanes based on the 3-to-1 mapping highlighted in blue). The description of the saturation densities and the vapour pressures as a function of temperature is illustrated for $n$-hexane, $n$-nonane, $n$-dodecane, $n$-pentadecane, and $n$-octadecane in Figure 3.2. The SAFT-$\gamma$ Mie EoS reproduces the experimental data with a high accuracy at low and mediate temperatures. Only close to the critical point, the theory exhibits slight deviations. The overshoot is a known feature of all mean-field theories [193], which do not account for fluctuations. To overcome this limitation, a special

treatment is required, as for instance the renormalisation-group theory [194]. Since the description of the critical region is not the main interest of the current work, those issues will not be addressed further.

Parameters derived with the SAFT EoS, are subsequently used as a direct input of the intermolecular potentials in MD simulation, along with the intramolecular potentials presented in Section 3.3.2. The molecular simulations reproduce the results obtained with the equation of state very accurately (Figure 3.2). Most remarkable, however, is the agreement of the MD simulations with experimental surface tensions if one considers that the surface tensions were not used in the parameter fitting procedure and are therefore true predictions of the model.

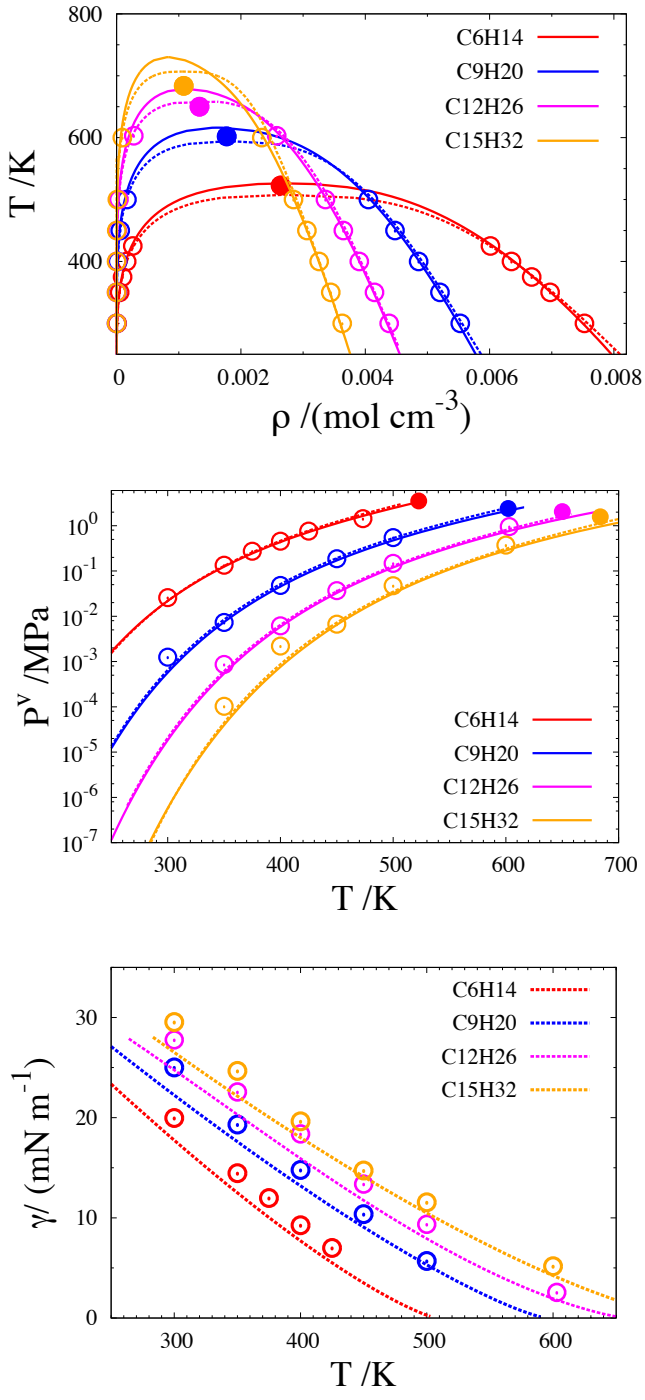



**Figure 3.2:** Saturation density (top), vapour pressure (middle) and surface tension (bottom) as a function of temperature for $n$-alkanes ($C_6H_{14}$, $C_9H_{20}$, $C_{12}H_{26}$, $C_{15}H_{32}$). Dashed lines denote the smoothed experimental data from NIST [195]; continuous lines are the results from SAFT-$\gamma$ Mie EoS based on the parameters from Table 3.4, open circles denote MD simulation results; the filled circles are critical points estimated obtained from the scaling laws.

The chosen CG mapping causes a limitation in the chain length of alkanes which can be represented through the model and hence the alkyl-group in the heteronuclear models, for instance in the surfactant molecules. In order to reproduce the $n-$ alkanes of any chain length, the 3-to-1 mapping has to be relinquished. Hence, alkanes $n-C_5H_{12}$ to $n-C_{14}H_{30}$ can be fitted individually with the number of beads fixed to the nearest integer from the division by three. As before, the Mie (15-6) potential is used for the sake of consistency and $\sigma$ and $\epsilon$ are optimised to reproduce the saturation densities and pressures. The parameters are given in Table 3.4 and the representation of the VLE properties with SAFT-$\gamma$ Mie EoS can be seen in Figure D.1 in Appendix.

**Table 3.4:** Intermolecular interaction parameters ($\sigma$ and $\epsilon$) and corresponding deviations in saturation liquid density and vapour pressure for homologous series of $n-$alkanes based on the Mie (15-6) potential. $m$ is the number of the tangent homonuclear spheres that build the molecule. $n-$alkanes which are based on the three-to-one mapping are highlighted in blue.

| Component | $m$ | $\sigma$ /Å | $\epsilon/k_B$ /K | $\Delta\rho$ % | $\Delta P^V$ % |
|---|---|---|---|---|---|
| $C_5H_{12}$ | 2 | 4.2449 | 310.78 | 0.51 | 0.48 |
| $C_6H_{14}$ | 2 | 4.5089 | 342.00 | 0.43 | 6.26 |
| $C_7H_{16}$ | 2 | 4.5736 | 380.87 | 11.46 | 17.24 |
| $C_8H_{18}$ | 3 | 4.2412 | 321.65 | 1.16 | 2.12 |
| $C_9H_{20}$ | 3 | 4.4212 | 344.83 | 0.30 | 3.85 |
| $C_{10}H_{22}$ | 3 | 4.4908 | 363.99 | 5.48 | 9.25 |
| $C_{11}H_{24}$ | 4 | 4.2212 | 328.18 | 1.11 | 3.04 |
| $C_{12}H_{26}$ | 4 | 4.3635 | 344.42 | 0.22 | 1.33 |
| $C_{13}H_{28}$ | 4 | 4.4680 | 354.25 | 0.42 | 2.06 |
| $C_{14}H_{30}$ | 5 | 4.2332 | 329.81 | 5.04 | 6.99 |
| $C_{15}H_{32}$ | 5 | 4.3286 | 344.85 | 0.71 | 2.13 |

### 3.4.2 Perfluoroalkanes

Perfluoroalkanes have gained a strong industrial interest due to their inert properties and a wide range of applications as refrigerants, solvents, and anaesthetics. A perfluoroalkane molecule has a much stiffer backbone compared to the alkanes and the molecules are much thicker since the fluorine atoms are larger and heavier than hydrogens. Additionally, fluorine atoms exhibit higher electronegativity so that the perfluoroalkanes are more polar than their hydrogen containing counterparts. Although the SAFT EoS does not account for polarity in an explicit fashion, the use of variable exponents in the Mie form does

allow the shape and range of the potential to be modified and to mimic the effect of polarity. Analogous to the homologous series of $n-$alkanes, the intermolecular forces in the $n-$perfluoroalkanes were described by the effective interactions, reproduced by the Mie (13-6) potential, which was found to provide a good representation for the studied components. Due to heavy fluorine atoms, the best aspect ratio for the CG mapping was found to be 2-to-1, i. e. with one bead represented by $-(CF_2)_2-$ group, in following referred to as $FF$-group. By fixing the exponents to 13 and 6, the size and the energy parameters for long chain perfluoroalkanes ($C_6F_{14}$, $C_8F_{18}$, $C_{10}F_{22}$) are then individually adjusted to reproduce the saturation densities and vapour pressures for the temperatures up to $0.9 \times T_c$, in an analogous manner as for $n-$alkanes. The parameter set with the corresponding deviations in target properties is given in Table 3.5. The subsequently performed MD simulations, based on the derived parameter set, illustrate a very good agreement with the results from the equation of state and experimental data for the liquid densities and vapour pressures for the studied components. A reasonable agreement with experiment is also obtained for the surface tensions, which were not used in the parameter optimisation (Figure 3.3).

**Table 3.5:** Intermolecular interaction parameters ($\sigma$ and $\epsilon$) and corresponding deviations in saturation liquid density and vapour pressure for homologous series of $n-$perfluoroalkanes based on Mie (13-6) potential.

| Component | $m$ | $\sigma$ /Å | $\epsilon/k_B$ /K | $\Delta\rho$ % | $\Delta P^V$ % |
|---|---|---|---|---|---|
| $C_6F_{14}$ | 3 | 4.3838 | 240.32 | 0.49 | 2.78 |
| $C_8F_{18}$ | 4 | 4.2789 | 238.63 | 2.19 | 7.21 |
| $C_{10}F_{22}$ | 5 | 4.2095 | 230.91 | 2.13 | 3.28 |

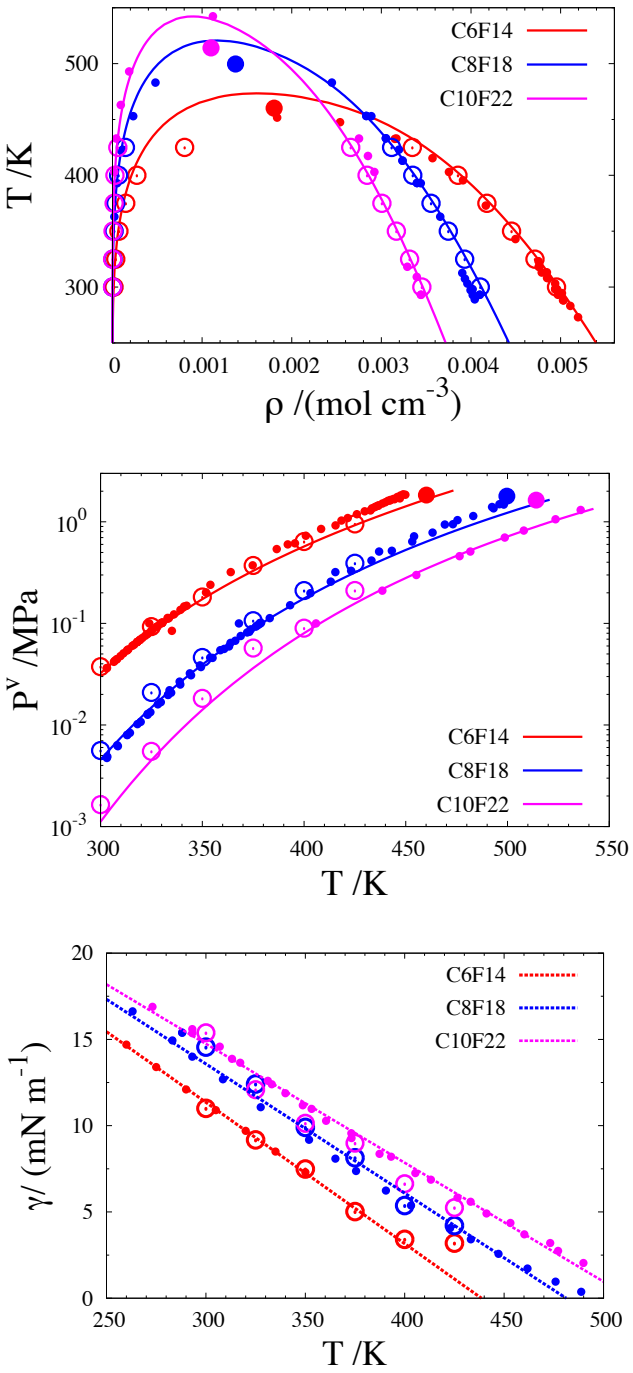



**Figure 3.3:** Saturation density (top), vapour pressure (middle) and surface tension (bottom) as a function of temperature for *n*-perfluroalkanes ($C_6F_{14}$, $C_8F_{18}$, $C_{10}F_{22}$). The small filled symbols represent the experimental data from NIST [195]; continuous lines are the results from SAFT-$\gamma$ Mie EoS based on the parameters from Table 3.5, open circles denote MD simulation results; the filled large circles are critical points estimated obtained from the scaling laws.

### 3.4.3 Ethers

For a later development of the CG models for the non-ionic surfactant solutions (Chapter 6), additionally to the non-polar alkyl groups, the representation of the polar ether groups is required. For this, the homologous series of long-chain ethers ($C_4H_{10}O_2$, $C_6H_{14}O_3$, $C_8H_{18}O_4$, and $C_{10}H_{22}O_5$) are parameterised, where the group ($-CH_2-O-CH_2-$) is referred to as *EO* and is represented as one CG bead. The aim of this section and Sec. 3.4.4 is to obtain a reliable parameter set which allows an accurate representation of the non-ionic surfactant solutions, therefore a detailed validation of properties of the chain molecules is omitted. To derive the intermolecular parameters, a similar assessment is undertaken as for the previously described homologous series of $n-$alkanes and $n-$perfluoroalkanes with no additional treatment for polarity. The exponent pair (19-6) is found to yield the best representation of the polar ether groups. By fixing exponents to (19-6), $\sigma$ and $\epsilon$ parameters are fitted individually to the saturation densities and vapour pressures for each component for the temperatures up to $0.9 \times T_c$. The parameter set for different ether chain length and the corresponding average deviations from the experimental VLE data are given in Table 3.6. A good agreement is achieved for the saturation densities, whereas the vapour pressures for long-chains exhibit higher deviations. This might be, at least partially, due to the very low vapour pressures of these longer molecules. The small (absolute) values of the vapour pressure makes both the experimental determination and the calculation using simulation prone to relatively large statistical deviations, which would show up as large absolute average deviations (AAD). The representation of the VLE properties as functions of temperature is given in Figure D.2 in Appendix.

**Table 3.6:** Intermolecular interaction parameters ($\sigma$ and $\epsilon$) and corresponding deviations in the target properties for the homologous series of ethers based on the Mie (19-6) potential.

| Component | m | $\lambda_r$ | $\sigma$ /Å | $\epsilon/k_B$ /K | $\Delta\rho$ % | $\Delta P^V$ % |
|---|---|---|---|---|---|---|
| $C_4H_{10}O_2$ | 2 | 19 | 4.2410 | 409.26 | 0.48 | 2.11 |
| $C_6H_{14}O_3$ | 3 | 19 | 4.1527 | 405.19 | 1.03 | 1.94 |
| $C_8H_{18}O_4$ | 4 | 19 | 4.0882 | 399.25 | 2.91 | 15.30 |
| $C_{10}H_{22}O_5$ | 5 | 19 | 4.0613 | 396.90 | 0.61 | 26.64 |

### 3.4.4 Triethylene glycol

The triethylene glycol molecule ($OH - CH_2 - CH_2 - O - CH_2 - CH_2 - O - CH_2 - CH_2 - OH$) is also used for a representation of the end-groups of non-ionic surfactants. The molecule is modelled as a chain of four CG beads, hence each pair of beads represents the terminal $-CH_2 - O - CH_2 - CH_2 - OH$ group. The exponent pair of (19-6) is chosen in order to simplify further search for the unlike interaction parameters with the neighbouring ether group. The best parameter set is obtained for $\sigma = 3.7244$ Å and $\epsilon/k_B = 461.11\ K$. The deviations compared to the experimental data are within 0.73 % for the liquid densities and within 12 % for the vapour pressures over a temperature range of a liquid fluid (see Figure D.3 in Appendix for representation of the target properties by the model).

### 3.4.5 Heteronuclear models

A strength of a group-contribution method is in its predictive power based on a possibility to calculate the physical properties of the entire component directly from the description of the comprising individual functional groups. Parameters for a chemical group derived from one particular moiety can be transferred in the description of a new molecule which contains this particular group along with other functional groups. The unlike interaction parameters can be obtained either from the properties of the mixture or from the properties of the pure component which comprises both types of groups.

Here, we demonstrate how the CG parameters derived from the homonuclear models of alkanes, perfluoroalkanes, ethers, and triethylene glycol, can be used to build new heteronuclear models. We showcase the moieties which contain groups of a very different chemical nature and different polarities. Different polarities generate components with an interface active behaviour which act as surfactants. Two examples illustrated here include semi-fluorinated alkanes which contain non-polar alkyl- and polar perfluoroalkyl-groups, and non-ionic surfactants from alkyl polyoxyethylene glycol family ($C_iE_j$), which contain non-polar alkyl- and polar ethoxy-groups.

#### Semi-fluorinated alkanes

Chemical differences in alkane and perfluoroalkane components cause weak unlike interactions, thereby leading to liquid-liquid immiscibility, large positive deviation from Raoult's

law, and large positive excess properties in the mixtures. Connecting both chemical groups by the covalent bonds in one molecule of a semi-fluorinated alkane (SFA) yields a number of unique properties, such as inertness, biocompatibility, ability to solubilise respiratory gases, and adsorb on the interfaces.

In order to describe the weak unlike interactions between the alkyl- and perfluoroalkyl-groups and therefore to ensure an accurate description of the highly non-ideal phase behaviour, the properties of mixture are used. The unlike interactions for size, exponent, and energy are obtained by using Eqs. 2.57, 2.59 and 2.60, respectively; the $k_{ij}$ parameter is adjusted directly to the target properties, in this case to the liquid-liquid equilibrium. The unlike parameter set of $\lambda_{ij} = 13.95$ and $\epsilon_{ij}/k_B = 265$ K was found to reproduce not only the liquid-liquid but also the vapour-liquid equilibria, simultaneously and at different temperatures for the mixture of $n-$hexane and $n-$perfluorohexane (Figure 3.4).

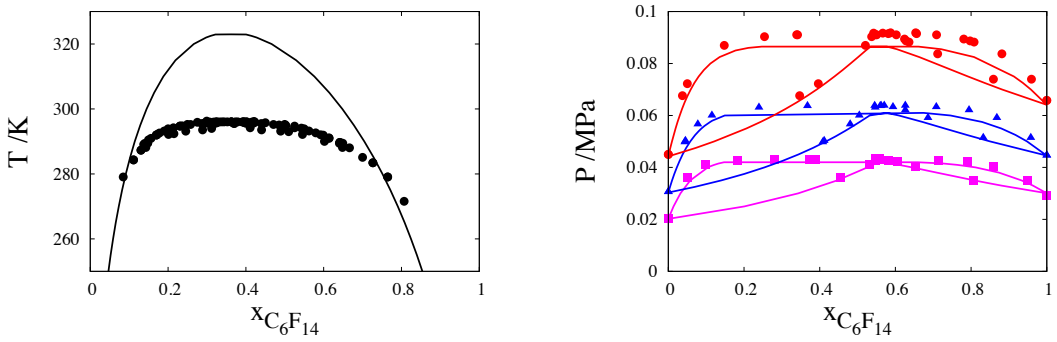



**Figure 3.4:** Liquid-liquid and vapour-liquid equilibria of $C_6F_{14} + C_6H_{14}$ mixture. The continuous line denotes the prediction with SAFT-$\gamma$ Mie EoS based on the unlike parameters $\lambda_{ij} = 13.95$ and $\epsilon_{ij}/k_B = 265$ K. Experimental data, denoted by the filled symbols, is taken from DETHERM [196] for LLE and Ref. [197]; VLE is given at $T = 298.15$ K (circles), $T = 308.15$ K (triangles), and $T = 318.15$ K (squares).

By having a chemical structure $F_3C-(CF_2)_i-(CH_2)_j-CH_3-$ (or simply $F_iH_j$), SFA exhibit a strong dipole moment on the boundary between the alkyl and perfluoroalkyl groups. Modelling of this behaviour requires a special treatment, which in this work is represented by introducing a polarised linker group $-CH_2-CH_2-CF_2-CF_2-$ (referred to as $CF$-group), containing two edge methyl- and two difluoromethylene groups. The representations of the alkyl ($CM$) and perfluoroalkyl ($FF$) groups were taken from

the homonuclear models of $n-$alkane and $n-$perfluoroalkane containing four segments, namely $n-$dodecane and $n-$perfluorooctane (Tables 3.4 and 3.6, respectively). Available experimental densities and vapour pressures for components $F_4H_5$, $F_4H_8$, $F_6H_8$, $F_8H_8$ taken from work by Morgado *et al.* [198, 199] were considered in the paramererisation of heteronuclear models.

The like interactions for the new linker group $CF$ and its unlike interactions with $CM$ and $FF$ groups were obtained simultaneously by adjusting to experimental saturation densities and vapour pressures of the named SFA components. The resulting parameter set is given in Table 3.7. The quality of representation of the target properties is illustrated in Figure 3.5. The experimental data is given for a low temperature region (below 350 K), which is correctly represented by the models. The liquid densities of $F_4H_5$, $F_4H_8$, and $F_8H_8$ are in a good agreement with experimental results, only $F_6H_8$ shows minor deviations. The experimental vapour pressures are given only for $F_4H_5$ and $F_4H_8$ components, which are accurately reproduced by the equation of state. The vapour pressures for $F_6H_8$ and $F_8H_8$ are considered as predictions and require a verification by the experiment.

The consequently performed MD simulations reveal a good agreement with the theoretical representation for the liquid densities. The vapour pressures, especially for longer chains, exhibit relatively large deviations from the theoretical predictions. This can be explained that the pressures are quite low and hence are difficult to obtain accurately. The experimental surface tensions of semi-fluorinated components roughly represent an average between the $n-$alkanes and $n-$perfluoroalkanes. The measured values for SFA differ only slightly from each other. The predictions via MD simulations are of the same order of magnitude but cannot exactly reproduce the experiment.

In general, our simple models for semi-fluorinated alkanes based on a very coarse description and a limited parameter set, are able to reproduce the complex interactions within the $F_iH_j$ molecule, which cause a number of important properties, including inertness and ability to adsorb on the interface.

**Table 3.7:** Unlike interaction parameters between the CG groups of semi-fluorinated alkanes.

| Bead | Bead | $\sigma_{ij}$/ Å | $\epsilon_{ij}/k_B$ /K | $\lambda_{r,ij}$ |
|---|---|---|---|---|
| $CF$ | $CF$ | 5.3087 | 522.33 | 13.95 |
| $CM$ | $CF$ | 4.8361 | 315.37 | 14.46 |
| $CF$ | $FF$ | 4.7938 | 219.23 | 13.46 |
| $CM$ | $FF$ | 4.3212 | 265 | 13.95 |

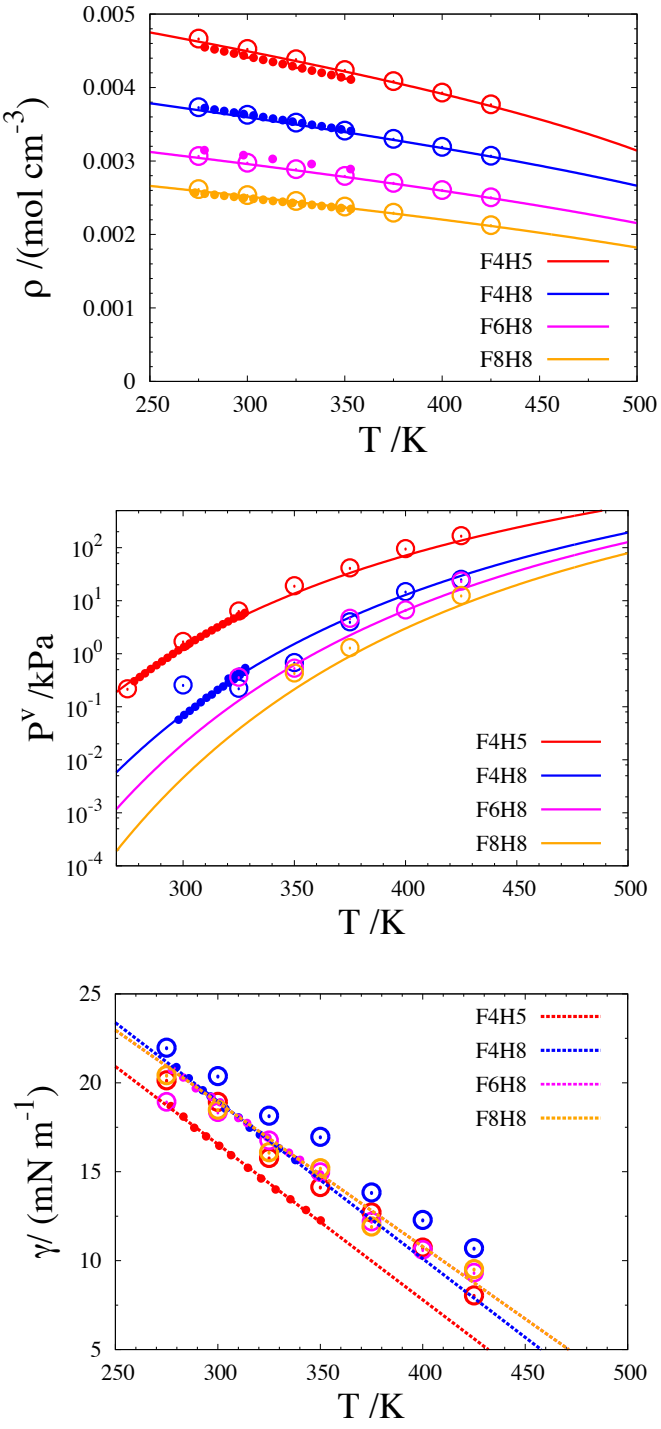



**Figure 3.5:** Atmospheric liquid density (top), vapour pressure (middle) and surface tension (bottom) as a function of temperature for semi-fluorinated alkanes ($F_4H_5$, $F_4H_8$, $F_6H_8$, $F_8H_8$, where $F$ denotes the number of fluorinated groups and $H$ denotes the number of methyl groups). The small filled symbols represent the experimental data taken from Morgado *et al.* [198, 199]; continuous lines are the results from SAFT-$\gamma$ Mie EoS based on the parameters from Table 3.7, open circles denote MD simulation results; the filled large circles are critical points estimated obtained from the scaling laws.

### Non-ionic surfactants

The non-ionic alkyl polyoxyethylene glycol surfactants $C_iE_j$ can be described within the group-contribution method by the groups derived from the homologous series of alkanes, ethers, and triethylene glycol. $C_{10}E_4$ surfactant is used as an example to visualise the break-down into the corresponding types of groups ($CM$, $EO$, and $OA$) (Figure 3.6).

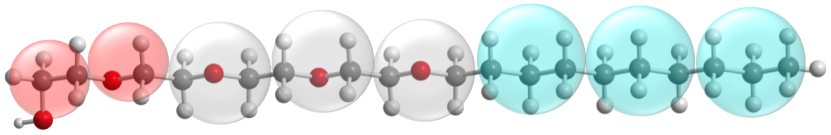


**Figure 3.6:** CG molecular model for $C_{10}E_4$ surfactant ($CH_3 - (CH_2)_9 - (OC_2H_4)_4 - OH$). Transparent spheres represent CG particles. Turquoise beads denote the alkyl group ($CM$), grey beads the ethoxy group ($EO$) and the red beads the head group ($OA$).

The unlike interaction parameters between the surfactant beads are parameterised to reproduce the VLE properties of chemical components that contain both types of beads. Moieties of different chain length are used simultaneously in the parameter optimisation procedure by means of the SAFT-$\gamma$ Mie EoS. In this manner, a universal parameter set is obtained that is applicable to the surfactants of different chain length.

The only new fitted parameter is the unlike energy ($\epsilon_{ij}$). The interaction between $CM$ and $EO$ groups are fitted to the VLE data of following chemical components: 1-methoxybutane ($C_5H_{12}O$), containing one $CM$ and one $EO$ group, 1-methoxyheptane ($C_8H_{18}O$), containing two $CM$ and one $EO$ group, and 1-(2-methoxyethoxy)butane ($C_7H_{16}O_2$), containing one $CM$ and two $EO$ groups. In order to parameterise the interactions between $OA$ and $CM$ groups, the VLE data of butyl glycol ($C_6H_{14}O_2$), containing two $OA$ and one $CM$ groups, and ethylene glycol monoheptyl ether ($C_9H_{20}O_2$), containing two $OA$ and two $CM$ groups, is used. The VLE properties of the named surfactants calculated by the SAFT-$\gamma$ Mie EoS appear to be in a good agreement with experimental data (Figure D.4 in Appendix).

In order to ensure the highly non-ideal phase behaviour, the interactions between $OA$ and $EO$ groups are fitted to the mixture properties of $C_7E_5$
($CH_3 - [CH_2]_6 - [CH_2 - O - CH_2]_4 - CH_2 - O - CH_3$) in $n$-dodecane ($C_{12}H_{26}$). This particular mixture involves all three type of groups ($OA$, $EO$ and $CM$). With the unlike interaction parameters $CM$-$EO$ and $OA$-$CM$ obtained from the pure component VLE

data, the only interactions which remain unknown for the mixture are between $OA$ and $EO$ groups and hence are adjusted to reproduce the solubility gap. The best possible agreement with experimental data is obtained with the unlike interactions $\lambda_{r,ij}$ =19 and $\epsilon_{ij}/k_B$ =392 K and is visualised in Figure 3.7. The obtained parameters can also reproduce the pure component VLE data of substances, which consist of $OA$ and $EO$ groups fairly reasonably (Figure D.4 in Appendix). In this manner, the derived parameter set, which is given in Table 3.8, is able to describe both, the VLE pure component properties of surfactants of different chain length and also properties of a mixture. Instead of a detailed description of the pure component non-ionic surfactants, we will focus on the properties of aqueous mixtures, which are discussed in detail in Chapter 6.

**Table 3.8:** Unlike interaction parameters between the CG groups of $C_iE_j$ surfactants.

| Bead | Bead | $\epsilon_{ij}/k_B$ /K | $\lambda_{r,ij}$ |
|---|---|---|---|
| $OA$ | $EO$ | 392.00 | 19 |
| $OA$ | $CM$ | 375.50 | 16.86 |
| $EO$ | $CM$ | 352.00 | 16.86 |

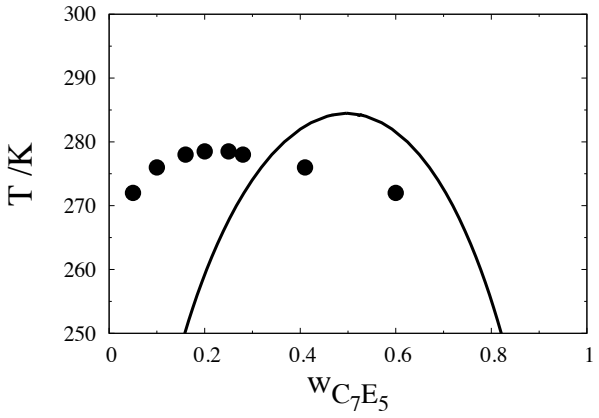



**Figure 3.7:** Liquid-liquid equilibrium of $C_7E_5$ surfactant in $n$-dodecane at $P = 0.1$ MPa. The filled circles denote the experimental data from Ref. [200] and the continuous line the results from SAFT-$\gamma$ Mie EoS. Abscissa is the weight fraction of $C_7E_5$.

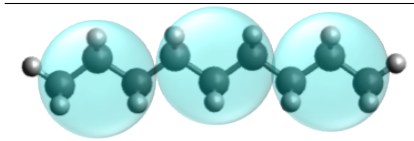
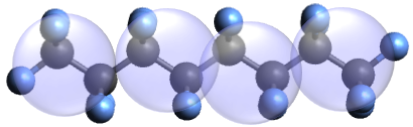
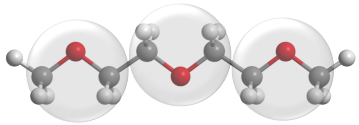
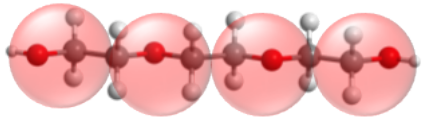
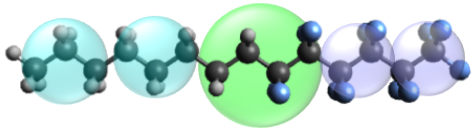
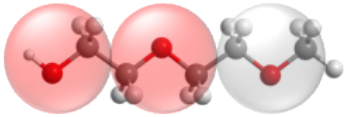
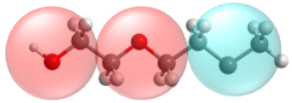
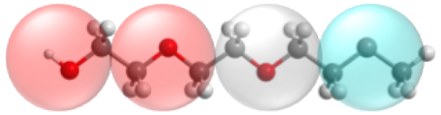
## 3.5 Conclusions

The SAFT-$\gamma$ Mie EoS is used to develop CG models for the chain molecules from different chemical families. The homonuclear models for the homologous series of $n-$alkanes, $n-$perfluoroalkanes and $n-$ethers are parameterised to reproduce the saturation densities and vapour pressures. All studied homologous series are treated within a single framework, where the interactions are described by means of the Mie intermolecular potential. Due to its flexibility and versatility, the Mie potential allows to capture the liquid densities and vapour pressures simultaneously, both for polar and non-polar components. The prediction is accurate over a wide range of temperatures, with the exception of the near-critical region, due to the mean-field nature of the equation of state. For the substances studied in this work, the exponents of the Mie potential differ significantly from each other due to different mapping and chemical nature of the components. Generalising these trends would require a more consistent study, which might be the subject of future work.

The parameters for the named groups are used within the group-contribution framework to derive new heteronuclear models. The pure component VLE and mixture property data is used to derive the unlike interactions between the CG beads. The heteronuclear models are represented based on the limited set of parameters, which appears to be transferable to represent both, the pure component and the mixture properties for the compounds of different chain lengths.

Molecular Dynamics simulations are performed in order to verify the results from the equation of state and also to predict the properties which are not accessible to the theory, such as surface tensions. Our CG force fields include the intramolecular potentials derived from the structure-based bottom-up approach and the intermolecular potentials obtained with the SAFT-$\gamma$ Mie EoS. MD simulations based on the derived parameter set illustrate a very good agreement with the results from the theory and therefore experimental data. Even surface tensions which are not used in the parameter optimisation procedure and therefore are true predictions of the model show a remarkable agreement with experimental data. In this manner, the SAFT-$\gamma$ Mie EoS allows to link the molecular simulations and experiments, thereby providing a very powerful predictable tool.

**Table 3.2:** CG representation for chain molecules.

| CG representation | Compound | Atomistic definition | CG beads |
|---|---|---|---|
|  | $n-$nonane | $CH_3-(CH_2)_7-CH_3$ | $CM-CM-CM$ |
|  | $n-$perfluorooctane | $CF_3-(CF_2)_6-CF_3$ | $FF-FF-FF-FF$ |
|  | (2-Methoxyethyl) ether | $CH_3-(O-CH_2-CH_2)_2-O-CH_3$ | $EO-EO-EO$ |
|  | triethylene glycol | $OH-CH_2-(CH_2-O-CH_2)_2-CH_2-OH$ | $OA-OA-OA-OA$ |
|  | perfluorohexyloctane | $CH_3-(CH_2)_7-(CF_2)_5-CF_3$ | $CM-CM-CF-FF-FF$ |
|  | 2-(methoxyethoxy) ethanol | $CH_3-(O-CH_2-CH_2)_2-OH$ | $OA-OA-EO$ |
|  | butyl glycol | $CH_3-(CH_2)_3-O-CH_2-CH_2-OH$ | $OA-OA-CM$ |
|  | 2-(butoxyethoxy) ethanol | $CH_3-(CH_2)_3-(O-CH_2-CH_2)_2-OH$ | $OA-OA-EO-CM$ |

# Chapter 4

# Water

## 4.1 Chapter overview

In this chapter we develop coarse-grained force fields for liquid water by using the SAFT-$\gamma$ methodology [68]. The corresponding models are based on an isotropic Mie potential characterised with temperature-dependent size and energy parameters owing to the averaging out of the strong directional interactions between water molecules. When each CG bead is taken to represent a single water molecule, we show that not all of the thermodynamic properties can be described simultaneously with the desired accuracy. Two models of water based on a one-to-one mapping are developed which can be used to reproduce target properties with a high accuracy over a wide temperature range. One is designed to reproduce the fluid-phase equilibria including saturated liquid density and the vapour pressure, and the other to capture the saturated liquid density and the vapour-liquid surface tension. At higher levels of coarse-graining we examine models comprising up to four molecules per bead (including two-to-one and four-to-one mapping schemes) and make some recommendations and comments on the virtues and deficiency of the models.

## 4.2 Introduction

Water is perhaps the most important liquid in our world and is the most common solvent in biological and industrial systems. Despite the enormous research effort expended over more than a century, modelling aqueous systems still remains a challenge [201]. A reliable force field is crucial for an accurate description of properties of water and its mixtures.

Numerous intermolecular potential models have been developed at different levels of resolution by matching specific structural and/or thermodynamic properties. Without the intension of being exhaustive, we briefly overview some of the popular molecular models for water. The reader is directed to the excellent reviews with a focus on the quantum-level description of water by Paesani and Voth [202], Baer *et al.* [203], and Santra *et al.* [204]; classical atomistic water models by Vega and co-workers [205] and by Shvab and Sadus [206]; the coarse-grained representations of water by Dill et al. [207], Chaimovich and Shell [208], Hadley and McCabe [209], and Knight and Voth [210]; and on the experimental validation of the various models by Striolo [211], and Meadley and Angell [212].

In *ab initio* approaches, the total potential energy surface of the system is calculated by solving the Schrödinger equation using the Born-Oppenheimer approximation. As a result of the computational complexity, these calculations are restricted to relatively small clusters of water molecules. Even then, many of the predicted bulk thermodynamic properties agree only qualitatively with experimental observables [203, 204]. One of the best known force fields derived from this type of calculations is the Matsuoka-Clementi-Yoshimine (MCY) model [213] obtained by parameterising *ab initio* quantum mechanical calculations for the water dimer in different relative molecular positions and orientations. Although the structural properties are well reproduced with the MCY model, significant deviations are observed for liquid density. In 1985, Car and Parrinello [214] combined Molecular Dynamics with quantum Density Functional Theory to simulate the fluid properties of water, explicitly introducing the electronic degrees of freedom as dynamical variables rather than using Born-Oppenheimer MD where the electron density is solved at each iteration. The quality of the description largely depends on the the choice of functional and basis set used in the quantum simulation. The vapour-liquid coexistence properties, which are particularly sensitive to details of the computation, can be off the experimental values by several orders of magnitude [203, 215]. Notwithstanding, the methodology made it possible to simulate clusters of over 100 molecules at the time, allowing for an exploration of the liquid state [216, 217].

At the classical atomistic level of description, experimental data for the bulk-phase properties and liquid phase structure is commonly used to parameterise the empirical pairwise additive potentials. These potentials employ analytical functions to give a simplified clas-

sical representation of repulsive, dispersive, and electrostatic forces. Rigid non-polarisable models with partial point charges are commonly employed in this context.

The classical intermolecular force fields in modern use for water are invariably based on the distributed-charge Lennard-Jones models proposed early on by Bernal and Fowler [218], Rowlinson [219, 220], Pople [221], and Bjerrum [222]. The first atomistic molecular simulation study of water was reported by Barker and Watts [223], who used the model of Rowlinson [220] in their simulations. The model consists of a Lennard-Jones spherical site and four electrostatic charges positioned to reproduce the dipole moment of water. A similar framework was used by other groups in subsequent studies to develop more refined water models with differing numbers of charge sites and geometries, including the BNS model of Ben-Naim and Stillinger [224], the ST2 model of Stillinger and Rahman [225, 226], the simple point charge (SPC) models of Berendsen and co-workers [77, 227] and the transferable interaction potential (TIP) models of Jorgensen and co-workers [76, 228–236]; in this context is it also important to mention the distributed charge exp-6 model of Errington and Panagiotopoulos [237], based on the modified Buckingham exponential-6 potential.

The TIP and SPC families of atomistic models generally offer a good overall description of the behaviour of water, even though they cannot be used to capture all of the properties and anomalies simultaneously [205, 238]. For instance, the TIP4P model does not allow one to represent the phase diagram in a quantitative agreement [239, 240], and the predicted dielectric constant is highly underestimated compared to the experimental value [205]. The TIP3P, TIP4P, SCP and SPC/E models are unable to provide an accurate representation of the experimental oxygen-oxygen radial distribution function [77, 241, 242], and unphysical clusters are found with the TIP3P model in the gas phase [243]. The SPC, TIP3P and TIP4P models also underestimate the melting point, corresponding to the values of 190 K, 146 K and 232 K, respectively [205]. An unsatisfactory description of the vapour-liquid, critical point, vapour pressure, and second virial coefficient is obtained with the TIP4P and TIP4P-Ew models (the latter of which is refined for use with Ewald summations of the electrostatic contributions) [244]. Probably the best overall description of the thermodynamic and structural properties of water is provided by TIP4P/2005 and TIP4P/Ice for the condensed liquid and solid states, respectively [233–235]; the recently re-parameterised TIP4P model of Huang *et al.* [236] provides a particularly good descrip-

tion of the saturation pressure and heat of vaporisation, including the near critical region. Despite some of their inadequacies, the SCP and TIP point charge models of water are still in ubiquitous use as they provide a predictive platform for a broad variety of properties including complex aqueous systems of biomolecules.

Polarisable force fields first introduced in the late seventies [245, 246] and developed extensively since account for the many-body polarisation effects to improve the description of the dielectric properties by, for example, including polarisable Gaussian charges [247–251], fluctuating charges [252, 253], polarisable point charges and multipoles [254–258], moving charged shells [259], or molecular flexibility [260–262]. Taking many-body polarisability effects into account also allows one to reproduce the properties of the high- and low-density states simultaneously at the expenses of more complexity and computational cost. The application of these kind of models in large scale simulations, for example when water acts as a solvent, is extremely time consuming.

Potential models can be considered that do not incorporate the electrostatic interactions explicitly but instead treat the strong short-range directional hydrogen bonding by including a number of off-centre sites with the spherical molecular core [263]. Accurate versions of the SAFT equation of state have been used to parameterise Wertheim association model for water that is particularly suited for the simulation of confined systems [164] where the use of Ewald summations for evaluating the long-range electrostatic interactions can be problematic.

Before we describe the use of coarse-graining methodologies to develop simple, typically spherically symmetric, intermolecular potentials for water and aqueous systems, designed to provide an accurate quantitative representation of target structural and thermodynamic properties, it is important to acknowledge the large body of work on so-called “toy models”. In contrast to CG force fields, “toy models” are developed to capture the qualitative underlying physics of the interactions in water with the aim of capturing distinctive features of the system’s behaviour, such as the anomalous low density of ice, the density maximum, heat capacity, and compressibility minima, without attempting to reproduce the aqueous properties faithfully. One of the first toy models of water was developed by Ben-Naim [264] based on a Lennard-Jones molecular core with directional attractive sites characterised by an angular dependent Gaussian cutoff; “Mercedez-Benz” variants of the Ben-Naim model have been used extensively to understand the fundamentals of the

anomalous behaviour [207, 265, 266]. Simple isotropic potentials with two characteristic length-scales [267–271], modified van der Waals [266, 272], and core-softened [273–280] toy models are also commonly employed to reproduce the water-like anomalies.

Modelling water molecule as a single spherically symmetric coarse-grained interaction site is a considerable challenge, because the strong anisotropy and short-ranged nature of the interactions are represented in an average effective manner. Yet, the development of simple, typically spherically symmetric, coarse-grained models of water without an explicit treatment of the electrostatics has gained much popularity in recent years [78, 82, 84, 89, 125, 159, 208, 281–299].

It comes as no surprise to find that these simple isotropic models cannot simultaneously capture all of the thermodynamic and structural properties of water. Notwithstanding the shortcomings due to the rather crude representation of the force field, these coarse-grained models are very useful in the simulation of large macroscopic systems dominated by solvent effects, e.g. biological systems, solvent-mediated phase or microphase separation, etc. In this scenario, the use of a simplified coarse-grained model offers a great advantage in computational time and is therefore paramount as a platform for multiscale modelling. The main purpose of CG models is to reproduce basic structural, dynamical, thermodynamic properties, and phase equilibrium properties in reasonable agreement with the target experimental data, at a low computational cost. It is clear that by reducing the resolution through the coarse-graining procedure one's capability of accurately describing the systems's behaviour must be diminished with respect to the atomistic models, which already fail to reproduce some of the key properties. The key to a successful coarse-graining procedure is to establish correct balance between simplicity and accuracy.

Coarse-grained models of water are commonly developed by lumping together several molecules into a single CG bead. Different levels of CG mapping have been reported depending on the purpose of study, ranging from one [82] and to five [291] water molecules per bead. Electrostatics and hydrogen-bonding interactions are often treated implicitly by considering a spherically symmetric potential.

"Bottom-up" techniques such as Iterative Boltzmann Inversion [285] and force matching [125] have been used to parameterise the CG potentials of water from a detailed atomistic description. Some empirical CG water models based on a "top-down" approach have also been reported : in such an approach the parameters that characterise the potential func-

tions are adjusted to match one or more of the macroscopic properties using an iterative procedure. One of the most commonly employed CG models of water is the Martini CG force field, which is based on the ubiquitous Lennard-Jones (12-6) potential [78]. The Martini model can be used to capture some of the salient structural properties of aqueous systems comprising lipids or surfactants, which has made the force field very popular in studies involving biological systems. The model is highly computationally efficient owing to the "aggressive" level of coarse-graining - with four water molecules per CG bead. The strong polar interactions present in water and aqueous systems are represented with deep energetic wells, which can unfortunately cause freezing at physiological temperatures [159, 287], making the model suitable only for mixtures with other components (that act as anti-freezing agents). The Martini model is found to significantly underestimate the surface tension and overestimate the compressibility of the solution [82]. The representation of the properties of pure water is only possible by employing after "softer" Mie or Morse potential along with a re-adjustment of the model parameters, as it has been suggested by Chiu *et al.* [159].

Shelley *et al.* [84] developed a CG model of water based on a three-to-one mapping scheme to study the self-assembly of aqueous surfactant solutions. The model is based on the "soft-core" Mie interaction potential characterised by (6-4)-exponents thereby avoiding the problem of the premature freezing. In a subsequent study, He *et al.* assessed several models at various levels of coarse-graining ranging from one to four molecules per bead by using different soft-core Mie or Morse potentials [82]. These CG models are parameterised to reproduce the liquid density, compressibility, and interfacial tension of water at ambient temperature; inevitably it was impossible to describe all three properties simultaneously at this level of coarse-graining. As we will show later in our work these models do not allow one to accurately predict the vapour pressure, vaporisation enthalpy or heat capacity (Table 4.4). Hadley and McCabe [89] have investigated different levels of coarse-graining of water. These authors used $k$-means algorithm to identify the spacial coordinates and the number of molecular clusters $k$ in the system. In this manner, an average of the most appropriate number of water molecules per CG bead could be estimated. Hadley and McCabe concluded that the four-to-one mapping represents the best balance between accuracy and computational efficiency in this case. Freeing themselves of the restriction of a spherically symmetric form of interaction, Molinero and co-workers [286, 300] proposed

the monatomic water (mW) CG model based the on Stillinger-Weber silicon two- and three-body potential that favours a tetrahedral coordination of the atoms. The model was parameterised to match the experimental vaporisation enthalpy, melting point, and density of liquid water at ambient conditions, and with it one retains the capability of describing some of the key structural properties, such as radial and angular distribution functions. Unfortunately, the mW model incorporates three-body contributions, which are costly to compute during the simulations. An additional inherit problem with the aggressive coarse-graining of several molecules into a single bead is that much of the molecular identity is lost; molecular details of interfacial densities and configurations are smeared out and crucially the vapour phase and properties associated with vapour-liquid equilibria become ill-defined.

In our current work, we first focus on the development of a simple isotropic CG model of water based on a one-to-one mapping. We take a leaf out of the book of Klein and co-workers [82] and employ a Mie form of interaction, allowing the exponents to differ from the usual Lennard-Jones (12-6) prescription. The target thermodynamic properties in our case are the saturated liquid density, the vapour pressure and the vapour-liquid interfacial tension of water. The intermolecular parameters of the Mie potential are adjusted to provide an optimal description of the target properties. By contrast to the work of Klein and co-workers [82], we ensure that a good description is obtained for a wide range of temperatures; this is achieved by introducing a temperature dependence for the size and energy parameters of the model. In essence our CG Mie force field for water represents a simple potential of mean force that can be used to faithfully describe the thermodynamic properties over a broad range of gaseous and condensed states. The molecular CG models are illustrated in Figure 4.1. Other levels of coarse-graining are considered within the SAFT-$\gamma$ Mie framework as described in the following section.

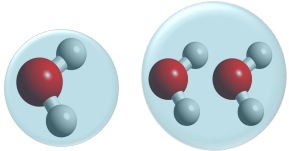


**Figure 4.1:** CG molecular model for water, based on the one-to-one and two-to-one level of coarse-graining.

## 4.3 Methods

### 4.3.1 Molecular simulation details

As mentioned in the introductory section, a simple Mie (generalised Lennard-Jones) isotropic potential form is employed to capture the effective intermolecular interactions between water molecules. The size and energy parameters of the models and the exponents that characterise the softness/ hardness and range of the interactions are parameterised to reproduce target thermodynamic properties of water with the aid of the SAFT-$\gamma$ equation of state [24, 55]. In our current study the target properties are the vapour-liquid coexistence properties and interfacial tension. The fluid-phase equilibria of our SAFT-$\gamma$ Mie CG models of water is determined using Molecular Dynamic simulation in the canonical ensemble, corresponding to a constant number of particles $N$, volume $V$, and temperature $T$. The overall density of the system is chosen such that it lies inside the coexistence envelope. A system of $N$ =5000 water molecules is arranged in an orthorhombic simulation box with the usual periodic boundary conditions, where the box length in $z$ direction is approximately three times longer than that in the $x$ and $y$ directions. In this configuration, a liquid slab of water with two planar interfaces is formed in contact with low density vapour. The simulations are carried out using DL_POLY package, version 2.0 [301], and the equations of motion are solved using the leap-frog algorithm with a time step size of 10 fs. The system temperature is maintained constant using Nosé-Hoover thermostat [188, 189] with coupling constant being 1.0 ps. The initial 300000 time steps are performed to ensure equilibration and the equilibrium properties are sampled for additional 300000 time steps to obtain time averages of the properties of interest.

The densities of the coexisting liquid and vapour are determined from the density profiles at the corresponding temperature. The vapour pressure is calculated as a component of the pressure tensor normal to the interface. The enthalpy of vaporisation is found as the difference of the single-phase enthalpies evaluated for the vapour and liquid phases at the corresponding bulk densities.

### 4.3.2 Interfacial tension by the mechanical and test-area methods

A common method to calculate the interfacial tension in molecular simulation is by means of a mechanical route, which requires the evaluation of forces [302] using Eq. 3.2. It

has been shown that there can be issues with the determination of the pressure or the interfacial tension with the mechanical route for interactions with a (short) cutoff in the potential [303]; the discontinuity in the potential at the cutoff leads to an impulse in the force which has to be taken into account explicitly in order to accurately determine the mechanical properties. In order to ensure that the pressure and interfacial tension is reliably estimated for our CG models of water we employ an alternative method by means of the test-area (TA) thermodynamic route [304]. In the TA method, the tension is computed from infinitesimal perturbations in the interfacial area, thereby inducing changes in the configurational energy and Helmholtz free energy of the system. The cutoff in the potential does not lead to computational issues in methods based on the calculation of the energy of the system. In the canonical ensemble, the surface tension of a planar interface can be obtained from the following thermodynamic definition

$$\gamma = \left(\frac{\partial A}{\partial a}\right)_{NVT} = \lim_{\Delta A \to 0} \left(\frac{\Delta A}{\Delta a}\right)_{NVT}, \tag{4.1}$$

where $A$ is the Helmholtz free energy and $a$ is the interfacial area. To perform a perturbation from the reference system 0 to a perturbed state 1, the box dimension in the $x$ and $y$ in-plane directions are scaled such that $L_{x1} = L_{x0}\sqrt{1+\psi}$ and $L_{y1} = L_{y0}\sqrt{1+\psi}$, respectively, where $\psi$ is a perturbation parameter corresponding to a change in the original area of $\Delta a_{0\to1} = a_0\psi = L_{x0}L_{y0}\psi$. Since the volume of the system has to remain constant, the corresponding change in $z$ normal direction is $L_{z1} = L_{z0}/(1+\psi)$. The molecular coordinates are scaled accordingly in such a way that the relative positions along each axis remain unchanged. The free energy difference $\Delta A_{0\to1}$ between the reference and perturbed states can be estimated in terms of the Boltzmann factor of the corresponding change in the configurational energy according to the Zwanzig perturbation expression [305]:

$$\Delta A_{0\to1} = -\frac{1}{\beta}\ln\langle\exp(-\beta\Delta U_{0\to1})\rangle_0. \tag{4.2}$$

10 The perturbations can involve both an increase (1) or a decrease (-1) in the interfacial area and the interfacial tension can then be obtained from a central difference scheme:

$$\gamma = \frac{A_{0\to1} - A_{0\to-1}}{2\Delta a}. \tag{4.3}$$

Adequate statistics is crucial for an accurate estimate of the surface tension by the TA method: the averages are computed over 1000000 time steps, for configurations every 100 time steps.

Three different values of the perturbation parameter $\psi$ are employed ($\psi$=0.001, 0.0005, and 0.0001) in the perturbation procedure and the results are extrapolated to the limit of $\psi$=0.

In our current study, the surface tension of water is calculated via the mechanical route and validated with the TA method. The values determined for interfacial tension strongly depends on the cutoff radius; if the cutoff radius is not long enough, the results obtained with the mechanical and TA approaches differ from each other, the latter being the more reliable.

## 4.4 Results and discussion

### 4.4.1 The effect of the cutoff of the potential on the thermodynamic properties

Before we discuss the parameterisation of the various models for water in detail it is important to assess the effect of the cutoff of the potential on the vapour-liquid equilibria and interfacial tension. The cutoff radius ($R_c$) is usually introduced in molecular simulation to reduce the computational time by reducing the number of interactions to those between the neighbours within the cutoff distance. The truncation of the interactions is commonly performed either by using spherically truncated (ST) or spherically truncated and shifted (STS) potentials. In MD calculations, this gives rise to a discontinuity in forces (which are the derivatives of the potential) at the cutoff distance. Depending on the molecular simulation software that is employed, long range corrections are treated in different approximate ways giving rise to possible sources of errors. Trokhymchuk and Alejandre [303] have shown that the vapour-liquid surface tension of the LJ fluid increases by $\sim$ 35 % when $R_c$ is increased from $R_c = 2.5\sigma$ to $4.4\sigma$ for the ST model and $\sim$ 60% for the STS model. The change in the density of the coexisting liquid phase is found to be $\sim$ 10 % and 15 % for the ST and STS models, respectively. Similar findings on the importance of the cutoff radius have been described by Wang *et al.* [306], who also employed a mechanical route to determine the surface tension. The interfacial tension is seen to have converged

to that of the full LJ potential for a cutoff radius of about $R_c = 10\sigma$.

The SAFT EoS accounts for a full potential, i.e. an infinite cutoff. However a suitable cutoff has to be imposed in the calculation of the energy and force within a finite simulation cell. It is therefore essential to determine which $R_c$ is appropriate to reproduce the target thermodynamic properties of a full potential for our coarse-grained model of water. As we will see in Section 4.4.2, the Mie (8-6) potential is found to be particularly appropriate to represent the average interactions over the entire phase envelope of water. The Mie (8-6) potential is "softer" than the LJ (12-6) potential, and as a consequence larger cutoffs are expected to be required in order for the thermodynamic properties to converge to a full potential model: the potential takes a value of $u(R_c)/\epsilon = 4.8 \times 10^{-4}$ at $R_c = 5\sigma$ and $u(R_c)/\epsilon = 7.8 \times 10^{-6}$ at $R_c = 10\sigma$. Despite these apparently small energies, the discrepancy in the resulting macroscopic properties is non-negligible.

The effect of the cutoff radius on selected thermodynamic properties of interest is investigated. The vapour-liquid coexistence and interfacial properties obtained from direct Molecular Dynamics simulation of the Mie (8-6) system for values of the cutoff radius ranging from $R_c \sim 5\sigma$ to $12\sigma$ are reported in Table E.1. We start by examining the Mie (8-6) model of water developed in our current work to reproduce the fluid-phase equilibria (Model 1). The densities of the coexisting liquid and the vapour states and the vapour pressure of Model 1, described in detail in Section 4.4.3, are displayed in Figure 4.2 for a temperature of $T = 393$ K. It is apparent that a cutoff radius of at least $R_c = 30$Å $\sim 10\sigma$ is required to reproduce the vapour-liquid coexistence properties of the full potential model.

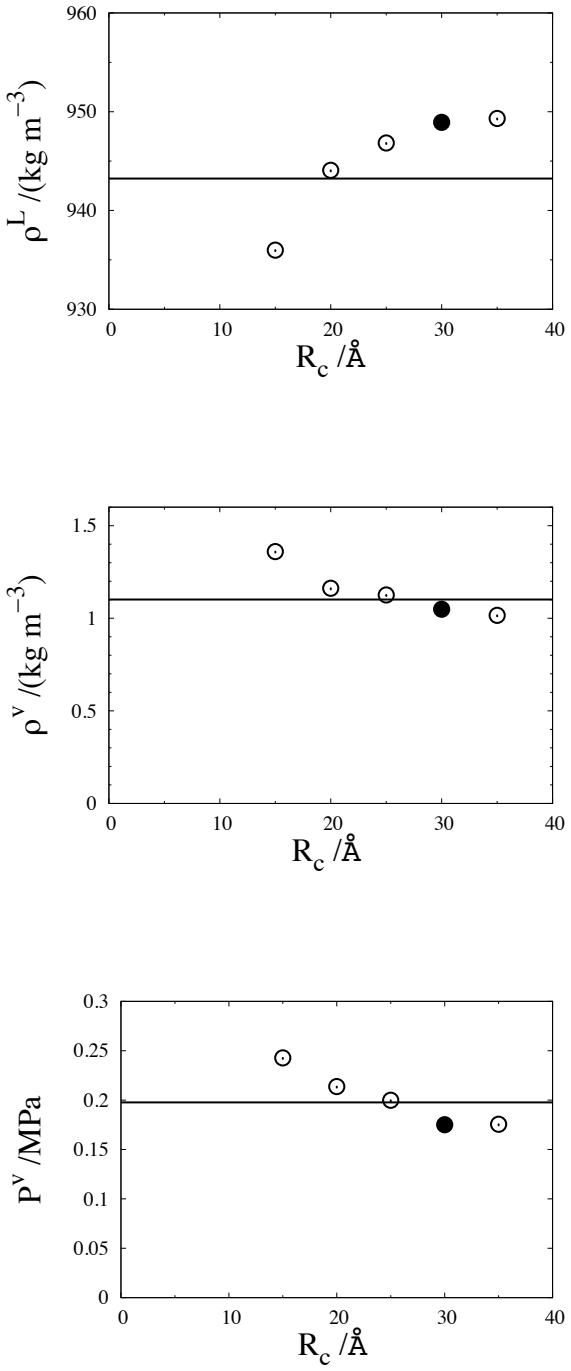



**Figure 4.2:** The vapour-liquid coexistence properties including the saturated liquid and vapour densities, and vapour pressure for the Mie (8-6) Model 1 of water at $T$=393 K as a function of the cutoff radius $R_c$. The continuous black line denotes the results obtained with SAFT-$\gamma$ Mie EoS. The open circles are the corresponding MD simulation data and the filled circles are the results for cutoff radius chosen for Model 1 in the current study.

A second Mie (8-6) model (Model 2) is parameterised to reproduce the saturated liquid density and the surface tension of water. The surface tension of the Model 2, which is described in detail in Section 4.4.4, is presented as a function of the cutoff radius in Figure 4.3. As for other thermodynamic properties, the surface tension is found to converge to those of the "full" potential only for a large cutoff radius ( $R_c \geq 10\sigma$). The converged values obtained with the mechanical and test-area approaches are seen to be identical. The deviations found from the full potential model for short cutoff radii are larger when estimated from the mechanical route. Hence, unless a very large cutoff radius is chosen, it is advisable to compute the surface tension by means of the test-area method, in order to minimise the effects of the truncation of the full potential. In a typical CG simulation of a biological system, the cutoff radius is commonly chosen to be between 9 and 15 Å [78, 85], which is rather short in terms of reproducing the properties of the equivalent full potential. One should bear in mind that the cutoff radius is an important parameter of the model, and it is therefore crucial to use the same value to insure reproducibility of the results.

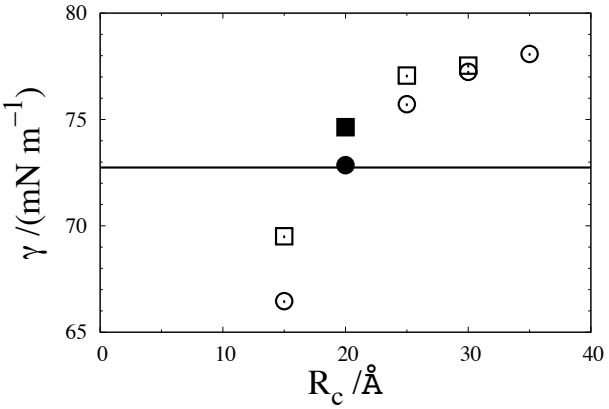



**Figure 4.3:** Vapour-liquid interfacial tension of our Mie (8-6) Model 2 of water at a temperature of T=293 K as a function of the cutoff radius $R_c$. The continuous black line denotes the experimental surface tension of water. The circles are the MD simulation data obtained with the mechanical route, and the squares are the corresponding data obtained with the test-area method; the filled symbols denote the values for the cutoff radius chosen for Model 2 in the current study.

### 4.4.2 Transferability and representability issues

#### Transferability

It is well known (cf. Section 2.6.2) that coarse-graining of a intermolecular potential leads to issues of representability and transferability. Problems in the transferability of the model occur if an effective potential developed to reproduce a given state point does not perform well at another state point, so that the potential has to be refined for the new state point. As was mentioned earlier in the thesis, the potential of mean force $U_{CG,T}(\mathbf{R})$ (Eq. 2.71) is necessarily a function of the thermodynamic state and thereby depends on the temperature and density. In contrast to the weakly polar chain molecules studied in Chapter 3, this effect is much more pronounced for a strongly polar fluid with highly directional interactions such as water [8, 208] since many-body interactions play a significant role and cannot be effectively averaged over all thermodynamic states. The averaging of atomistic interactions into effective interactions for a CG bead leads one to neglect some important degrees of freedom. The Mie potential (like the LJ potential) belongs to the group of spherically symmetric (isotropic) potentials. The intermolecular forces which are responsible for the complex behaviour exhibited by cannot be fully captured with a spherically symmetric potential that is capable of reproducing its physical properties at different conditions; a universal parameter set fails to reproduce various target properties if a wide range of thermodynamic conditions are considered. For example, if we use the parameter set of the Mie (8-6) potential with $\sigma$=3.0085 Å and $\epsilon/k_B$=497.84 K, developed to provide a good description of the saturated liquid density and vapour pressure of water at ambient temperature, the prediction of the vapour-liquid equilibria is very poor at high temperatures as the critical temperature and liquid densities are markedly overestimated and the vapour pressure is underestimated (cf. Figure 4.4).

To overcome the problem of transferability of the intermolecular potential, we consider temperature dependent segment size and energy parameters. Our effective intermolecular potential thus can be considered as a free energy (or potential of mean force). The re-parametrisation of the potential in this manner is usually a very inefficient procedure involving a number of iterative simulations. The use of the algebraic SAFT-$\gamma$ Mie equation of state in this regard greatly facilitates the process as a wide range of thermodynamic states can be considered for a fraction of the computational cost.

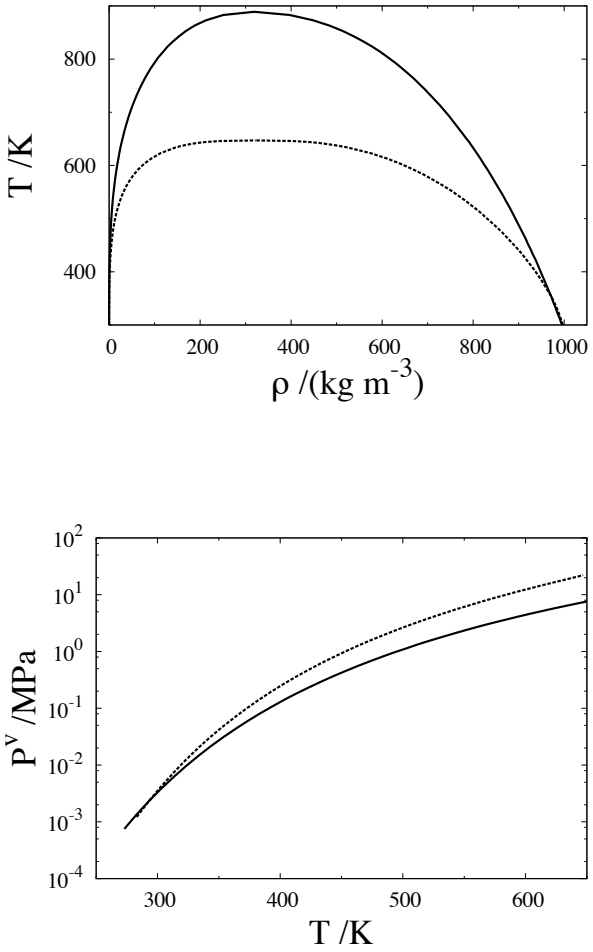



**Figure 4.4:** The saturated liquid density and vapour pressure of water, predicted with a universal parameter set, estimated from the corresponding properties of the state at a temperature of 298 K: the model is a Mie (8-6) potential with $\sigma = 3.0085$ Åand $\epsilon/k_B$=497.84 K. Continuous curves denote the experimental values from NIST [195] and dashed curves give the prediction with SAFT-$\gamma$ Mie EoS [24, 55] with the parameters obtained for a unique state point.

### Representability

The issue of representability is associated with the fact that a given coarse-grained potential cannot be used to simultaneously represent all the properties of the system at the given state point. A spherically symmetric Mie potential which averages out all of the atomistic interactions into an effective CG interactions is characterised by four parameters: attractive and repulsive exponents, size and energy parameters. A central question

is how many properties can be captured by tuning the four parameters to experimental data. If one assumes that one degree of freedom allows one to determine a given property, it is useful to assess which parameter determines which property.

Considering the first term (mean-attractive energy) (cf. Eq. A.13) in the Barker and Henderson perturbation expansion [178] for a Mie fluid in the mean-field limit (corresponding to a uniform structure $g^{HS}(r) = 1$), we obtain

$$a_1 = 2\pi\rho\sigma^3\epsilon\alpha, \tag{4.4}$$

where a dimensionless van der Waals constant $\alpha$ can be defined as

$$\alpha(\lambda_a, \lambda_r) = \frac{1}{\epsilon\sigma^3}\int_\sigma^\infty u(r)r^2dr = \mathcal{C}(\lambda_a, \lambda_r)\left[\left(\frac{1}{\lambda_a - 3}\right) - \left(\frac{1}{\lambda_r - 3}\right)\right], \tag{4.5}$$

with $\mathcal{C}(\lambda_a, \lambda_r)$ defined in Eq. 2.18.

At the mean-field level at least, the mean-attractive energy is therefore a function of only three parameters $\epsilon$, $\sigma$, and $\alpha(\lambda_a,\ \lambda_r)$, the later of which is itself a function of the attractive and repulsive exponents. As a consequence, two fluids with the same $\sigma$, $\epsilon$, and $\alpha$ will exhibit the same thermodynamic properties at this level of approximation. Fluids with the same value of $\alpha$ have a conformal intermolecular potential $u(r)$ [53, 60], implying that the exponents $\lambda_a$ and $\lambda_r$ are not independent and provide only one additional degree of freedom.

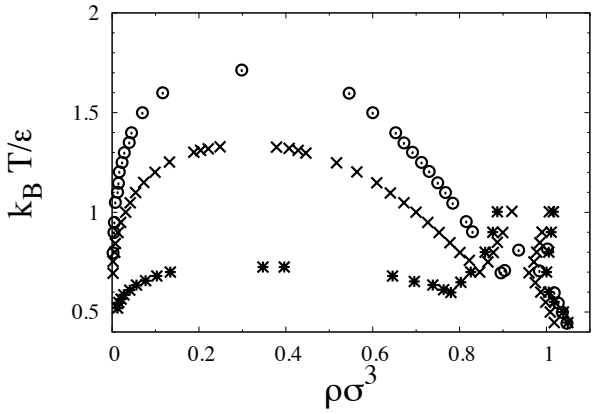



**Figure 4.5:** Temperature-density phase envelops for Mie particles with different values of the van der Waals constant $\alpha$ including the fluid and solid phases. The stars denote $\alpha = 0.40$ (Mie 19-8.8), the crosses represent $\alpha = 0.89$ (Mie 12-6), and the circles are $\alpha = 1.26$ (Mie 8-6). Adopted from Ramrattan *et al.* [307].

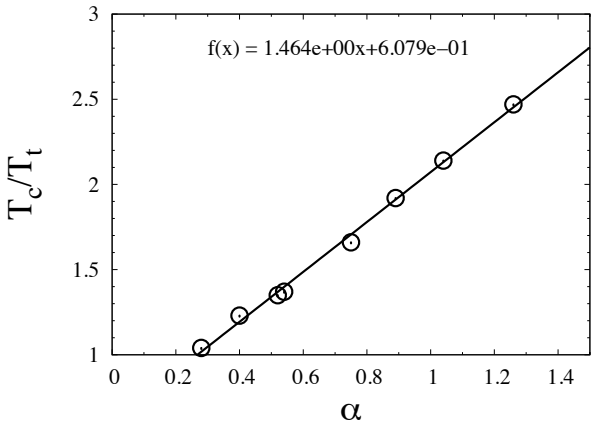



**Figure 4.6:** The ratio of the critical and triple point temperatures as a function of the van der Waals constant $\alpha$ corresponding to different Mie fluids. Adopted from Ramrattan *et al.* [307]

Due to conformality of the potentials, once the value of the energy parameter $\epsilon$ is fixed, for a given set of exponents, both the critical temperature $T_c$, and the triple point $T_t$ are then implicitly fixed. The corresponding correlation between the ration of the critical to the triple point temperatures and the van der Waals attractive parameter $\alpha$ can be expressed

as $T_c/T_t = 1.464\alpha + 0.6079$ [307]. For the common (12-6) combination of the LJ potential, the fluid range is given by $T_c/T_t = 1.9$, since for this fluid $T_c^* = 1.3$ and $T_t^* = 0.7$ [307]. One could guarantee a sensible description of the critical point by matching the experimental critical temperature ($T_c = 647$ K) to the model with $T_c^* = k_B T/\epsilon = 1.3$, suggesting an $\epsilon = 498$ K. This choice then fixes the triple point of the model at $T_t = 348$ K (rather than the experimental melting point of 273 K), rendering any simulation at ambient temperature prone to premature freezing. Conversely, fixing the triple point gives a cohesive interaction which is too week compared to experiment and a fluid range which is unsatisfactorily small. A sensible choice of the exponents for water is one that would match the experimental ratio of $T_c/T_t = 647/273 = 2.4$ is the one that guarantees $T_c/T_t = 2.4$, which according to Figure 4.6 would correspond to a van der Waals integrated energy of $\alpha \sim 1.22$. There are of course an infinite number of pairs of exponents that satisfy this condition (c.f. Eq. 4.5). Retaining the London form of the attractive contribution, we chose the relatively soft potential with the pair of exponents (8-6), corresponding to a value of the van der Waals constant of $\alpha = 1.264$ which is close to that expected for water from our simple scaling analysis. This result is consistent with empirical observations made by He *et al.* [82], who assessed a number of models and recommend the use of (9-6) and (12-4) Mie potentials for one-to-one CG description of water, which would correspond to $\alpha = 1.125$ and $\alpha = 2.309$, respectively.

If one simply frees all of the parameters available as degrees of freedom with the Mie potential (i.e., exponents, size, and energy parameters) to optimise the description of a maximum number of target thermodynamic properties with the SAFT-$\gamma$ Mie EoS, the model corresponding to Mie (40-6) potential with $\sigma = 3.1065$ Å and $\epsilon/k_B = 777$ K is found to provide a good representation of the critical temperature, saturated liquid densities, and vaporisation enthalpies over a range of temperatures. Unfortunately, the model with such a repulsive interaction and deep energetic well would correspond to a van der Waals constant of $\alpha = 0.50$ implying that the system would exhibit a very narrow liquid range and marked premature freezing with a triple point of $T_t \sim 483$ K.

The various intermolecular potentials discussed in the context of a coarse-grained representation of water are compared in Figure 4.7. The issue of the premature freezing can be explained by assessing the well-depth and the overall shape of the potential (particularly the steep repulsive region). The deepest well and therefore the highest melting

point is expected for Mie (40-6) model, followed by the Martini model. Our Mie (8-6) Model 1 (parameterised to optimise the description of the fluid-phase equilibria of water, as described in the following section) and the standard LJ model have a comparably large well-depth, and also give rise to freezing at ambient conditions: the Mie (8-6) potential is "softer" and exhibits a more distinct attraction range than the LJ model. Our Mie (8-6) Model 2 (parameterised to optimise the description of the saturated-liquid density and vapour-liquid surface tension) has a form which is nearly identical to the Mie (9-6) and (12-4) models suggested by He *et al.* [82], which is reassuring as the same properties of water are being targeted: liquid density and interfacial tension. The relatively soft nature of the potential and its shallow depth ensures that Model 2 remains in a stable liquid state throughout the experimental liquid fluid range. An entirely different form of potential is obtained from the Iterative Boltzmann Inversion (IBI) of the structure of the SPC water model [227]. Unlike the models based on a CG potential characterised by a single well, the CG IBI potential is seen to exhibit multiple wells that account for the short-range hydrogen bonding and the long-range attractive interactions; the relatively shallow well means that the model will not suffer from premature freezing. It is important to point out, however, that the aim of a IBI coarse-graining methodology is principally to reproduce the structural properties of fluid; the thermodynamic properties are not considered explicitly with this approach and cannot be accurately reproduced [283]. A further disadvantages of the IBI is that the potential is not based on the close algebraic form and instead has to be obtained with an iterative procedure at each state point, making it computationally demanding.

After having chosen our designated Mie (8-6) form of the potential for our coarse-grained representation of water, the remaining size and energy parameters can be estimated by matching the appropriate target properties of water. A key quantity that has been a target for the majority of models is the (saturated) liquid density. The absolute value of the density of a model is closely linked to the magnitude of the size parameter $\sigma$. As the representation liquid density is generally a prerequisite this further reduces the degrees of freedom of the model, so that only the energy scale $\epsilon$ can be refined to capture the remaining properties. In the next section we will show that an attempt to match a given target property (e.g., the enthalpy of vaporisation) can lead to a significant deterioration in the prediction of another properties (e.g., interfacial tension). A compromise to reproduce

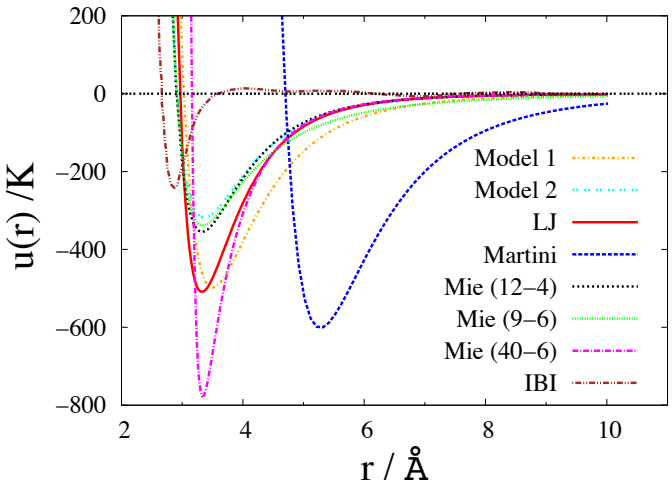



**Figure 4.7:** The various CG intermolecular potentials used to represent water at 298 K: Model 1, Mie (8-6) potential, developed with the SAFT-$\gamma$ Mie EoS to represent the saturated liquid density and vapour pressure; Model 2, Mie (8-6) potential, developed with the SAFT-$\gamma$ Mie EoS to represent the saturated liquid density and vapour-liquid interfacial tension; the Mie (40-6) potential developed with the SAFT-$\gamma$ Mie EoS as a compromise to reproduce the saturated liquid density, the vapour pressure, enthalpy of vaporisation, and vapour-liquid interfacial tension to a lower level of accuracy, freezes at ambient conditions; the Mie (9-6) and (12-4) potentials proposed by He *et al.* [82] capture the saturated liquid density and the vapour-liquid interfacial tension but fail to describe the vapour pressure and the enthalpy of vaporisation; the Martini model [78], which corresponds to a simple LJ (12-6) potential form at the four-to-one molecules per bead level of coarse-graining, captures the liquid density but freezes prematurely; the LJ (12-6) potential at the one-to-one level of coarse-graining reproduces the liquid density but fails to describe the vapour pressure and enthalpy of vaporisation, and freezes at ambient conditions; the IBI model is obtained by mapping the structural properties of SPC model to a CG bead using iterative Boltzmann inversion, without a consideration of the thermodynamic properties.

several properties simultaneously with a lower degree of accuracy is found to be unsatisfactory. Thus, two single-site Mie (8-6) CG models of water are developed in our current work with the aid of the SAFT-$\gamma$ Mie equation of state: the first Model 1 is parameterised to reproduce the saturated liquid density and the vapour pressure; the second Model 2 to reproduce the saturated liquid density and the vapour-liquid interfacial tension.

### 4.4.3 Model 1 with the liquid density and vapour pressure as target properties

The first Mie (8-6) model considered (Model 1), is designed to provide an optimal description of the experimental saturated liquid density and vapour pressure for the entire fluid phase range. To that effect, the size and energy parameters $\sigma$ and $\epsilon$ are estimated with the SAFT-$\gamma$ Mie EoS by minimising the difference between the experimental and theoretical values of these properties at each temperature. The parameter estimation procedure is undertaken at a each state point, so that there is a parameter set for each temperature studied. Temperature dependent parameters are reported in Table C.3 in Appendix and plotted in Figure 4.8. Simple correlations for $\sigma(T)$ and $\epsilon(T)$ of our Mie (8-6) Model 1 of water (for use with the relatively long cutoff of $R_c = 30$ Å, which essentially corresponds to the full potential) can then be developed as simple polynomial functions of the temperature:

$\sigma/\text{Å} = 1.262 * 10^{-9}T^3 - 8.72 * 10^{-8}T^2 - 4.554 * 10^{-4}T + 3.119$, and

$\epsilon/k_b/K = 1.105 * 10^{-5}T^2 - 0.3077T + 586.8$.

It can be seen that $\epsilon$ is characterised by a near-linear behaviour with the temperature, while $\sigma$ exhibits a minimum at a temperature of $\sim$ 350 K.

The corresponding temperature-density vapour-liquid envelope for the Mie (8-6) Model 1 is displayed in Figure 4.9. The triple point was determined by Ramrattan *et al.* [307] as the intersection of the bubble point curve for the liquid with the solidification curve at $T_t^* = 0.71$. One observes that the state at 273 K, which corresponds to the experimental triple point ($T_t$) of water, is below the triple point estimated for the Mie (8-6) CGW1-vle ($T_t^* = T/(\epsilon/k_B) = 0.54$), suggesting the triple point of the model is

$T_t$(CGW1-vle)$\sim$ 343 K and hence premature freezing. This overestimation is unavoidable because of the high values of the energy parameters which are required for an accurate description of the vapour-pressure curve over the entire fluid range. The high values of well-depth $\epsilon$ are attributed to the effective averaging of the hydrogen bonding in our isotropic coarse-grained force field; the effective attractive energy is seen to decrease with increasing temperature which is consistent with corresponding breaking of hydrogen bonds expected from entropic considerations. The physically meaningful temperature range for our Mie (8-6) Model 1 is therefore from the triple point of the model to the near-critical

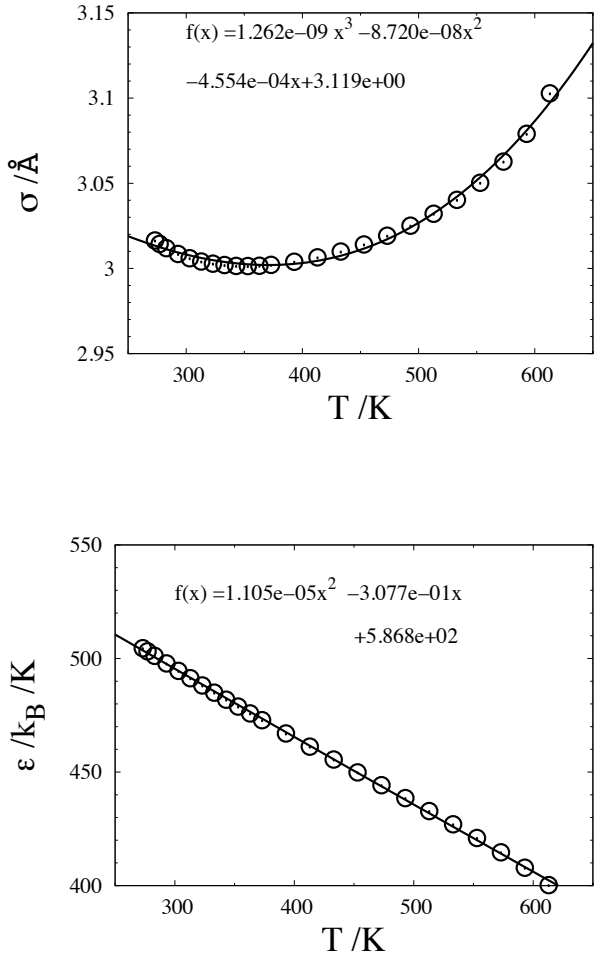



**Figure 4.8:** The temperature dependence of the size and energy parameters $\sigma$ and $\epsilon$ for Mie (8-6) Model 1 of water developed to reproduce the experimental saturated liquid density and vapour pressure, estimated with SAFT-$\gamma$ Mie EoS

region (343-643 K).

In order to retain a close link with the theory and faithfully represent full potential, a relatively large value of $R_c$ = 30 Å (corresponding to $R_c \sim 10\sigma$ for our models of water) is chosen for the cutoff radius as indicated in Figure 4.2. Overall, the accuracy of the prediction of the fluid-phase equilibria with the SAFT-$\gamma$ EoS [24] for the Mie (8-6) system compared to the MD simulation data for the full potential ($R_c$=30 Å) (reported in Table E.2 in Appendix) is within 1% for the saturated liquid density and 4% for the vapour pressure as it is apparent from Figure 4.10. As our Model 1 is parameterised to reproduce the saturated liquid density and vapour pressure at each temperature along the

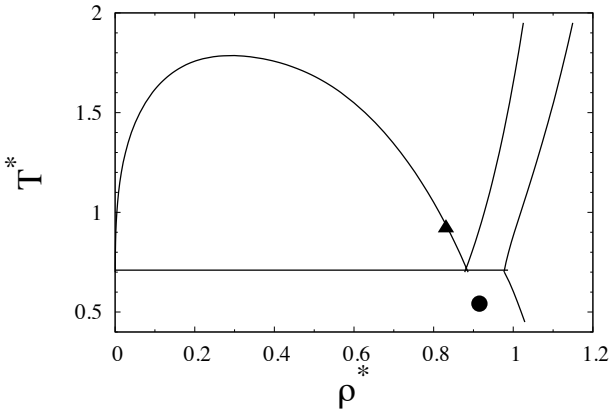



**Figure 4.9:** Adopted from Ramrattan *et al.* [307]. The temperature-density coexistence envelope in dimensionless units $T^* = k_B T/\epsilon$ and $rho^* = \rho\sigma^3$ for the Mie (8-6) system is given by the continuous curves, the filled circle indicates the state point corresponding to 273 K for Model 1 and filled triangle is the state point corresponding to 273 K for the Model 2.

vapour-liquid envelope, the prediction of other properties can unfortunately deviate from experimental values due to the aforementioned issues of representability. For instance, the enthalpy of vaporisation is underestimated by about 14 % and the vapour-liquid interfacial tension is overpredicted by about 100% at low temperatures (see Figure 4.11). A major drawback of Model 1 is however its high melting point which causes an unphysical freezing at ambient conditions.

Notwithstanding these deficiencies, the Model 1 is suitable for the description of any type of fluid-phase equilibria including the challenging phase behaviour of aqueous mixtures, especially at high temperatures and pressures, since the density and vapour pressure of the system are accurately reproduced. The one-to-one mapping can be used to account for the behaviour of water molecules in the vapour phase in a reasonably realistic manner, in contrast to CG models of water involving more than one water molecule per bead, which imply an unphysical clustering in the vapour phase. The performance of the Model 1 for binary aqueous systems is illustrated for the mixtures with carbon dioxide and $n-$alkanes in Chapter 5.

In the study of biological systems, most of the relevant phenomena, such as the phase morphologies of aqueous solutions of amphiphilic molecules or the configurations of macromolecular structures, take place in the liquid phase. In order to best capture the physical

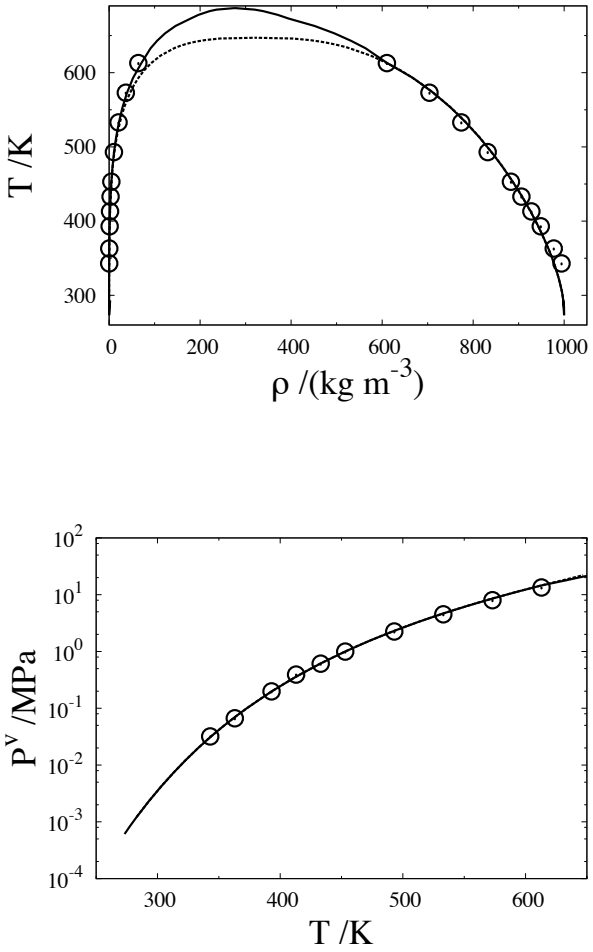



**Figure 4.10:** The saturated liquid density and vapour pressure of water as represented with our Mie (8-6) Model 1. The dashed curves denote the correlated experimental values from NIST [195], the continuous curves are the description with SAFT-$\gamma$ Mie EoS [24, 55] (indistinguishable from the experimental values of the vapour pressure), and the symbols are the corresponding data obtained for the system by Molecular Dynamics simulation for the model.

properties of these systems, an accurate description of the interfacial tension is crucial, as the phase morphologies of the microphase separated domains are very sensitive to details of the interfacial properties [85]. Therefore, we propose an alternative parametrisation focussing of the accurate reproduction of the liquid density and surface tension.

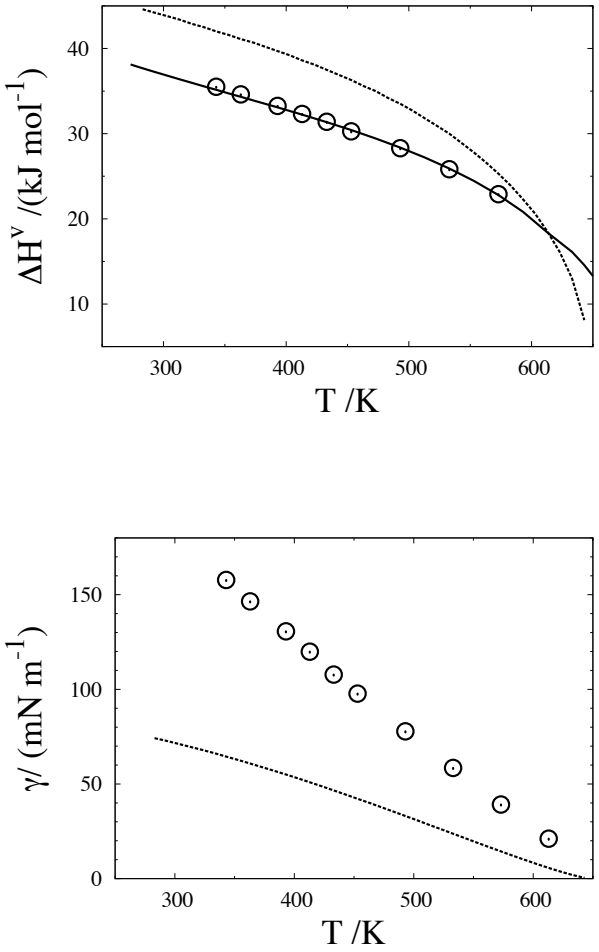



**Figure 4.11:** The enthalpy of vaporisation and vapour-liquid interfacial tension of water as represented with our Mie (8-6) Model 1. The dashed curves denote the correlated experimental values from NIST [195], the continuous curves are the description with SAFT-$\gamma$ Mie EoS [24, 55] and the symbols are the corresponding data obtained for the system by Molecular Dynamics simulation for the model.

### 4.4.4 Model 2 with the saturated liquid density and vapour-liquid interfacial tension as target properties

For our Model 2 of water, the exponent pair characterising the Mie potential is kept as (8-6) for the sake of consistency, and the size and energy parameters $\sigma$ and $\epsilon$ are estimated directly from the molecular simulation data data for the saturated liquid density and the vapour-liquid interfacial tension at each temperature, respectively, since the prediction of the latter is not directly accessible from the equation of state (unless a suitable treatment

of the inhomogenous properties of the system is made [308–310].

The parameterisation is undertaken to match the saturated liquid density and vapour-liquid interfacial tension (as determined with the mechanical route), using the cutoff radius of $R_c$= 20 Å as a compromise between an accurate representation of a "full" potential (corresponding to $R_c$= 30 Å) and the computational effort associated with the long-ranged interactions; the corresponding simulation data for $R_c$= 20 Å is reported in Table E.3 in Appendix. The vapour-liquid interfacial tension obtained with the mechanical and thermodynamic routes for $R_c$ = 20 Å are indicated on Figure 4.3 (cf. discussion in Section 4.4.1).

The temperature-dependent parameter set for $\sigma$ and $\epsilon$ of our Mie (8-6) Model 2 is given in Table C.4 in Appendix and plotted in Figure 4.12. Corresponding correlations for $\sigma(T)$ and $\epsilon(T)$ of our Mie (8-6) Model 2 of water (for use with the cutoff of $R_c$= 20 Å) can also be developed as simple polynomial functions of the temperature:

$\sigma/\text{Å} = -6.455 * 10^{-9}T^3 + 9.1 * 10^{-6}T^2 - 4.291 * 10^{-3}T + 3.543$ and

$\epsilon/k_b/K = -4.806 * 10^{-4}T^2 + 0.6107T + 165.9$.

The range of values of the parameters at ambient conditions is similar to that of the one-to-one coarse-grained Mie (12-4) or (9-6) models of water developed by He *et al.* [82], which were also parameterised to reproduce the liquid density and surface tension. Because of the relatively low values of the well-depth $\epsilon$, the system remains in the liquid state over the whole experimental fluid temperature range considered. In the global phase diagram, the state point of the model system corresponding to a temperature of 273 K is above the experimental triple line (cf. Figure 4.9); the triple point of Model 2 corresponds to a temperature of $T_t$ = 193 K in real units. The saturated liquid density and the surface tension obtained by direct Molecular Dynamics simulation are in excellent agreement with the experimental data over the entire fluid temperature range as can be seen from Figure 4.13. As discussed earlier, there can be problems of representability with coarse-grained models of this type for properties not employed in the parameterisation: at the low temperatures, the enthalpy of vaporisation and vapour pressure predicted with Model 2 deviates significantly from the experimental data (Figure 4.14). Our Mie (8-6) Model 2 is therefore designed for studies of the condensed liquid state and interfacial properties of aqueous solutions at the moderate temperatures, e.g., for the research area on aqueous solutions of surfactants, biological systems, and macromolecules in general.

The performance of Model 2 as a solvent for aqueous solutions of non-ionic surfactant systems is illustrated in Chapter 6.

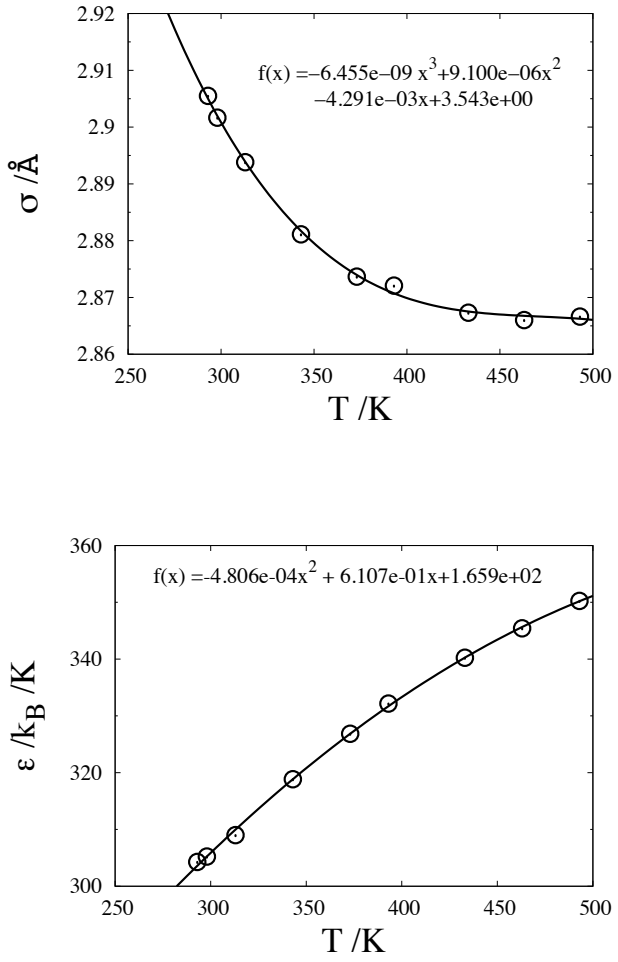



**Figure 4.12:** The temperature dependence of the size and energy parameters $\sigma$ and $\epsilon$ parameters of our Mie (8-6) Model 2 of water developed to reproduce the experimental saturated liquid density and vapour-liquid interfacial tension as obtained from the corresponding simulation data for the model.

### 4.4.5 Different levels of coarse-graining of water

In this section, we explore different levels of coarse-graining for the Mie (8-6) model of water, ranging from one to four molecules per bead. In previous work, He *et al.* [82] observed that at the higher levels of coarse-graining the water models give rise to unphysical crystalline states at ambient conditions when parameterised to reproduce liquid density

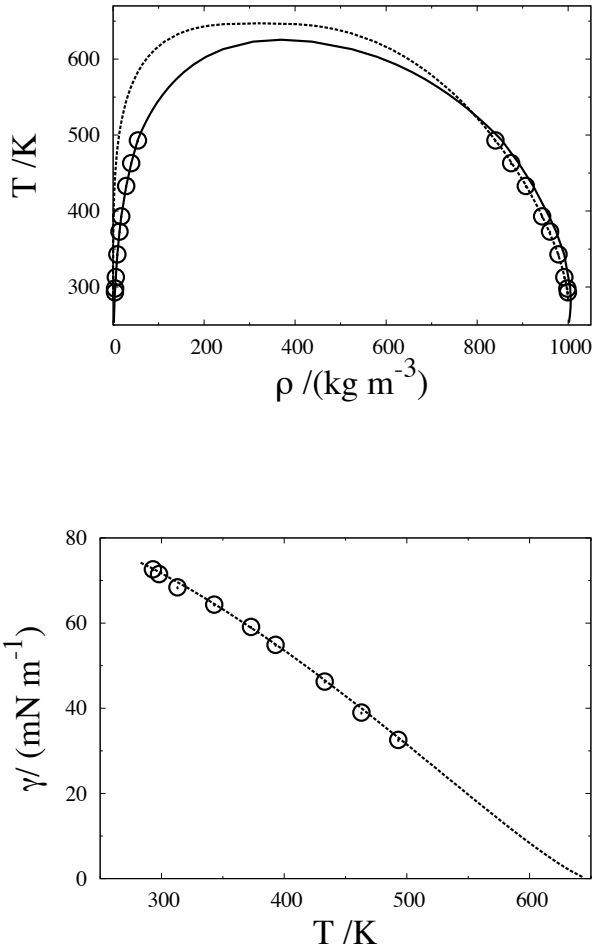



**Figure 4.13:** The saturated liquid density and surface tension of water as predicted with our Mie (8-6) Model 2. The dashed curves denote the correlated experimental values from NIST [195], the continuous curve are the prediction with SAFT-$\gamma$ Mie EoS [24, 55], and the symbols are the corresponding data obtained for the system by Molecular Dynamics simulation for the model.

and surface tension. For our CG models, the melting temperature can be estimated from the principle of conformality ($T_t^* = T_t/\epsilon$) [307]; the triple point for the Mie (8-6) model is at $T_t^* = 0.71$ as indicated on Figure 4.9 and the energy parameter can be obtained from the representation of the liquid density and surface tension as explained in the previous section for the simple one-to-one mapping. The parameters obtained in this manner for different levels of coarse-graining are reported in Table 4.1 together with the corresponding melting points. It is apparent that a consequence of coarse-graining more than two water

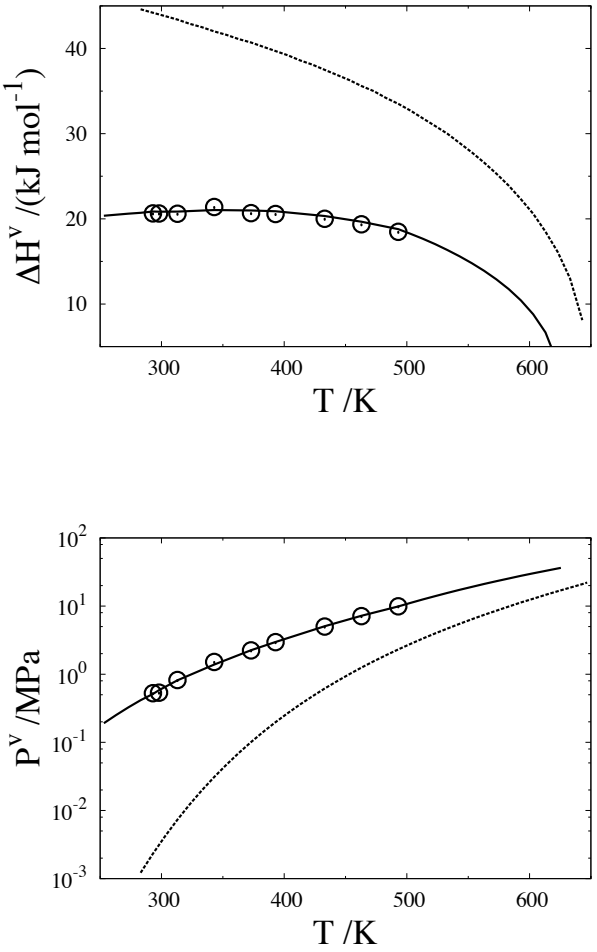



**Figure 4.14:** The enthalpy of vaporisation and vapour pressure of water as predicted with our Mie (8-6) Model 2. The dashed curves denote the correlated experimental values from NIST [195], the continuous curve are the prediction with SAFT-$\gamma$ Mie EoS [24, 55], and the symbols are the corresponding data obtained for the system by Molecular Dynamics simulation for the model.

molecules per bead is a marked overestimation of the melting point (at least in the case of the Mie (8-6) model).

**Table 4.1:** Mie (8-6) models of water (with a cutoff of $R_c = 20$ Å) at different molecule-to-bead levels of coarse-graining developed to reproduce the saturated liquid density and vapour-liquid interfacial tension at $T = 298$ K with the corresponding melting temperatures.

| CG level | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\sigma$ /Å | 2.9017 | 3.7467 | 4.3310 | 4.8295 |
| $\epsilon/k_B$ /K | 305 | 400 | 481 | 573 |
| $T_m$ /K | 199 | 284 | 341 | 407 |

The coarse-graining level should therefore not be chosen higher than two water molecules per bead to avoid unphysical freezing at ambient temperatures. The best choice for the large scale simulation of aqueous surfactants or macromolecular biological systems would be 2-to-1 mapping. In this regard a Mie (8-6) model of water at the two-to-one level of coarse-graining is developed, our "biowater model", which is parameterised to reproduce the saturated liquid density and surface tension at ambient conditions. A unique set of temperature-independent size and energy parameters $\sigma$ and $\epsilon$ is employed for temperature range between 293 and 313 K, unavoidably leading to small deviations in liquid densities and surface tensions at the limits of the given temperature range (Table 4.2). The model will be particularly useful for large scale simulation of aqueous surfactants or biomolecular systems to improve the computational efficiency. A further advantage is that the vapour pressure decreases on increasing the level of coarse-graining, so that the vapour pressure of the two-to-one CG model is improved by one order of magnitude compared to the CG models based on a one-to-one mapping. The "biowater" model has already been successfully implemented in the study of the air-water interfacial tensions and surface excess adsorption of dilute non-ionic surfactant solutions [311], a challenging task owing to the low critical micelle concentrations exhibited by the system.

**Table 4.2:** Biowater model, with a two-to-one molecule-to-bead mapping, parameterised to reproduce the saturated liquid density and vapour-liquid interfacial tension at ambient temperatures.

| T /K | 293 | 298 | 303 | 313 |
|---|---|---|---|---|
| $\sigma$ /Å | | 3.7467 | | |
| $(\epsilon/k_B)$ /K | | 400 | | |
| $\gamma/(\mathrm{mNm}^{-1})$ | 73.43 | 71.88 | 71.12 | 69.17 |
| $\rho_l/(\mathrm{kgm}^{-3})$ | 1003.89 | 999.18 | 995.47 | 986.31 |
| $P^v$/bar | 0.217 | 0.303 | 0.347 | 0.479 |
| | Expt. | | | |
| $\gamma/(\mathrm{mNm}^{-1})$ | 72.74 | 71.97 | 71.19 | 69.6 |
| $\rho_l/(\mathrm{kgm}^{-3})$ | 998.16 | 997.00 | 995.6 | 992.17 |
| $P^v$/bar | 0.023 | 0.031 | 0.042 | 0.074 |

### 4.4.6 Comparison of the Mie (8-6) CG models of water with existing models

As a final assessment of the various force fields for water we compare the performance of our Mie (8-6) CG models with some of the popular atomistic and other CG models. The molecular parameters for the various models are summarised in Table 4.3.

The computational performance of our Mie (8-6) Model 2 in typical Molecular Dynamics simulation is compared to the atomistic SPC model in Figure 4.15. As evident from the figure, the representation of the water molecule as a single coarse-grained bead with no explicit electrostatics leads to compelling savings in computational time. The expected gain in computational efficiency is about one order of magnitude in terms of the system size (for a fixed CPU time per time step) and about two orders of magnitude in the simulation speed-up for a fixed system size. This performance makes our CG model attractive for calculations of large systems over long times.

A comparison of the accuracy in representing the thermodynamic properties with the various atomistic and CG models of water at room temperature ($T$ = 298 K) and at a high temperature ($T$ = 450 K) is presented in Table 4.4: our Mie (8-6) CG Models 1 and 2 and the "biowater" model are assessed along with the Martini [78] and Klein [82] CG models, as well as the popular atomistic models (namely, models from TIP [76] and SPC [312] families). One should recall at this point that $T$ = 298 K is below the

**Table 4.3:** Parameter sets for the atomistic and coarse-grained water models, including the size $\sigma$ and energy $\epsilon$ parameters, the repulsive and attractive exponents $\lambda_r$ and $\lambda_a$ of the Mie potentials, and the proton charges $q_H$ of the atomistic models.

| Water models | $\sigma$ /Å | $\epsilon/k_B$ /K | $\lambda_r$ | $\lambda_a$ | $q_H/e$ |
|---|---|---|---|---|---|
| CG models | | | | | |
| Model 1 (T=298 K) | 3.0089 | 496.09 | 8 | 6 | |
| Model 1 (T=450 K) | 3.0114 | 450.57 | 8 | 6 | |
| Model 2 (T=298 K) | 2.9016 | 305.21 | 8 | 6 | |
| Model 2 (T=450 K) | 2.8666 | 343.39 | 8 | 6 | |
| Martini | 4.7 | 601.36 | 12 | 6 | |
| Biowater | 3.7467 | 400 | 8 | 6 | |
| He *et al.* Mie (9-6), | 2.908 | 354.77 | 9 | 6 | |
| atomistic models | | | | | |
| TIP3P | 3.1506 | 76.54 | 12 | 6 | 0.417 |
| TIP4P | 3.1540 | 78.02 | 12 | 6 | 0.52 |
| TIP4P/2005 | 3.1589 | 93.20 | 12 | 6 | 0.5564 |
| SPC | 3.1656 | 78.20 | 12 | 6 | 0.41 |
| SPC/E | 3.1656 | 78.20 | 12 | 6 | 0.423 |

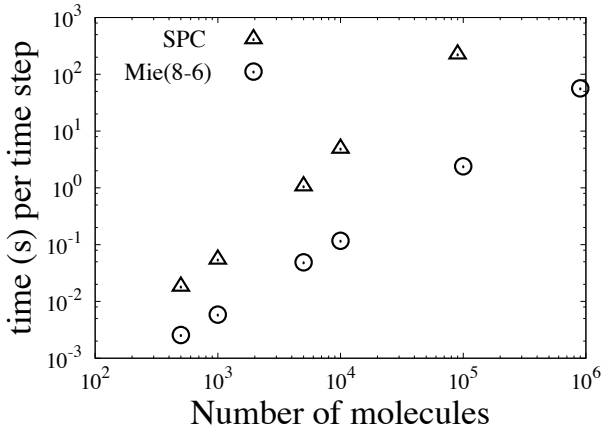



**Figure 4.15:** Comparison of the CPU time per time step in typical Molecular Dynamics simulations as a function of number of molecules for our Mie (8-6) Model 2 and the atomistic SPC water model. The state point is at 298 K. The simulations are performed in the canonical $NVT$ ensemble on 8 CPU 3 GHz processors employing the DL_POLY 2.0 package [301].

triple point of Model 1; this represents a metastable fluid state which we asses by means of the SAFT-$\gamma$ Mie equation of state [24, 55]. The same applies to the Martini CG

force-field of water, where the calculation of properties of the pure liquid at ambient temperature is only possible following the modification proposed by Chiu *et al.* [159]. We examine representative thermophysical properties including the saturated liquid density, vapour pressure, enthalpy of vaporisation, vapour-liquid interfacial tension, isobaric heat capacity, isobaric thermal expansivity, isothermal compressibility, self-diffusion coefficient, and the second virial coefficient in Table 4.4. The corresponding ambient melting points and the extent of the vapour-liquid phase envelope measured in terms of the ratio of the critical and the melting temperatures are also compared. The results for the atomistic models are taken from the corresponding literature [205, 230, 234, 235, 291, 313–316]. For the CG models, the second derivative thermodynamic properties are estimated using the SAFT-$\gamma$ Mie EoS. The melting points are also obtained from a corresponding states analysis as described in Section 4.4.2. The boiling points are estimated from the Clausius-Clapeyron equation; the critical points are obtained from the scaling laws (Eqs. 3.3 and 3.4). All other properties are calculated by means of molecular simulation. As one would expect, the performance of the high-level resolution atomistic models of water is generally better than for the CG models. It is however still important to note that, none of the atomistic models can reproduce all of the properties with a high accuracy despite a much more significant computational costs, and the description with our CG models is very satisfactory considering their simple nature.

The isobaric thermal expansivity of water, which corresponds to the temperature derivative of the volume at constant pressure, for our Mie (8-6) Model 1 is found to exhibit a better performance than most of the atomistic models. The isothermal compressibility is however largely underestimated with our CG models, as is heat capacity, which is underestimated by approximately 33 %.

The self-diffusion coefficient determined from the mean-square displacement by Molecular Dynamics simulation of our CG models and the various atomistic models is represented as a function of the inverse temperature in Figure 4.16. It is well known that the self-diffusion coefficient is overestimated relative to the experimental value for most CG models; the mobility of the CG beads is faster because the molecule is not slowed down by the re-orientation of the hydrogen atoms and the formation / break-up of hydrogen bonds [286]. Our Mie (8-6) Model 2 indeed conforms with this behaviour: the diffusion coefficients at low temperatures are markedly overestimated and become comparable with experiment

only in the high-temperature limit. By contrast, the self-diffusion coefficient determined with our Mie (8-6) Model 1 is underestimated relative to experiment; this could be attributed to the large values of the energetic well of the potential which results from the use of the vapour pressure as the target property. The use of the vapour pressure to estimate the effective cohesive forces holding CG molecules together leads to interactions which appear to be stronger than the real forces, reducing the mobility of the CG sites.

At high temperature, our CG models represent the liquid density more accurately than all of atomistic models considered here. The vapour pressure and the second virial coefficient determined with our Mie (8-6) Model 1 are also in closer agreement to the experimental values than those corresponding to the other models. This is not surprising as Model 1 was specifically designed to capture the properties of the liquid and the vapour phase simultaneously. It is well known that the rigid non-polarisable atomistic models, which are parameterised to reproduce the properties of the liquid phase cannot be used to reproduce properties of the vapour phase reliably [205, 317].

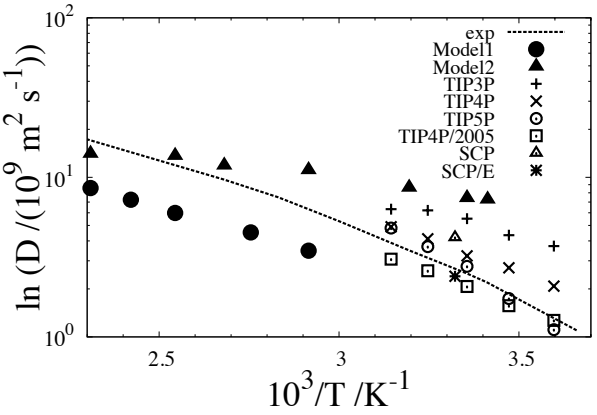



**Figure 4.16:** The self-diffusion coefficient as a function the inverse temperature. The dashed curve denotes the correlated experimental data, and the symbols represent the corresponding values obtained for the various atomistic and coarse-grained models by Molecular Dynamics simulation; the data for our Mie (8-6) CG models are shown as filled symbols. Experimental values and the data for the TIP and SPC atomistic models are taken from Ref. [205].

**Table 4.4:** Comparison of the thermophysical properties of water at 298 K or at 300 K* (indicated with asterix) and 1 atm for the coarse-grained and atomistic models: the values are given in dimensional form for the ratio of the melting and critical temperatures $T_m/T_c$, melting $T_m$/K, boiling $T_b$/K, and critical temperature $T_c$/K, saturated liquid density $\rho_L/(\mathrm{kgm}^{-3})$, vapour pressure $P^v/(\mathrm{kPa})$, enthalpy of vaporisation $\Delta H^v/(\mathrm{kJmol}^{-1})$, surface tension $\gamma/(\mathrm{mNm}^{-1})$, isobaric heat capacity $c_p/(\mathrm{Jmol}^{-1})$, isobaric thermal expansivity $\alpha_P/(10^{-4}\mathrm{K}^{-1})$, isothermal compressibility $\kappa/(\mathrm{GPa}^{-1})$, self-diffusion coefficient $D/(10^{-9}\mathrm{m}^2\mathrm{s}^{-1})$, and the virial coefficient $B_2/(\mathrm{cm}^3\mathrm{mol}^{-1})$. The values without the references were estimated in this work, either via MD simulations or with the SAFT-$\gamma$ Mie EoS[24, 55]. The experimental data is taken from NIST [195].

| | Coarse-Grained Models | | | | | Atomistic Models | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Biowater | Martini | He *et al.* (9-6) | TIP3P | TIP4P | TIP4P/2005 | SPC | SPC/E | Expt. |
| $T_m/T_c$ | 0.514 | 0.293 | 0.407 | 0.532 | 0.451 | 0.25 [205] | 0.394 [205] | 0.394 [205] | 0.321 [205] | 0.337 [205] | 0.422 |
| $T_m$ | 343 | 193 | 284 | 412 | 253 | 146 [205] | 232 [205] | 252 [205] | 190 [205] | 215 [205] | 273.15 |
| $T_b$ | 373 | 235 | 335 | 368 | 262 | 367 [205] | 364 [315] | 401 [315] | 368 [314] | 398 [314] | 373 |
| $T_c$ | 667 | 659 | 697 | 774 | 561 | 578 [205] | 588 [205] | 640 [205] | 592 [205] | 638 [205] | 647 |
| $T$ =298 K (or $T$ =300 K*) | | | | | | | | | | | |
| $\rho_L$ | 997 | 998 | 999 | 1005 [159] | 999 | 1002 [230] | 988 [235] | 993 [235] | 972 [291] | 994 [291] | 997 |
| $P^v$ | 3.1 | 536 | 30 | 7.9 | 370 | 5.2* [205] | - | 7.8* [315] | 4.6 [314] | 0.34 [314] | 3.14 |
| $\gamma$ | 179 | 72 | 72 | 32 [159] | 72 | 52.3* [205] | 59* [205] | 69* [235] | 53 [291] | 64* [235] | 72 |
| $\Delta H^v$ | 37 | 21 | 29 | 30.2 [159] | 23 | 44 [230] | 45 [230] | 46 [234] | 41 [313] | 41 [313] | 44 |
| $c_p$ | 49 | 50 | 48 | 58 | 51 | 78 [235] | 84 [235] | 88 [235] | 64 [291] | 87 [235] | 75 |
| $\alpha_P$ | 4.2 | 11 | 6 | 5.6 | 9.8 | 9.2 [205] | 4.4 [205] | 2.8 [205] | 7.4 [316] | 5 [316] | 2.6 |
| $\kappa$ | 0.14 | 0.42 | 0.43 | 0.9 [159] | 0.35 | 0.64 [205] | 0.59 [205] | 0.46 [205] | 0.53 [316] | 0.47 [316] | 0.47 |
| $D$ | 1.7 | 7.4 | 3.8 | 1.6 [159] | 5.8 | 5.5 [205] | 3.2 [205] | 2.1 [205] | 4.2 [291] | 2.4 [291] | 2.3 |
| $T$ =450 K | | | | | | | | | | | |
| $\rho_L$ | 887 | 888 | 864 | 887 | 793 | 790 [205] | 823 [235] | 879 [235] | | 860 [235] | 890 |
| $P^v$ | 849 | 6016 | 1272 | 721 | 9144 | 1195 [205] | 1330 [235] | 446 [235] | | 580 [235] | 935 |
| $\gamma$ | 101 | 43 | 40 | 19 | 23 | 24.7 [205] | 27.5 [235] | 41.8 [235] | | 36.7 [235] | 43 |
| $B_2$ | -173 | -82 | -276 | -1760 | -74 | -476 [235] | -396 [235] | -635 [235] | | -653 [235] | -238 |

## 4.5 Conclusions

We have introduced coarse-grained single-site force fields for water, which have been developed for targeted thermodynamic properties with the aid of the SAFT-$\gamma$ equation of state based on the Mie (generalised Lennard-Jones) interaction potential [24, 55]. We have parameterised simple isotropic single-site models based on a "soft" Mie (8-6) intermolecular potential. One finds however that fixed forms of intermolecular potential (i.e., fixed repulsive and attractive exponents) cannot be used to accurately represent all of the thermodynamic properties of water simultaneously. We have designed two different one-to-one molecule-to-bead coarse-grained Mie (8-6) models of water to target specific thermodynamic properties: Model 1 is optimised to reproduce the saturated liquid density and the vapour pressure, as a force field for use in the simulation of the fluid-phase equilibria of aqueous mixtures; Model 2 is designed to accurately capture the saturated liquid density and vapour-liquid interfacial tension precisely, and is therefore particularly suited for the prediction of the interfacial properties of aqueous solutions in the dense liquid phase. Since the effective coarse-grained interactions vary significantly with the temperature and density, temperature dependent size and energy parameters have been introduced to provide a reliable representation of the thermophysical properties over a broad range of conditions.

Our CG models compare favourably with the commonly used CG models as at least two generic properties are accurately reproduced with our Mie (8-6) force fields (which is not generally true of existing models) over the entire temperature range of the liquid fluid. Our Mie (8-6) Model 1 is developed at the one-to-one mapping level of coarse-graining and can be used to reliably represent the vapour-liquid equilibrium of aqueous mixtures, while force fields that result from an aggressive coarse-graining, such as Martini model, produce unphysical clustering and are unsuitable to represent the vapour phase.

To a certain extent all coarse-grained models suffer from issues of representability for the different types of thermodynamic properties and transferability to different thermodynamic states; this is particularly true for water and aqueous systems. The advantage of the methodology employed in our work is that it is straightforward to implement an alternative parameterisation if different thermodynamic properties or conditions are of interest. The level of coarse-grained mapping can also be tuned for specific applications. In this

vein, we have also proposed a Mie (8-6) model of water at the two-to-one molecule-to-bead level ("biowater"), parameterised to reproduce the liquid density and surface tension at ambient conditions. Our Mie (8-6) biowater model is therefore designed for use in large scale simulations of aqueous solutions of amphiphiles, lipid membranes, proteins and other challenging biomolecular systems. The relative simplicity of the coarse-grained representation means that fewer unlike intermolecular parameters have to be specified in mixtures. Furthermore, the unlike interactions between the various chemical moeities in the system can also be estimated reliably with the aid of the SAFT-$\gamma$ Mie EoS from appropriate macroscopic experimental data for the given mixture [68].

# Chapter 5

# Binary and ternary mixtures comprising carbon dioxide, water and $n-$alkanes

## 5.1 Chapter overview

In this chapter, we study the phase behaviour of the binary and ternary mixtures comprising carbon dioxide, water and $n-$alkanes. The models for the $n-$alkanes and water have been developed in Chapters 3 and 4, respectively, whereas the model for carbon dioxide was taken from the previous publication on the SAFT-$\gamma$ Mie force field [22]. The modelling is performed by using the SAFT-$\gamma$ Mie EoS. The unlike interactions between the components are described by a single adjustable parameter, which appears to be transferable over a wide range of conditions. The derived CG force fields are used in Molecular Dynamics simulations in order to validate the predictions by the theory and also to calculate the properties that cannot be accessed via the equation of state, such as interfacial phenomena. The quality of predictions is comparable to the description by the atomistic models and other well-established equations of state.

## 5.2 Phase behaviour of binary mixtures

Due to a large variety of possible intermolecular interactions between two components, binary mixtures exhibit a very rich phase behaviour, which is commonly described in

the classification of Scott and van Konynenburg [32, 33]. Binary mixtures are commonly subdivided into six main types, which differ from one another by the presence or absence of the three phase line and the way the critical line connects with it. In the types I, II, and VI mixtures, the components exhibit similar critical temperatures, therby resulting in a continuous vapour-liquid critical line, which connects both pure components. In contrast to this, in the mixtures of types III, IV, and V, the critical temperatures are very different, thereby triggering the discontinuity of the critical line. The corresponding $PT$-projections are illustrated in Figure 5.1. The original classification included five types of mixtures (from I to V) which were described by the van der Waals EoS by choosing various $a$ and $b$ parameters. Type VI was proposed later by Gubbins and co-workers [318, 319]. In recent literature, other sub-types are proposed as different combinations of the main six types.

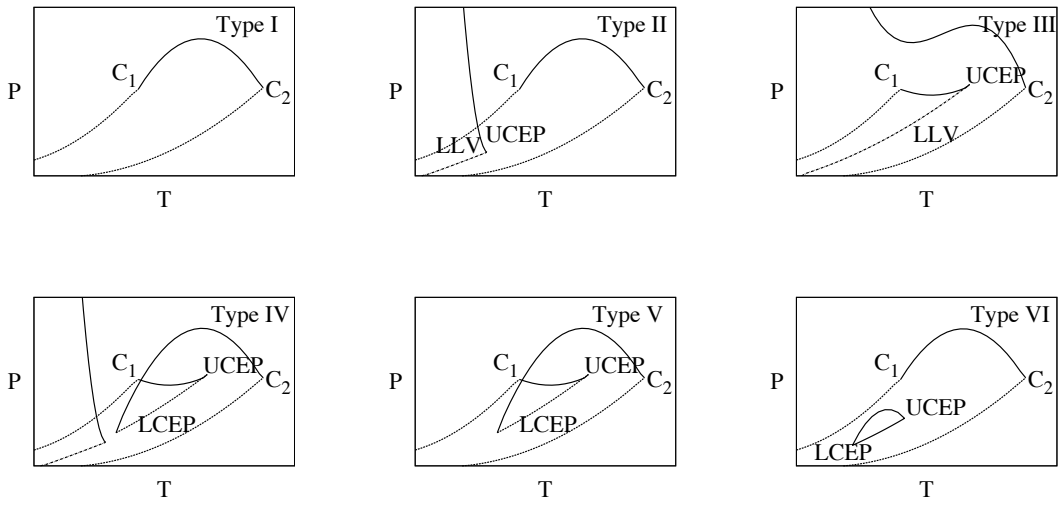



**Figure 5.1:** $PT$-projections illustrating the critical phase behaviour according to the classification by Scott and van Konynenburg [32, 33]. Solid curves denote the critical lines, dashed curves are the pure component vapour pressures and dash-dotted curves are the three-phase lines. $C_1$ = critical point of component 1; $C_2$ = critical point of component 2; LLV = liquid-liquid-vapour curve; UCEP= upper critical end point; LCEP= lower critical end point.

The simplest case of the mixture is where two components have similar molecular sizes, type of interactions, and critical properties, so that the mixture exhibits a relatively ideal behaviour. Referred to as the type I mixture, it is described by a continuous vapour-liquid critical line and no liquid-liquid immiscibility occurs. Typical examples include the mixtures of nobel gases such as argon with krypton, or mixtures of carbon dioxide with

the short-chain alkanes up to $n-$hexane [320].

As the differences between both components increase, the liquid-liquid immiscibility appears at low temperatures. Since the liquid-liquid critical line does not expand beyond the critical temperatures of any of the components, the vapour-liquid critical line remains continuous connecting the critical points of both pure components. The liquid-liquid critical line merges with the three phase line in the upper critical end point (UCEP). This type of mixture is referred to as the type II and its typical examples include mixtures of carbon dioxide with middle-length alkanes from $n-$heptane to $n-$dodecane [320].

The most non-ideal case with the largest disparities in the intermolecular forces is given by the mixtures of type III. The liquid-liquid immiscibility region expands to the temperatures above the critical point of at least one component. By penetrating the region of the vapour-liquid critical line, the critical line becomes discontinuous. Its upper branch starts at the critical point of the less volatile component, and proceeding to higher pressures finally merges with liquid-liquid critical line. The lower branch of the vapour-liquid critical line is usually much shorter; it starts at the critical point of the more volatile component and merges with the three phase line at the UCEP. The type III behaviour is quite common and occurs for systems including alkanes and water, water and carbon dioxide, and carbon dioxide with the long-chained alkanes (from $n-$tetradecane onwards) [320]. Those systems are the main focus of this chapter. Type IV occurs as a transition between the type II to type III behaviour in the same homologous series, as for example in the carbon dioxide and $n-$tridecane system [320]. The vapour-liquid critical line starts at the critical point of a less volatile component and merges with the three phase line in the lower critical end point (LCEP). In this manner, the $Tx$-projection exhibits a closed loop of the liquid-liquid immiscibility. Additionally, the second liquid-liquid immiscibility region occurs at the low temperature and the three phase line merges with the liquid-liquid critical line in the upper critical end point as in the type II mixture. This is the only difference to the type V mixture, which does not exhibit the liquid-liquid immiscibility at low temperatures. The type VI is also characterised by the closed loop immiscibility in the liquid-liquid region, however it occurs at the temperatures below the critical points of both components so that the vapour-liquid critical line remains continuous.

## 5.3 Introduction to the systems

The mixtures of carbon dioxide, water and $n-$alkanes have received a particular interest in the literature due to their central role in petrochemical industry. The range of applications spans enhanced oil recovery, $CO_2$ storage, natural gas transport and processing, and many other processes. An accurate description of the phase behaviour of these mixtures is crucial for the optimal design of a chemical process. From the modelling perspective, the mixtures that $n-$alkanes form with $CO_2$ and water are challenging because they exhibit a high degree of non-ideality due to quite different nature of interactions between the components. While $n-$alkanes are non-polar, $CO_2$ shows a significant quadrupole moment and water has a large dipole moment and forms hydrogen bonds. The resulting mixtures are therefore highly asymmetric, both in size and energetic interactions. The disparity of the those interactions produces different mixture fluid phase behaviour (Section. 5.2).
In the mixtures of carbon dioxide with the middle-length alkanes a type II behaviour is observed, whereas the mixtures with the long alkanes display a type III behaviour [321, 322]. The type III behaviour is also observed in $CO_2$+water [323, 324] and $n-$alkane+water mixtures [325, 326].
In terms of modelling, the use of a cubic equation of state (EoS), such as Peng-Robinson (PR) [9] or Soave-Redlich-Kwong (SRK) [10] remains a common practice in reservoir engineering. Yet, since this type of the equation of state is mostly suitable for the spherical components and mixtures with non-associating interactions, an accurate description of the phase equilibria is only possible if a large number of adjustable parameters is introduced, thereby leading to the reduction of the predictive power of the model. As a recent example one can refer to recent work of Cismondi *et al.* [327], who accurately described phase diagrams of the entire homologous series of $n-$alkanes with carbon dioxide using a RK-PR equation of state at a cost of employing forty adjustable parameters.
In the study of highly associating mixtures, using an equation of state that can directly account for association has been proven advantageous. The SAFT EoS has been successfully applied to study the phase behaviour of mixtures containing $n-$alkanes, water and carbon dioxide. Some of distinctive properties of those systems such as transition in the phase behaviour from type II to type III for $CO_2$+$n-$alkanes [328–330], barotropic density inversion in $CO_2$+water mixture and solubility minima of water in $CO_2$ [172, 331–

336], solubility minima for hydrocarbons in water [337, 338] and extreme immiscibility of $n-$alkane+water mixtures [58, 59, 339–345] can be accurately reproduced using various versions of the SAFT equation of state. Some models are more complex than others, for instance, when the quadrupole moment is included into the $CO_2$ model [172, 335]. Additionally, controversy in the literature has aroused on whether the cross-association is needed between $CO_2$ and water components to account for the change of the slope in solubility near the critical temperatures [331, 346] or whether this behaviour can be reproduced without any additional interactions [332–334].

Although an equation of state is still a preferable tool in the petrochemical community due to its efficiency, in the last few decades the molecular simulation has emerged as an alternative for a number of industrial applications [3, 4]. The mixtures containing $n-$alkanes, carbon dioxide and water have therefore received a lot of attention also from the molecular simulation perspective. Different atomistic force fields have been employed to study those mixtures. To name just a few examples, Cui and Cummings [62] used the EPM2 model for carbon dioxide and the model developed by Siepmann, Karaborni, and Smit (so-called SKS model) for $n-$alkanes [181] to describe vapour-liquid equilibrium of $CO_2$+$n-$alkane mixtures. Potoff and Siepmann [347] developed the $CO_2$ model within all-atom TraPPE formalism to calculate VLE of binary mixtures with $n-$alkanes and nitrogen. Potoff and Panagiotopoulos [63] studied phase behaviour of various polar and non-polar components including $CO_2$, $n-$alkanes and water mixtures using exp-6 potential models. Johansson *et al.* [348] studied equilibrium solubilities and structures of water in $n-$alkanes using a modified TraPPE-SPC force field. Ferguson *et al.* [349] explored the water-rich liquid phase by studying the solubility and molecular conformations of $n-$alkane chains in water using a TraPPE model for alkanes and the SPC/E model for water. Recently, a comprehensive study of $CO_2$+$H_2O$ mixture has been carried out by Liu *et al.* [350], where the authors compared the performance of several existing water (SPC, TIP4P, TIP4P2005, and exp-6) and $CO_2$ models (EPM2, TraPPE, and exp-6) to study the fluid phase behaviour at moderate and high temperatures. It has been concluded that none of the tested models were capable of reproducing the experimental data over the entire temperature and pressure range despite the high resolution of the models. Consequently, a drawback of the atomistic models is not only their high computational demand but there are still open questions with respect to their limitations in terms of transferability

and robustness. In order to enhance the computational efficiency and to improve the transferability, representability, and robustness, we aim to develop a new SAFT-$\gamma$ Mie CG force field to study properties of binary and ternary mixtures.

## 5.4 Methods and models

### 5.4.1 SAFT-$\gamma$ force field

For the sake of completeness, the CG model of carbon dioxide, developed by Avendaño *et al.* [22] is briefly summarised here. The $CO_2$ model is represented by a single sphere based on the Mie (23-6.66) potential, without any additional electrostatic interactions. The size and the energy parameters are given by $\sigma$ =3.741 Å and $\epsilon/k_B$ =353.55 K, respectively. Despite the simplicity of the model, it is capable to accurately reproduce the entire phase envelope with a single parameter set. The prediction of properties which were not used in the parameterisation procedure such as enthalpy of vaporisation, interfacial tension, supercritical density, as well as the second derivative thermodynamic properties (thermal expansion, isobaric heat capacity, isothermal compressibility, speed of sound and Joule-Thomson coefficient) were found to be in a very good agreement with experimental data. In a comprehensive study, which compared the thermodynamic properties of seven different carbon dioxide force fields, Aimoli *et al.* [351] found that the SAFT-$\gamma$ Mie force field delievers a comparable accuracy as higher level resolution force fields, such as rigid and fully flexible three-site models of $CO_2$.

The $n-$alkanes are modelled as the homonuclear chains of the tangential Mie spheres based on the parameter set given in Table 3.4. Model 1 ( Section 4.4.3), which provides an accurate representation of the VLE properties, is used to represent water in the binary and ternary mixtures with $CO_2$ and $n-$alkanes. The model is based on the Mie (8-6) potential with temperature dependent size and energy parameters, $\sigma/\text{Å} = \sigma(T) = 1.262 \times 10^{-9}T^3 - 8.720 \times 10^{-9}T^2 - 4.554 \times 10^{-4}T + 3.119$ and $(\epsilon/k_B)/K = (\epsilon/k_B)(T) = 1.105 \times 10^{-5}T^2 - 0.3077T + 586.8$. The interfacial properties of water mixtures however cannot be captured by the model and hence will not be treated in this chapter. The combining rules to describe the unlike interactions between the pure components are given by Eqs. 2.57, 2.59, and 2.60. The SAFT-$\gamma$ Mie EoS is used to adjust the $k_{ij}$ parameter in Eq. 2.60 in order to account for the deviations from the geometric combining rule between

two components. The $k_{ij}$-values are obtained by adjusting to one isotherm in the $Px$ diagram for the corresponding binary mixture, with the derived values are given in Sections 5.5.1, 5.5.2, and 5.5.3. Although adjusted to solubility only at one temperature, the $k_{ij}$-parameter appears to be transferable to different state points and is therefore applied to study the entire phase diagram. Since the temperature dependence is already incorporated in the water model, it appears to account for a correct behaviour in description of water mixtures at different temperatures. Derived from the SAFT EoS, the obtained force fields are then used as a direct input in molecular simulation in order to validate the predictions by the theory and to investigate the adsorption properties.

### 5.4.2 Simulation details

The vapour-liquid equilibria is calculated using Molecular Dynamics (MD) simulation in the canonical ensemble ($NVT$), with constant number of particles $N$, temperature $T$ and volume $V$, whereas the liquid-liquid-phase equilibria simulations are carried out in the isobaric-isothermal $NPT$ ensemble, with constant number of particles $N$, temperature $T$, and pressure $P$. The system temperature is maintained using a Nosé-Hoover thermostat [188, 189] with the coupling constant of 1.0 ps, whereas the pressure is maintained by means of a Parrinello-Rahman barostat with a relaxation constant of 10.0 ps. The equation of motion is integrated by using the leap-frog algorithm [18] with a time step size of 5 fs. The simulations are carried out with the Gromacs package, version 4.5.5 [190]. For the mixtures, the same cutoff radius is used as for the pure components of $n-$alkanes and water, namely 3.0 nm. The global density and the composition of the systems is chosen inside the coexistence region. As an initial configuration in the simulation of liquid-liquid equilibrium, a phase separated system is assembled in the orthorhombic simulation box, which dimensions are constrained to minimum 6 nm in $x$ and $y$ directions due to the chosen cutoff radius. $z$ dimension is approximately four times larger than the $x$ and $y$ dimensions. The system size is chosen accordingly to the box dimensions. During the LLE simulation in $NPT$ ensemble at a pressure which corresponds to the liquid phase, the equilibrium box dimensions are established. In vapour-liquid equilibrium simulation, $z$ dimension is chosen so that enough vapour phase at a pressure corresponding to the two phase region can be formed. The standard periodic boundary conditions [18] are applied. After the equilibrium has been established, the configurations are sampled for further

20 ns. Two phases are formed between two planar interfaces. The composition of the fluid phases is estimated from the corresponding density profiles. The same unlike binary interaction parameters are used to account for the interactions between three components in the ternary mixture.

## 5.5 Results and discussion

The Results section is organised as following: for each of the binary mixtures studied, the unlike interaction parameter is adjusted to the fluid phase equilibria at one temperature. It is important to note that only one adjustable parameter is required to account for the unlike interactions between the components over a wide range of conditions. Subsequently, the obtained parameter is used to investigate the entire phase behaviour using the SAFT-$\gamma$ Mie EoS, including VLE, LLE, critical lines, and three phase lines. Three dimensional diagrams are produced to visualise and to provide a better understanding of the phase behaviour. For the mixtures with $n-$alkanes, the same parameter set is used to predict the phase equilibria of the components of different chain length. Additionally, the MD simulations based on the derived CG force fields are performed to compare with the experimental and theoretical results.

### 5.5.1 $CO_2$+$H_2O$ mixture

The unlike interaction parameter in the $CO_2$+$H_2O$ mixture is adjusted to the liquid-liquid equilibrium at $T = 473$ K using SAFT-$\gamma$ Mie EoS, where $k_{ij} = -0.07$ is found to provide the best possible agreement with experimental data. This result is consistent with the study of carbon dioxide+water mixture by the SAFT-VR EoS, performed by Forte *et al.* [336]. The quality of representation is illustrated in Figure 5.2. The solubilities in both phases are reasonably well reproduced, including the solubility minimum of water in $CO_2$-rich phase. The developed force field is validated by means of MD simulation. At low pressures, MD simulation results accurately reproduce the prediction made by the SAFT EoS and therefore the experiment. The solubility minimum of water in $CO_2$-rich phase is captured correctly by the CG models. By approaching the critical region, the simulations show larger deviations from the theory due to the divergence of the correlation length. However, the shape of the phase envelope is qualitatively reproduced.

The transferability of the model can be validated by applying the same parameter set at temperatures which were not considered in the parameterisation procedure. The adjustable parameter $k_{ij}$ is treated as temperature-independent and it appears indeed to be transferable to different temperatures (see for instance the isotherm at $T = 383$ K in Figure 5.2). At lower pressures, the phase compositions alongside with the solubility maximum are accurately captured. At high pressures, however, the deviations become more apparent. Molecular Dynamics simulations, performed at the same conditions, agree well with theoretical predictions.

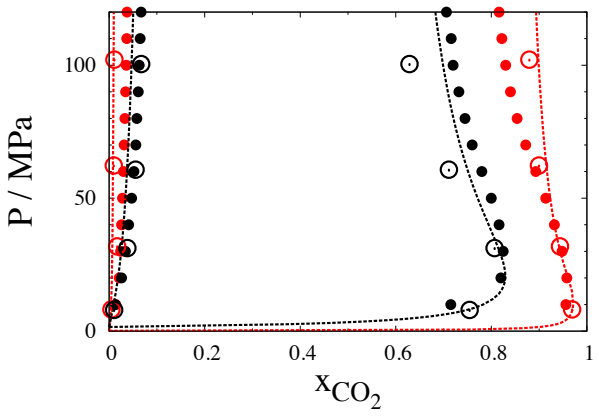



**Figure 5.2:** Isothermal pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at $T$ =383 (in red) and 473 K (in black). The filled circles correspond to experimental fluid-liquid equilibrium data [323]. The dashed curve corresponds to the SAFT-$\gamma$ calculation with $k_{ij} = -0.07$. The open circles are the MD simulation results for the SAFT-$\gamma$ Mie CG force field.

Using the unlike interaction parameter $k_{ij} = -0.07$, the entire phase diagram can be studied quite efficiently by means of the SAFT-$\gamma$ Mie EoS. The carbon dioxide+water mixture exhibits the type III behaviour in the classification of Scott and van Konynenburg, which is characteristic of highly immiscible fluids. The global features of the phase behaviour, which are visualised in Figure 5.3, illustrate that the phase diagram is indeed largely dominated by the immiscibility region.

The discontinuous vapour-liquid critical line can also be seen in the $PT$ projection of the $PTx$ surface in Figure 5.4. The critical line at the intermediate pressures shows a good agreement with experimental data but there are larger deviations at higher pressures.

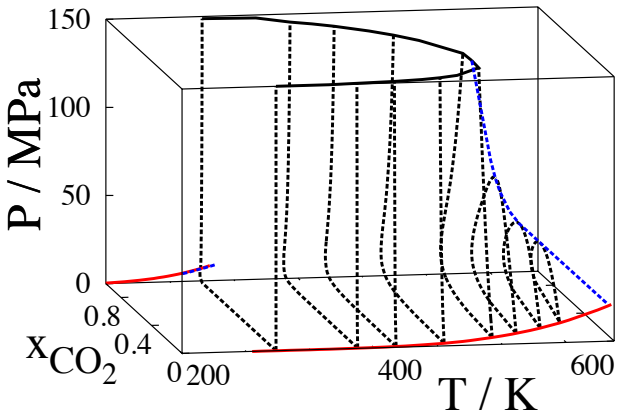



**Figure 5.3:** Three-dimensional ($PTx$) phase diagram for carbon dioxide+water system. The continuous red lines are the SAFT-$\gamma$ calculations for the vapour pressures of the pure components, the dashed blue lines are the SAFT-$\gamma$ predictions for the vapour-liquid critical line and three phase line, the dashed black lines correspond to ($Px$) sections.

The lower branch of the vapour-liquid critical line merges with the three phase line in the UCEP; the three phase line appears at the pressures slightly lower than the vapour pressure of carbon dioxide. The critical point of the $CO_2$ model is overestimated [22] due to the nature of a mean field theory, which leads to a slight overestimation of the three phase pressures by SAFT-$\gamma$ Mie EoS. The zoomed representation of the three phase line and the lower branch of the critical line is illustrated in the inset of Figure 5.4.

The three phase line separates the liquid-liquid equilibrium at the higher pressures from the vapour-liquid equilibrium at the lower pressures, *i.e.* at the temperatures and pressures of the three phase line, two immiscible liquid phases coexist with the vapour phase. In Figure 5.5, the isothermal pressure-composition ($Px$) at $T = 298$ K, predicted by the SAFT-$\gamma$ Mie EoS is compared to experimental data. The vapour-liquid equilibrium is in a very good agreement with experiment, whereas at the liquid-liquid equilibrium conditions, the concentration of carbon dioxide is slightly underestimated in both phases. The three phase pressure is reproduced within 0.45 MPa deviation compared to the experimental value. A small vapour-liquid region with $CO_2$-rich vapour and $CO_2$-rich liquid phase is observed above the three phase line. The theory is able to capture the main features of the carbon dioxide-water system at this temperature, providing an explanation to the solubility minimum of water in the $CO_2$-rich phase at higher temperatures. The solubility minimum is ascribed to the presence of the three phase line. At low temperatures, the

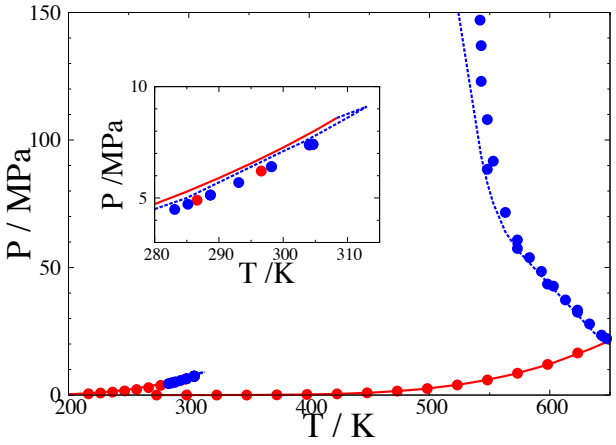



**Figure 5.4:** *PT* projection of the phase diagram for the carbon dioxide+water binary mixture. The filled red circles correspond to the experimental vapour pressures of the pure components [195], the filled blue circles correspond to the experimental vapour-liquid critical line and three phase line [323, 324]. The continuous red lines are the SAFT-$\gamma$ calculations for the vapour pressures of the pure components, the dashed blue lines are the SAFT-$\gamma$ predictions for the vapour-liquid critical line and three phase line. The inset shows the enlarged region close to the critical point of pure $CO_2$.

solubility curve exhibits a sharp discontinuity at the three phase coexistence pressure. As the temperature increases above the UCEP, the discontinuity disappears thereby causing a minimum at the pressures close the UCEP. This behaviour is described in greater details in Ref. [332]

MD simulations reproduce the theoretical predictions very accurately (Figure 5.5). The three phases: the water-rich liquid phase ($L_1$), the carbon dioxide rich liquid phase ($L_2$) and the gaseous phase $G$, can be clearly distinguished in the corresponding density profile and the snapshot, presented in Figure 5.6. The adsorption of carbon dioxide on the interface between $L_1$ and $L_2$ phases is distinctly seen from the density profile.

The adsorption phenomena along the interfacial region can be studied via MD simulations at different temperatures and pressures. Figure 5.7 presents the density profiles at temperatures $T$ = 383 and 423 K and four different pressures: $P$ =8.18 MPa, 31.87 MPa, 62.30 MPa, and 102.10 MPa. Carbon dioxide exhibits a positive surface activity ($\mathrm{d}\rho_{CO_2}/\mathrm{dz} = 0$, $\mathrm{d}^2\rho_{CO_2}/\mathrm{dz}^2 < 0$), whereas water does not show any surface activity. The surface activity of carbon dioxide increases as the pressure increases until the limit of the bulk liquid

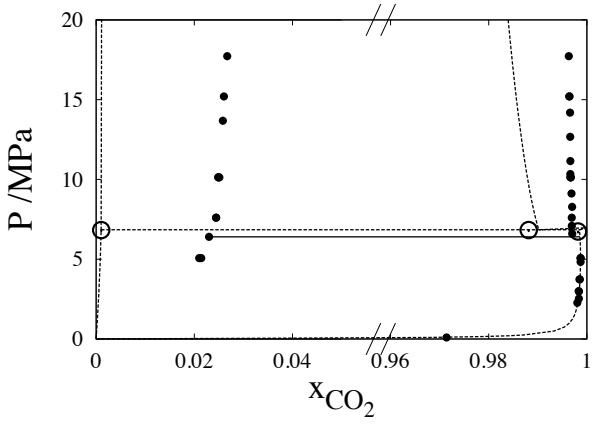



**Figure 5.5:** Isothermal pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at $T = 298$ K. The filled circles correspond to experimental vapour-liquid and liquid-liquid equilibria [352–357]. The solid line corresponds to the experimental three phase line. The dashed curves correspond to SAFT-$\gamma$ prediction with $k_{ij} = -0.07$. The open circles are the MD simulation results for the SAFT-$\gamma$ Mie CG force field.

density is reached and the saturation is established. With increasing temperature, a lower surface activity and the widening of the interface produced by the decreased slope of the $CO_2$ density curve is observed.

Apart from the density at the interface, the bulk densities of the carbon dioxide+water mixture as functions of pressure represent s substantial scientific interest since the mixture exhibits a barotropic density inversion. In this phenomenon, the supercritical $CO_2$ becomes denser than the water phase at high pressures. In Figure 5.8, the experimental densities of the mixture are compared to the prediction by the SAFT-$\gamma$ Mie EoS and MD simulations at $T = 298$, 304, 383, 423, and 473 K. The theoretically calculated densities of carbon dioxide phase are in excellent agreement with experiment, both in the sub- and the supercritical region, whereas the densities in the water rich phase are slightly underestimated. The pressures at which the barotropic density inversion occurs at the given temperature, are compared to the experimental values from Ref. [324] in Table 5.1. Taking into account that those are true predictions of very simple models with only one adjustable unlike interaction parameter, the agreement between theoretical predictions and experimental data can be considered remarkable. The MD simulations reproduce the prediction from SAFT-$\gamma$ Mie EoS and thereby the experimental data quite accurately.

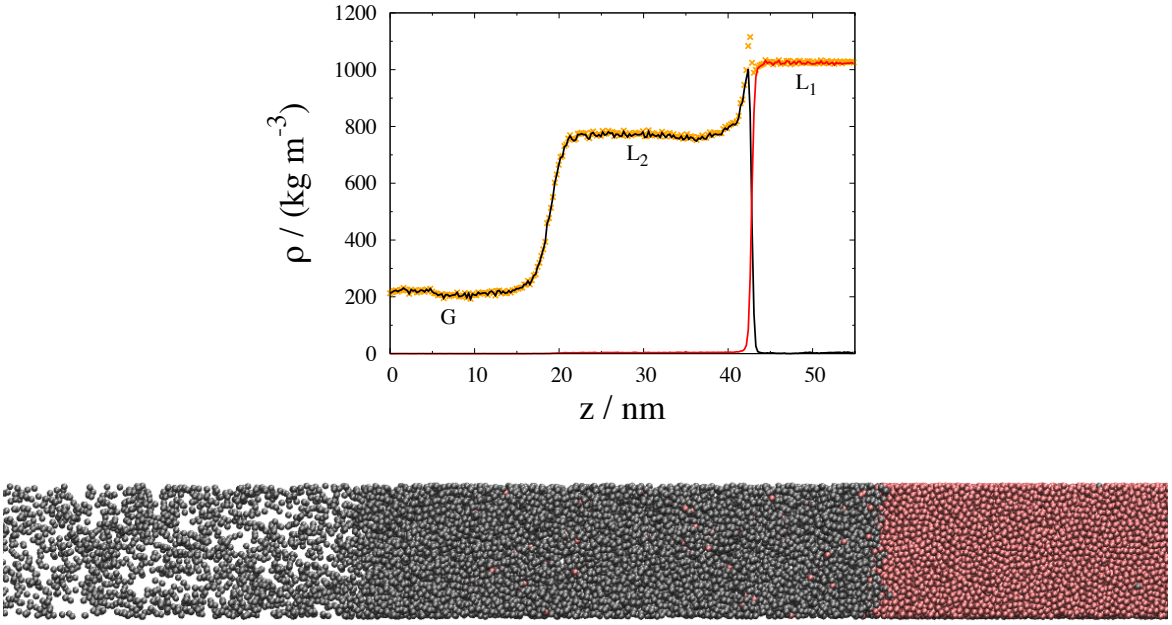



**Figure 5.6:** Top: total density profile of carbon dioxide +water mixture at the three phase point at $T = 298.15$ K and $P = 6.82$ MPa obtained via MD simulation. $L_1$ denotes water-rich liquid phase, $L_2$ denotes $n$-hexane rich phase, $G$ denotes the gas phase; Bottom: snapshot from Molecular Simulations, where water molecules are displayed in red and carbon dioxide in grey, corresponding to the lines of the same colour in the density profiles.

**Table 5.1:** Barotropic density inversion points for the carbon dioxide +water mixture, experimental [324] and predicted with SAFT-$\gamma$ Mie EoS.

| $T/K$ | $P_{exp}/MPa$ | $P_{SAFT}/MPa$ |
|---|---|---|
| 323 | 80 | 70 |
| 373 | 125 | 108 |
| 523 | 200 | 161 |

The results for our CG models are compared to the atomistic simulations performed by Liu *et al.* [350], who studied carbon dioxide+water system based on some popular force fields for water (SPC, TIP4P, TIP4P2005, and exponential-6) and carbon dioxide (EPM2, TraPPE, and exponential-6), using Lorentz-Berthelot combining rules to describe the unlike pair interactions. The authors used two different sets of the experimental data, by Takenouchi and Kennedy [323] and Todheide and Franck [324], to validate their results. In Figure 5.9, the CG simulations are compared to the atomistic results at two temperatures: away from the critical region ($T = 423$ K) and near the critical region ($T = 548$

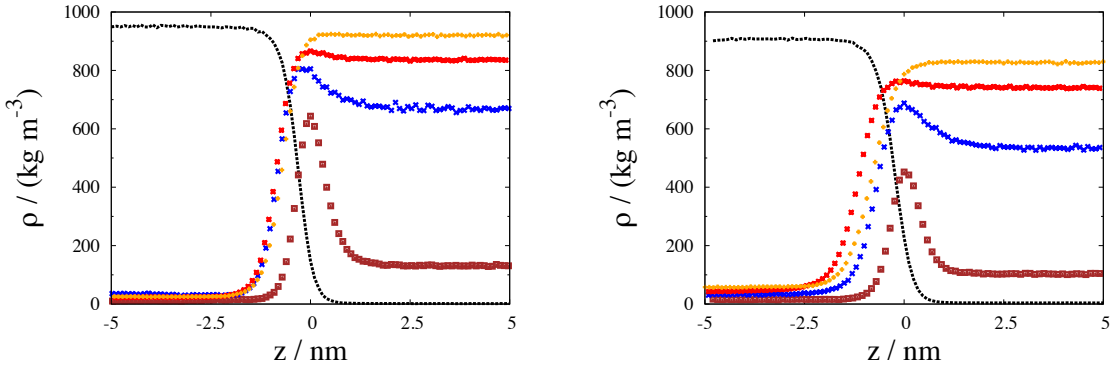



**Figure 5.7:** Density profiles at the interphase of the carbon dioxide+water mixture at $T = 383$ K (left) and 423 K (right) and different pressures calculated by MD simulations. The dashed line denotes water density profile at $P$ =8.18 MPa, the symbols represent the carbon dioxide profiles at $P$ =8.18 MPa (brown squares), 31.87 MPa (blue crosses), 62.30 MPa (red stars), and 102.10 MPa (yellow pluses).

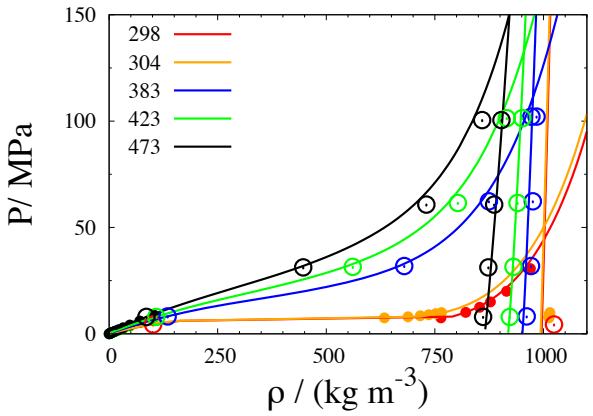



**Figure 5.8:** Pressure-density ($P\rho$) diagram for carbon-dioxide+water mixture at isotherms corresponding to $T$ =298, 304, 383, 423, and 473 K. The filled circles correspond to experimental density data for the saturated phases [196]. The curves correspond to the predictions obtained by SAFT-$\gamma$ with $k_{ij} = -0.07$, the open circles are the MD simulation results for our SAFT-$\gamma$ CG Mie force field.

K). In view of the disagreement between the different data sets, it is difficult to judge which model delivers the best performance. However, the experimental data by Todheide and Franck [324] were reported to exhibit large errors in composition (up to 1 mol %)

[358]. Neither captures this data set the solubility minimum at $T = 423$ K, whereas the experimental data from Takenouchi and Kennedy [323] undergoes a clear minimum at low pressures. The solubility minimum is not reproduced by any of the atomistic models. Compared to the experimental data from Ref. [324], the best performance from the atomistic force fields is achieved by the exponential-6 model for the $CO_2$-rich phase and TraPPE/TIP4P2005 model for the $H_2O$-rich phase at $T = 423$ K. The results from the CG models are closer to experimental data from Ref. [323], showing a solubility minimum at low pressures. At high temperatures, the exponential-6 models perform better than the other atomistic models. They can accurately reproduce the lower part of the phase envelope, whereas the critical pressure is largely underestimated compared to both sets of experimental data. The CG models show a similar behaviour, by exhibiting a good agreement with experiment at low pressures and larger deviations and an underestimation of the critical point at high pressures. Despite the high resolution, the existing atomistic models for carbon dioxide+water system cannot adequately reproduce the phase behaviour over a wide temperature and pressure range. At low temperatures, the solubility in the both phases cannot be reproduced simultaneously. At high temperatures, only a qualitative description can be achieved with atomistic models. On that account, the performance of the SAFT-$\gamma$ Mie CG models can be considered as competitive with respect to the atomistic models. Having a direct robust link between theory, experiment and simulation enables us studying the phase behaviour over a wide range of conditions with greater confidence.

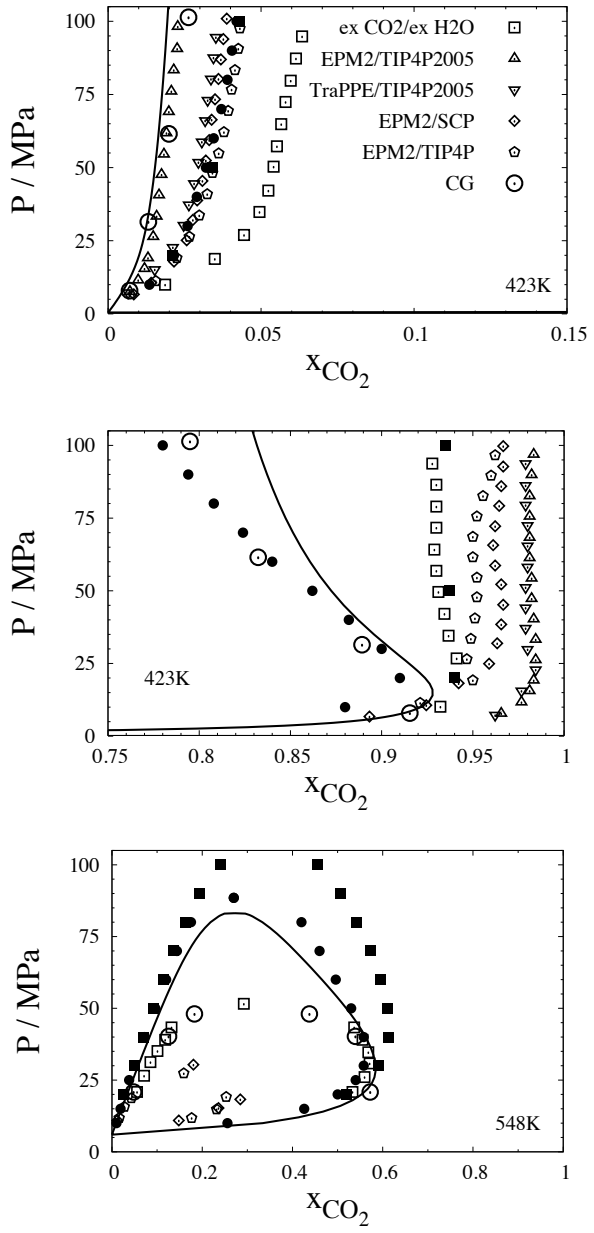



**Figure 5.9:** Isothermal pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at $T = 423$ and 548 K. The filled circles correspond to experimental data from Ref. [323], the filled squares correspond to experimental data from Ref. [324]. The dashed line corresponds to the predictions obtained by SAFT-$\gamma$ with $k_{ij} = -0.07$, the open circles are the MD simulation results for our SAFT-$\gamma$ CG Mie force field. The CG simulations are compared to some popular atomistic models: exponential-6 $CO_2$/exponential-6 $H_2O$ model (□); EPM2/TIP4P2005 model (△); TraPPE/TIP4P2005 model (▽); EPM2/SPC model (◇); EPM2/TIP4P model (pentagon). The unlike-pair parameters of the atomistic force fields are based on the conventional Lorentz-Berthelot combining rules. Atomistic simulations are taken from Ref. [350].

### 5.5.2 $n-$alkane + $H_2O$ mixtures

The same approach described above is applied to study a system with extreme immiscibility, $n-$alkanes and water. A value of $k_{ij} = 0.36$ is obtained from adjusting to the solubility of water in a hexane-rich phase at $T = 473.2$ K. The quality of the representation is illustrated in Figure 5.10. The calculation is then extended to low pressures to study the vapour-liquid equilibrium. The $n-$hexane+water mixture exhibits a heteroazeotrope, where the vapour phase coexists with two immiscible liquid phases. The azeotropic point is accurately captured by the SAFT-$\gamma$ Mie EoS both in terms of composition and pressure at the given temperature. MD simulations, which are carried out to verify the theoretical results, can correctly reproduce the solubility of water in hexane-rich phase. In vapour-liquid equilibria, the compositions in both phases are also accurately captured by the simulations. The result for the phase composition at the azeotropic point displays a remarkable agreement with the theoretical prediction and the experimental value. The corresponding density profile of the mixture at $T = 473.2$ K and $P = 3.33$ MPa is illustrated in Figure 5.11. Three phases (a water-rich liquid phase $L_1$, a hexane-rich liquid $L_2$, and a gas phase $G$) are clearly distinguished by the density, which is also apparent from the MD simulation snapshot.

Whereas the solubility of water in $n-$hexane-rich phase is accurately captured by the parameter set referred to above, the solubility of $n-$hexane in the water rich phase exhibits significant deviations. The experimental solubility values are extremely low: in the order of magnitude of $10^{-6}$. Calculation with the SAFT EoS results in a solubility of approximately $10^{-9}$. Whereas the solubility of $n-$alkanes in the water-rich phase is of a lesser importance for petrochemical applications, it is crucial for the estimation of solubility of the alkyl groups of more complex species which are relevant for the detergence and pharmaceutical industry. In this case, a separate value of $k_{ij}$ is required to describe the solubility of $n-$alkanes in the water rich phase. It has already been shown by Haslam *et al.* [61] that for the mixtures with extreme immiscibility, no single unlike interaction parameter $k_{ij}$ can be used to describe both phases with a high accuracy.

Using the same value of $k_{ij} = 0.36$ for the unlike interactions, we extend the prediction to study the rest of the phase diagram. The family of $n-$alkanes+water mixtures is known to exhibit the type III behaviour due to the mentioned immiscibility of the system [325].

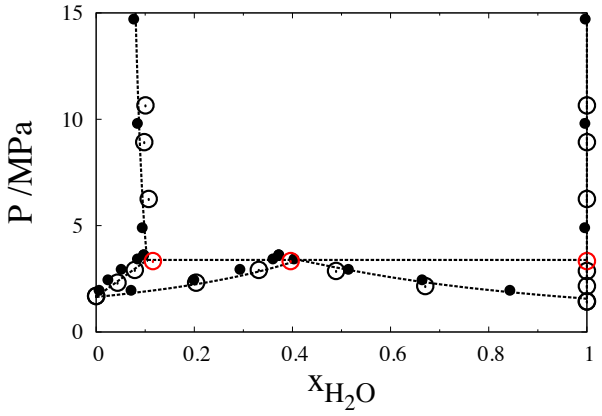



**Figure 5.10:** Isothermal pressure-composition ($Px$) phase diagram for water+ $n-$hexane mixture at $T = 473.2$ K. The filled circles correspond to experimental vapour-liquid and liquid-liquid equilibrium data [359–361]. The dashed lines correspond to the results obtained by SAFT-$\gamma$ Mie EoS with $k_{ij}$=0.36. The open circles are the MD simulation results for SAFT-$\gamma$ CG Mie force field. Heteroazeotropic point at $P = 3.33$ MPa with composition $x^{L_1}_{H_2O} = 1.0$, $x^{G}_{H_2O} = 0.396$, and $x^{L_2}_{H_2O} = 0.116$ is highlighted in red.

The extent of the immiscibility gap is visualised in the three dimensional phase diagram in Figure 5.12. The discontinuity of the vapour-liquid critical line is displayed in $PT$ projection in Figure 5.13. In the SAFT description, the lower branch of the critical line is slightly overestimated compared to experimental data due to the over prediction of the critical point for $n-$alkanes by the equation of state. Nevertheless, if the critical point is estimated from the scaling laws using the same model in the molecular simulation, the result is closer to the experimental value, as it has been shown in Figure 3.2 in Section 3.4.1. In the heteroazeotropic systems, the three phase line is located at pressures higher than the vapour pressure of both pure components. As it was already evident from Figure 5.10, the three phase line pressure is in excellent agreement with the experimental data. The upper branch of the vapour-liquid critical line should exhibit a minimum typical for the systems with the gas-gas immiscibility of the second kind. However, the SAFT EoS does not capture the change of the slope of the critical line thereby suggesting the gas-gas immiscibility of the first kind, characteristic for systems with even greater disparity of the intermolecular forces.

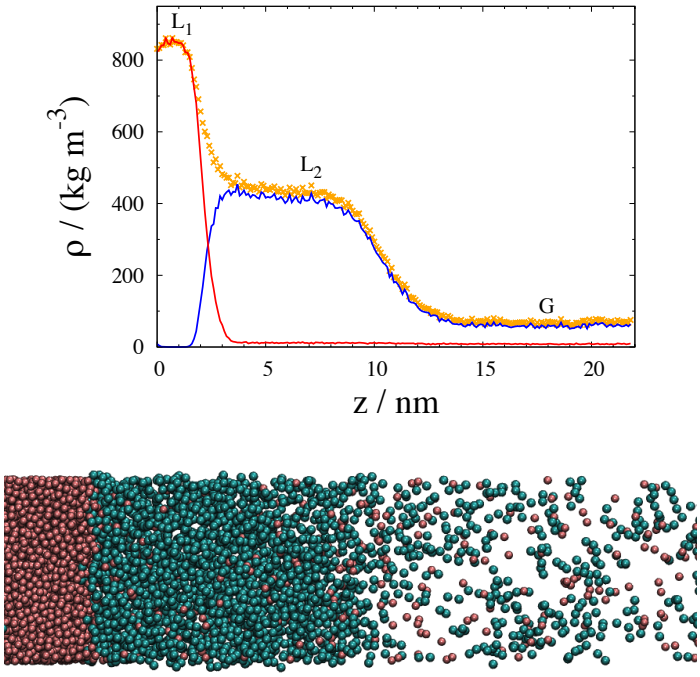



**Figure 5.11:** Total density profile of $n-$hexane+water mixture at the three phase point (heteroazeotrope) at $T = 473.2$ K and $P = 3.33$ MPa obtained via MD simulation. $L_1$ denotes water-rich liquid phase, $L_2$ denotes $n$-hexane rich phase, $G$ denotes the gas phase; composition of the heteroazeotrope is $x^{L_1}_{H_2O} = 1.0$, $x^{G}_{H_2O} = 0.396$, and $x^{L_2}_{H_2O} = 0.116$. In the snapshot from Molecular Simulations, water molecules are displayed in red and $n-$hexane in green, corresponding to the lines of the same colour in the density profiles.

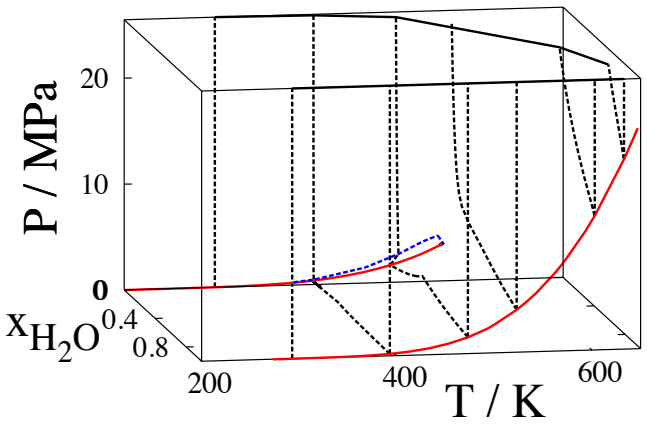



**Figure 5.12:** Three-dimensional ($PTx$) phase diagram for water+ $n-$hexane system. The continuous red lines are the SAFT-$\gamma$ calculations for the vapour pressures of the pure components, the dashed blue lines are the SAFT-$\gamma$ predictions for the vapour-liquid critical line and three phase line, the dashed black lines correspond to $(P, x)$ sections.

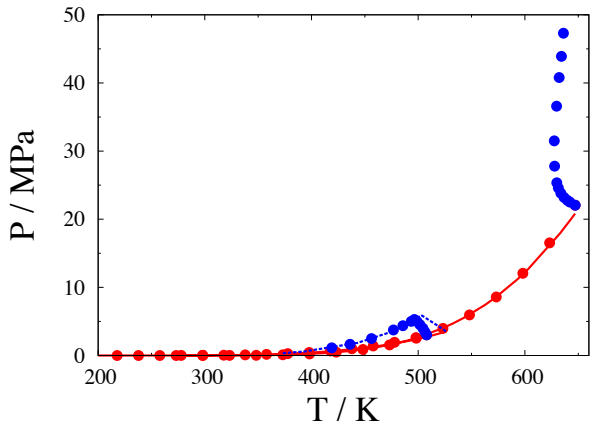



**Figure 5.13:** $PT$ projection of the phase diagram for water+ $n-$hexane binary mixture. The filled red circles correspond to the experimental vapour pressures of the pure components [195], the filled blue circles correspond to the experimental vapour-liquid critical line and three phase line [325]. The continuous red lines are the SAFT-$\gamma$ calculations for the vapour pressures of the pure components, the dashed blue lines are the SAFT-$\gamma$ predictions for the vapour-liquid critical line and three phase line.

The interfacial phenomena can be studied via MD simulations at pressures corresponding to the vapour-liquid and liquid-liquid equilibria. Figure 5.14 illustrates the interfacial region of the water+$n-$hexane mixture at $T = 473$ K and pressures below the three phase

line ($P$ =2.16 and 2.87 MPa) and above the three phase line ($P$ =6.25 MPa). As in the mixture with carbon dioxide, in mixtures with $n-$alkanes, water does not show any surface activity. At the pressures below the tree phase line, the $n-$hexane-vapour exhibits a positive surface activity which increases as the pressure increases. As the pressure increases above the three phase line pressure, $n-$hexane reaches the liquid bulk density and does not show any activity at the interface.

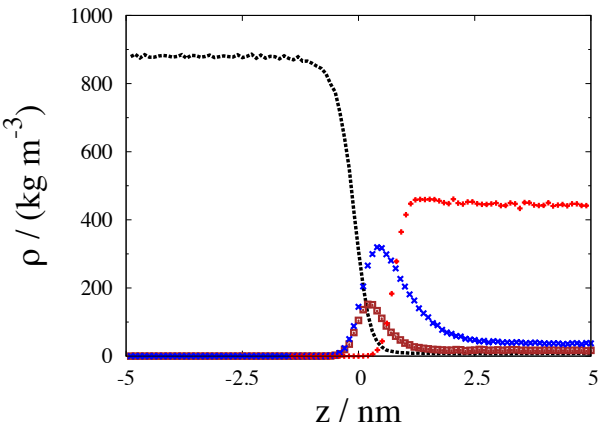



**Figure 5.14:** Density profiles at the interphase of the water+$n-$hexane mixture at $T = 473$ K at three different pressures calculated by MD simulations. The dashed line denotes water density profile at $P$ =2.16 MPa, the colour symbol represent $n-$hexane profiles at $P$ =2.16 MPa (brown squares), 2.87 MPa (blue crosses), and 6.25 MPa (red crosses).

The bulk density of $n-$hexane+water mixture can be predicted both by molecular simulation and the equation of state as shown in Figure 5.15. At $T = 473$ K, the MD simulations reproduce the theoretical results with a high accuracy. The barotropic density inversion for the $n-$alkane+water mixture was reported only for long-chain alkanes, starting from $n-$octacosane ($C_{28}H_{58}$) [325] and therefore is rightly not observed for the $n-$hexane+water mixture.

The force field can be tested for its transferability. If the model is transferable, the unlike interaction parameter which has been adjusted to reproduce the $n-$hexane+water mixture should also provide a quantitative agreement with experiment for any $n-$alkane+water mixture. This statement is verified by applying the same force field to the mixture of water with a long chain alkane, for instance $n-$dodecane. The experimental data for this mixture is selected at $T = 603$ K, which is above the UCEP. The theory captures the split into

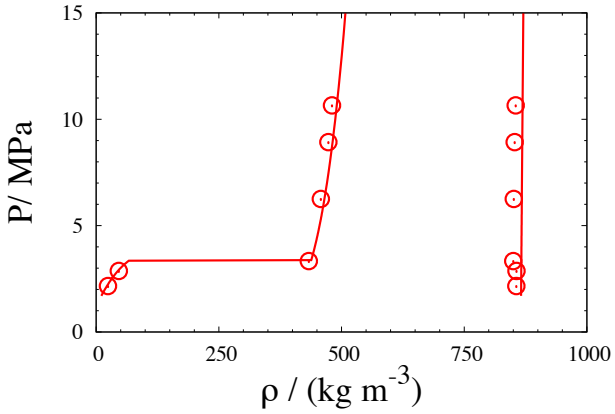



**Figure 5.15:** Pressure-density ($P\rho$) diagram for water+$n-$hexane mixture at the $T$ =473 K isotherms. The curves correspond to the predictions obtained by SAFT-$\gamma$ with $k_{ij} = 0.36$, the open circles are the MD simulation results for the SAFT-$\gamma$ CG Mie force field.

the vapour-liquid equilibrium at lower pressures and the fluid-liquid equilibrium at high pressures (Figure 5.16). MD simulations reproduce theoretical results both for the vapour-liquid equilibrium at the lower pressure and the fluid-liquid equilibrium at the higher pressure with high precision as can be seen in Figure 5.16. The three dimensional phase diagram is presented in Figure 5.17. The critical region of the vapour-liquid equilibrium is in line with the experimental data, whereas the fluid-liquid equilibrium shows greater immiscibility compared to the experiment. The discontinuity of the vapour-liquid critical line is visualised in $PT$-projection in Figure 5.18. Similar to the $n-$hexane+water mixture, the three phase line is in excellent agreement with experiment. Due to the over-prediction of the critical point of the $n-$alkanes, the critical pressure close to the critical point of $n-$dodecane is over-predicted. The minimum in the upper branch of the critical line is not reproduced. Nevertheless, the parameter set is capable of representing the main features of the phase diagram and is transferable not only to different state points but also to the related members of the homologous series.

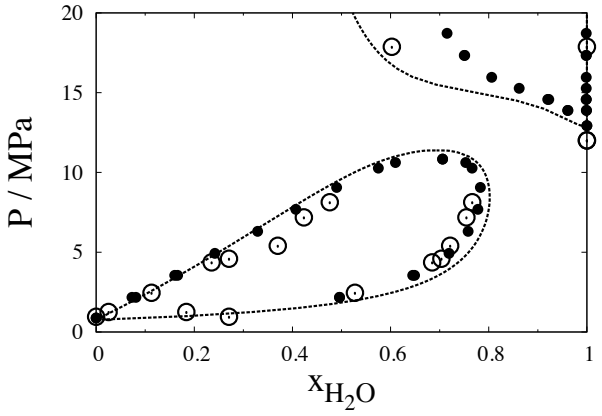



**Figure 5.16:** Isothermal pressure-composition ($Px$) phase diagram for water+ $n-$dodecane mixture at $T = 603$ K. Experimental data is taken from Ref. [326]. Legend as in Figure 5.10.

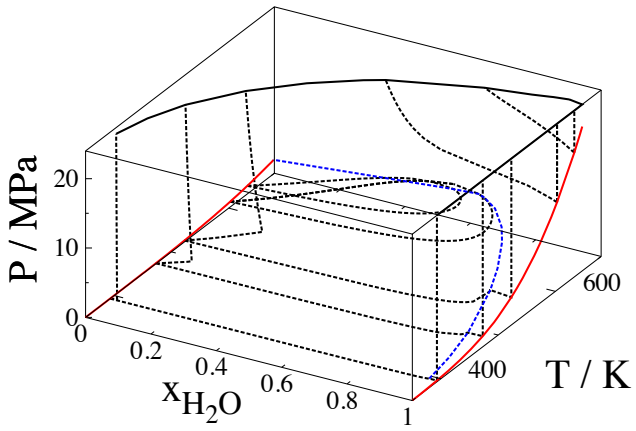



**Figure 5.17:** Three-dimensional ($PTx$) phase diagram for water+ $n-$dodecane system. Legend as in Figure 5.12.

### 5.5.3 $n-$alkane + $CO_2$ mixtures

The remaining binary system is the mixture of $n-$alkanes with carbon dioxide. This system is particular interesting because the phase behaviour changes from type II to type III with the increase of the chain length of the $n-$alkane. Capturing this behaviour with only one adjustable parameter for the entire homologous series is a challenging task for any equation of state [328]. The value of $k_{ij} = 0.08$ is found to provide the best match with experimental data of the vapour-liquid equilibrium data of $n-$hexane+$CO_2$ mixture

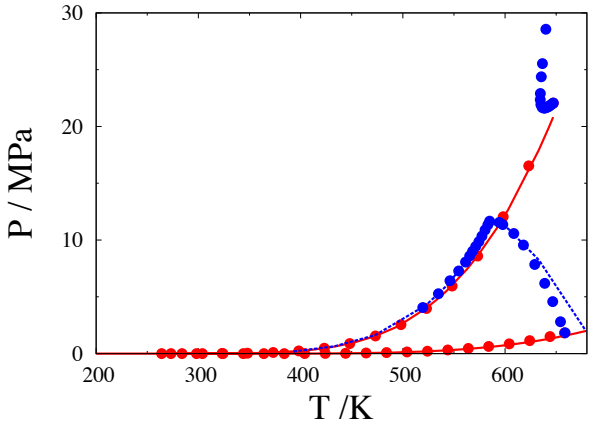



**Figure 5.18:** $PT$ projection of the phase diagram for water + $n-$dodecane binary mixture. Legend as in Figure 5.13.

at $T = 313$ K. This parameter set yields a good representation of the fluid-liquid equilibria at T=353 and 393 K for the pressures away from the critical point, whereas the critical point is overestimated by the theory (Figure 5.19). MD results performed at given temperatures compare well with the theory. The same force field is used to study the phase behaviour of the related members of the homologous series; the pressure-composition ($Px$) phase diagram for the long-chained $n$-pentadecane+carbon dioxide mixture is displayed in Figure 5.19 (bottom). The change of the phase behaviour compared to the shorter-chain alkanes is evident through the appearance of the liquid-liquid equilibrium at low temperature (here 292 K). In the theoretical prediction, the liquid-liquid immiscibility persists at high pressures, whereas the experimental results indicate that a total miscibility is obtained at pressures above 20 MPa. The vapour-liquid equilibrium is correctly described by the theory. At the temperatures above the critical point of carbon dioxide, the fluid-liquid equilibrium is accurately reproduced by the equation of state at low and intermediate pressures, the critical point of the mixture is however overestimated. Molecular Dynamics simulations performed based on the given force field reproduce the theoretical and experimental results with a high accuracy. The pressure-composition diagrams of carbon dioxide mixtures with other alkanes such as $n$-nonane, $n$-dodecane, $n$-octadecane are given in Figure F.2 in Appendix. For all members of the homologous series similar trends are observed: whereas the prediction by the theory is consistent with the experimental data

at low pressures, the critical point of mixtures are overestimated.

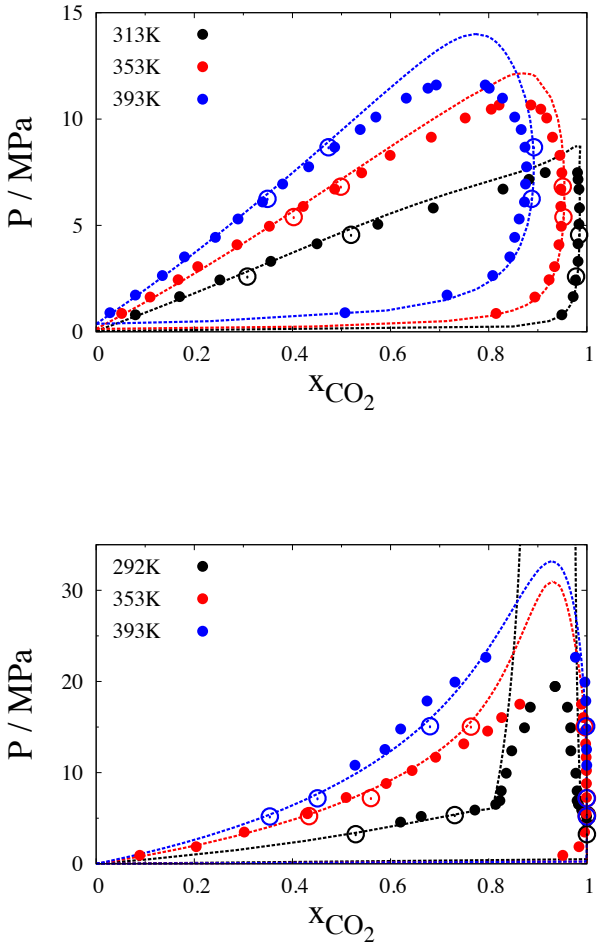



**Figure 5.19:** Pressure-composition ($Px$) phase diagram for mixtures of carbon dioxide with $n-$hexane and $n-$pentadecane at isotherms corresponding to $T$ =313, 353, 393 K and to $T$ =292, 353, and 393 K, respectively. The filled circles correspond to experimental vapour-liquid equilibrium data [196]. The dashed lines correspond to the predictions obtained by SAFT-$\gamma$ with $k_{ij} = 0.08$, the open circles are the MD simulation results for our SAFT-$\gamma$ CG Mie force field.

The change from type II behaviour for the short chain alkane to type III for the long chain alkanes is evident from the three dimensional representation of the mixtures of carbon dioxide with $n-$hexane and $n-$pentadecane in Figure 5.20 and the corresponding $PT$ projection in Figure 5.21. The phase behaviour of the $n-$hexane+carbon dioxide mixture is qualitatively reproduced, displaying the continuous vapour-liquid critical line,

which connects the critical points of the both components, and the liquid-liquid critical line, which merges with the three phase line in UCEP. Quantitatively, the vapour-liquid critical line is overestimated compared to the experimental data. By contrast, the discontinuous vapour-liquid critical line can be identified in $PT$ representation of $n-$pentadecane+carbon dioxide system, illustrating the type III behaviour. Starting from the critical point of $n-$pentadecane, the vapour-liquid critical line undergoes a maximum and a minimum with decreasing temperature until it finally merges with the liquid-liquid critical line. Again, only qualitative agreement with experiment is achieved: whereas the experimental data indicate that the critical line reaches its maximum at $\sim$ 385 K and 23 MPa and its minimum at $\sim$ 301 K and 12 MPa, the critical line predicted by the SAFT-$\gamma$ Mie EoS exhibits its maximum at $\sim$ 450 K and 35 MPa and its minimum at $\sim$ 350 K and 30 MPa.

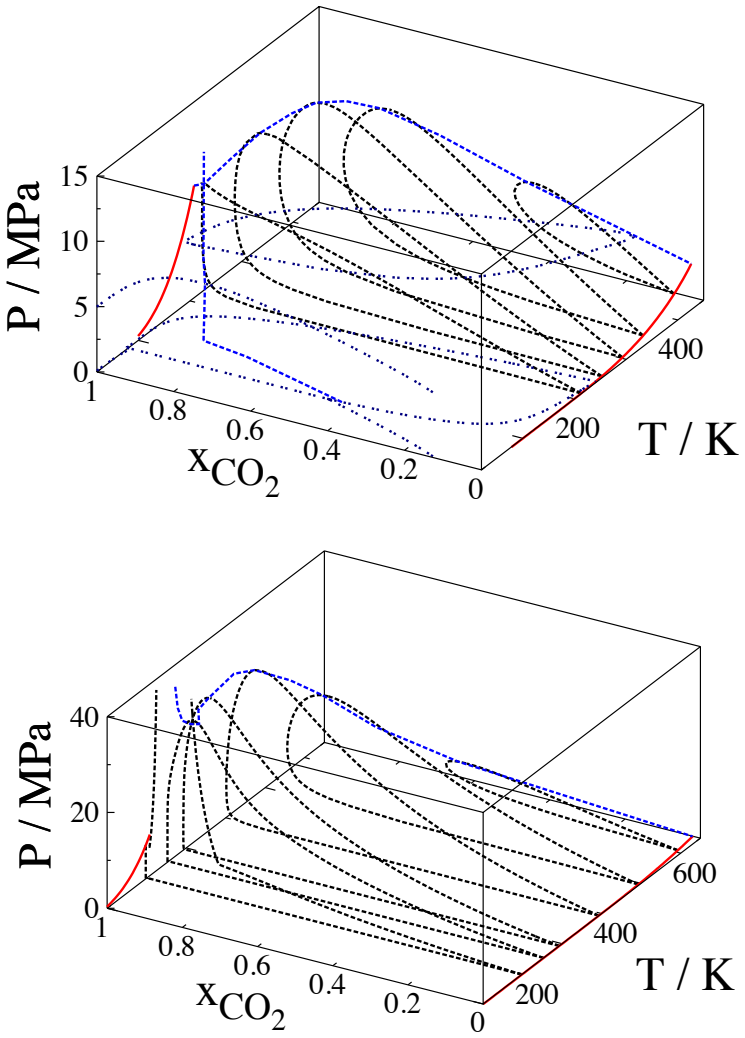



**Figure 5.20:** Three-dimensional ($PTx$) phase diagram for carbon dioxide+ $n-$hexane (a) and carbon dioxide+ $n-$pentadecane system (b). The continuous red lines are the SAFT-$\gamma$ calculations for the vapour pressures of the pure components, the dashed blue lines are the SAFT-$\gamma$ predictions for the vapour-liquid critical line and three phase line, the dashed black lines correspond to ($Px$) sections, the dotted blue line corresponds to ($Tx$) sections.

The adsorption phenomena along the interfacial region of the carbon dioxide+ $n-$pentadecane mixture is studied at the subcritical ($T = 292$ K) and supercritical ($T = 353$ K) temperatures of carbon dioxide (Figure 5.22). For all conditions, carbon dioxide adsorbs at the interface, whereas $n-$pentadecane does not exhibit any surface activity. At $P$ =5.2 MPa, the $CO_2$ density reaches approx. 600 kg$\times$m$^{-3}$ in the interface region at $T = 292$ K, whereas at $T = 353$ K the adsorption is reduced by factor 3. In this manner, the large excess adsorption at the low temperature acts as an indicator of being at the vicinity of

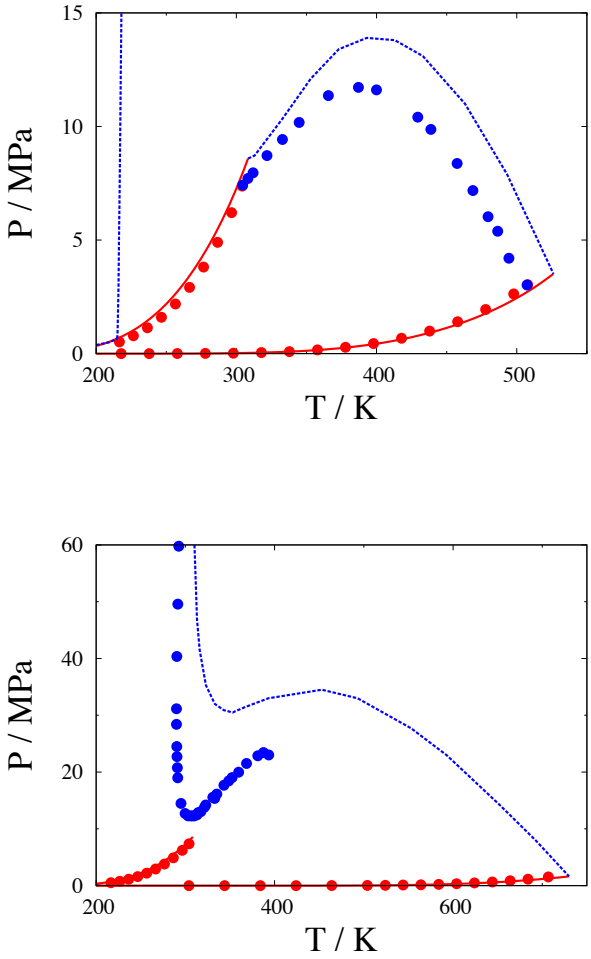



**Figure 5.21:** *PT* projection of the phase diagram for the carbon dioxide+$n-$hexane and carbon dioxide+ $n-$pentadecane binary mixture. The filled red circles correspond to the experimental vapour pressures of the pure components [195], the filled blue circles correspond to the experimental vapour-liquid critical line [362]. The continuous red lines are the SAFT-$\gamma$ calculations for the vapour pressures of the pure components, the dashed blue lines are the SAFT-$\gamma$ predictions for the vapour-liquid critical line and three phase line.

a liquid-liquid equilibrium (Figure 5.19). With increasing pressure, a higher $CO_2$ density in both phases and at the interface region is observed.

In following, the bulk density of the carbon dioxide+$n-$pentadecane mixture is studied at different temperatures and pressures and compared to the experimental data. The barotropic density inversion has been experimentally observed for mixtures of carbon dioxide with alkane homologous that exhibit the type III behaviour. The $n-$tetradecane

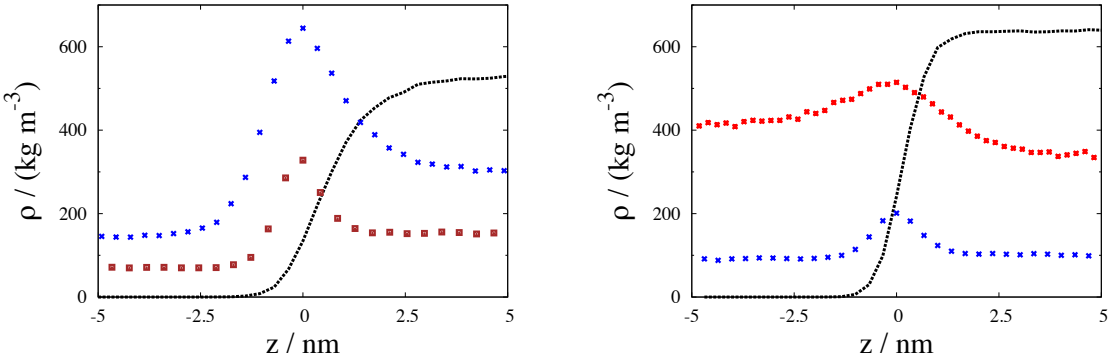



**Figure 5.22:** Density profiles at the interface of the carbon dioxide+ $n-$pentadecane mixture at $T = 292$ K (left) and 353 K (right) and different pressures calculated by MD simulations. The dashed line denotes the $n-$pentadecane density profile at $P$ =5.2 MPa, the symbols represent the carbon dioxide profiles at $P$ =3.2 MPa (brown squares), 5.2 MPa (blue crosses), and 15.4 MPa (red stars).

was reported as the shortest alkane which exhibits the barotropic density inversion in the mixtures with $CO_2$ [320]. The experimental densities reported by Tanaka *et al.* [363] at $T = 313$ K are compared to the prediction by SAFT-$\gamma$ Mie, thereby revealing an excellent agreement (Figure 5.23). Further isotherms at $T$ =292, 353, and 393 K are calculated by the equation of state and compared to the molecular simulation, also showing a very good agreement between the two. The barotropic density inversion is observed at low temperatures. The prediction by the SAFT EoS indicates the mass density inversion at $P$ =9.5 MPa and $T$ =292 K, and at $P$ =22 MPa and $T = 313$ K.

### 5.5.4 Ternary $n-$decane +$CO_2$+$H_2O$ mixture

One of the most recent and comprehensive studies of the ternary $n-$decane+carbon dioxide+water mixture is that of Forte *et al.* [364]. The authors pointed out that a large part of the phase diagram is occupied by the three phase region due to immiscibility of the binary systems comprising the mixture. In this work, we compare the experimental data with the results from the SAFT-$\gamma$ Mie EoS and the CG simulations. No additional parameter fitting is requited, as the same unlike interaction parameters are used as in the binary systems comprising the ternary mixture.

Table 5.2 and Figure 5.24 illustrate the comparison between the experimental [364] equilib-

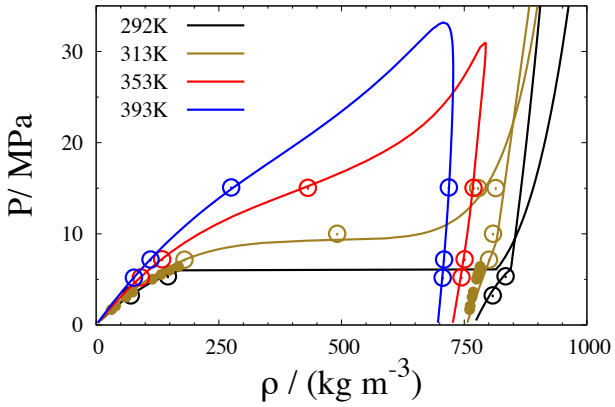



**Figure 5.23:** Pressure-density ($P\rho$) diagram for carbon dioxide+$n-$pentadecane mixture at the isotherms $T$ =292, 313, 353, and 393 K. The filled circles correspond to experimental density data for the saturated phases at $T$ = 313 K [363]. The curves correspond to the predictions obtained by SAFT-$\gamma$ with $k_{ij}$ = 0.08, the open circles are the MD simulation results for our SAFT-$\gamma$ CG Mie force field.

rium phase compositions and those predicted by the SAFT-$\gamma$ Mie EoS and MD simulations of the ternary mixture at $T$ = 393.14 K and four different pressures away from critical region: $P$ =6.02, 7.58, 9.59, and 11.81 MPa. The phase diagram is dominated by the three phase region, which decreases with increasing pressure. The SAFT-$\gamma$ Mie EoS reproduces the experimental measurements of the three phase region very accurately. Additionally to the large three phase region, the SAFT EoS predicts the substantial two phase regions. Only very small regions of single-phase equilibria can be detected in the margins of the diagram. The MD simulations of the three phase region reproduce the compositions of the $CO_2$ and $H_2O$ rich phase in a good agreement with theoretical predictions. However, the solubility of carbon dioxide in the alkane rich phase is overestimated for all pressures. Nevertheless, the trend of increasing solubility with pressure can be correctly captured by simulations.

The density profile at $P$ = 6.02 MPa is displayed in Figure 5.25. Three phases (water-rich phase $L_1$, $n-$decane-rich phase $L_2$, and carbon dioxide rich phase $G$) are clearly distinguished by the density, which is also apparent from the MD simulation snapshot. Carbon dioxide exhibits a large adsorption peak at the interface between the water-rich and $n-$decane-rich phases and a small adsorption peak at the interface between $n-$decane-

rich and carbon dioxide rich phases. With increasing pressure, the density and the width of the interfaces were also found to increase.

**Table 5.2:** Comparison between the experimental [364] equilibrium phase compositions and those predicted by the SAFT-$\gamma$ Mie EoS and MD simulations of the $n-$decane+carbon dioxide+water system at $T = 393.15$ K and different pressures.

| | | expt. | | | SAFT | | | MD | | |
|---|---|---|---|---|---|---|---|---|---|---|
| P/ Mpa | Phase | $x^{expt}_{C_{10}H_{22}}$ | $x^{expt}_{CO_2}$ | $x^{expt}_{H_2O}$ | $x^{SAFT}_{C_{10}H_{22}}$ | $x^{SAFT}_{CO_2}$ | $x^{SAFT}_{H_2O}$ | $x^{MD}_{C_{10}H_{22}}$ | $x^{MD}_{CO_2}$ | $x^{MD}_{H_2O}$ |
| 6.02 | I | 4.50E-06 | 0.0086 | 0.9914 | 0 | 0.0031 | 0.9969 | 0 | 0.0076 | 0.9924 |
| | II | 0.0097 | 0.9542 | 0.0361 | 0.0079 | 0.9454 | 0.0467 | 0.0137 | 0.9447 | 0.0416 |
| | III | 0.6477 | 0.3191 | 0.0332 | 0.6598 | 0.3197 | 0.0204 | 0.5932 | 0.3757 | 0.0311 |
| 7.58 | I | 2.70E-06 | 0.0107 | 0.9893 | 0 | 0.0038 | 0.9962 | 0 | 0.0074 | 0.9926 |
| | II | 0.0078 | 0.9605 | 0.0317 | 0.0084 | 0.9503 | 0.0413 | 0.0130 | 0.9473 | 0.0397 |
| | III | 0.5859 | 0.3795 | 0.0346 | 0.5924 | 0.3862 | 0.0214 | 0.4906 | 0.4751 | 0.0343 |
| 9.59 | I | 2.60E-06 | 0.0127 | 0.9873 | 0 | 0.0046 | 0.9954 | 0 | 0.0065 | 0.9935 |
| | II | 0.0112 | 0.9636 | 0.0252 | 0.0099 | 0.9522 | 0.0379 | 0.0142 | 0.9482 | 0.0377 |
| | III | 0.4857 | 0.4821 | 0.0322 | 0.5135 | 0.4636 | 0.0229 | 0.4077 | 0.5527 | 0.0396 |
| 11.81 | I | 8.30E-07 | 0.0145 | 0.9855 | 1.44E-09 | 0.0054 | 0.9946 | 0 | 0.0072 | 0.9928 |
| | II | 0.0184 | 0.9569 | 0.0247 | 0.0131 | 0.9500 | 0.0369 | 0.0134 | 0.9481 | 0.0386 |
| | III | 0.4006 | 0.566 | 0.0334 | 0.4349 | 0.5402 | 0.0250 | 0.3670 | 0.5961 | 0.0370 |

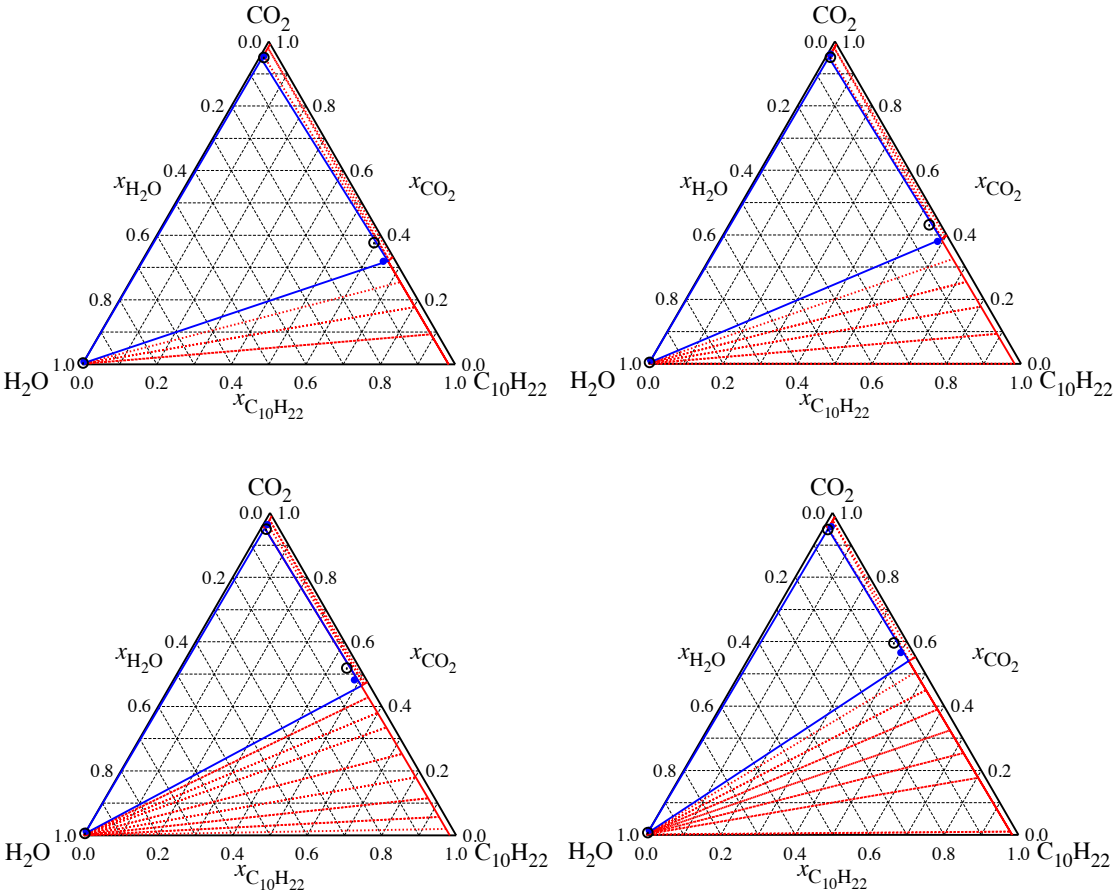



**Figure 5.24:** Phase composition diagram for the $n-$decane+carbon dioxide+water system at $T = 393.14K$ and $P$ =6.02, 7.58, 9.35, and 11.81 MPa, respectively. The filled circles are the experimental values for the three-phase equilibrium region from Ref. [364]; the continuous blue lines are SAFT-$\gamma$ Mie prediction of the three-phase equilibrium region, the dashed red lines are SAFT predictions of the tie-lines between the two coexisting phases , the continuos red lines separate the two phase from the one phase region. The open black circles are the CG simulation results for the three phase region.

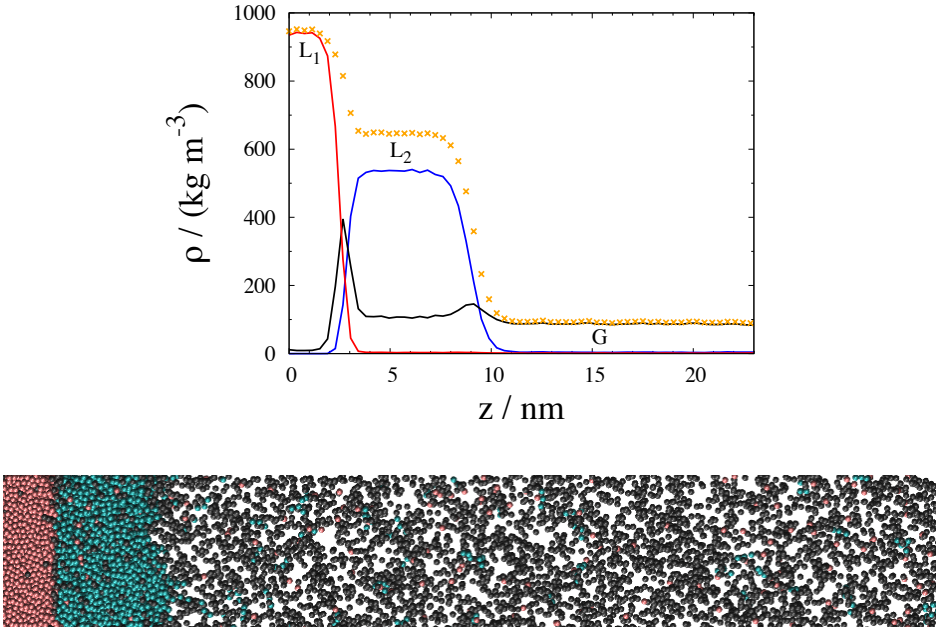



**Figure 5.25:** Total density profile of the tree-phase region of the ternary $n-$decane+carbon dioxide+water mixture at $T = 393.14$ K and $P = 6.02$ MPa and the global composition of $x_{C10H22} = 0.1$, $x_{CO2} = 0.5$ obtained via MD simulation. $L_1$ denotes water rich phase, $L_2$ denotes $n-$decane-rich phase, and $G$ denotes carbon dioxide-rich phase. In the snapshot from molecular simulations, carbon dioxide molecules are displayed in grey, $n-$decane in blue, and water in red, corresponding to the lines of the same colour in the density profiles.

## 5.6 Conclusions

In this chapter, the SAFT-$\gamma$ Mie force fields are developed to study binary and ternary mixtures comprising carbon dioxide, $n-$alkanes, and water. The models of pure components, developed in Ref. [22] and Chapters 3 and 4, respectively, are used to describe the like-like interactions. The unlike interactions, which are obtained by considering the fluid-phase equilibria of the corresponding binary mixture, can be described by only one additional adjustable parameter, which is found to be transferable to different state points, different properties and also to the components of related homologous. Molecular Dynamics simulations, which are carried out subsequently to verify the prediction of the phase equilibria by the equation of state, show an excellent agreement with the theory and the experiment.

In the studied mixture of carbon dioxide with water, an accurate description of the vapour-

liquid and fluid-liquid equlibria is obtained. The critical and three phase line are predicted in a quantitative agreement with experimental data. The developed CG force fields are shown to deliver a comparable performance when compared to the more sophisticated atomistic models. An accurate description of the vapour-liquid and fluid-liquid equlibria is also observed for the mixture of water with $n-$alkanes, which is characterised by the extreme immiscibility. The azeotropic points and the tree phase line are, again, predicted with a high accuracy. The last studied binary mixture, comprising carbon dioxide and $n-$alkanes, exhibit a transition from type II to type III with increasing chain length of an alkane, in line with the experimental findings. Finally, the ternary mixture, described only by the binary interaction parameters, is reproduced in a reasonable agreement with experimental measurements, recently performed by Forte *et al.* [364].

In this manner, the force fields can be considered as robust and transferable and can be applied to predict properties that were not used in the parameterisation procedure. The entire phase diagram is hardly accessible either to an experiment or molecular simulation. Whereas the molecular simulations are time consuming, the experiments may not be even possible to perform at extreme temperatures and pressures. On the other hand, the molecular simulations provide information about the structural and interfacial properties, which are not accessible to the equation of state. Since the same molecular models are used both in the molecular simulation and the equation of state, the theory-simulation-experiment relationship can be used in a very flexible way, thereby offering a unique possibility of the multi-scale modelling of complex fluids. Hence, the most appropriate modelling technique can be chosen depending on the properties of interest.

# Chapter 6

# Aqueous non-ionic surfactant mixtures

## 6.1 Chapter overview

In this chapter we apply the SAFT-$\gamma$ force field methodology to develop a force field for non-ionic surfactant solutions from the alkyl polyoxyethylene glycol family, abbreviated as $C_iE_j$. The surfactants are described within the group-contribution framework. The constructive moieties (alkanes, ethers and alcohol endgroup) and their cross mixtures are parameterised to reproduce the vapour-liquid equilibria properties of the pure components or the liquid-liquid equilibrium of the corresponding mixture, as described in Chapter 3. The unlike interactions between the surfactant groups and water are adjusted to the interfacial tensions for the insoluble components and to the enthalpy of mixture for the soluble components. The resulting model is robust and transferable, and allows for the prediction of key structural and interfacial properties. The Molecular Dynamics simulations reveal a spontaneous micelle formation at low surfactant concentrations and a self-assembly into a bilayer at high surfactant concentrations. Aggregation numbers, critical micelle concentrations, area per molecule, surface excess properties and bilayer thickness are found in a quantitative agreement with experimental data.

## 6.2 Introduction

Surfactants embrace a wide range of applications from textile manufacturing, emulsification, drug delivery to electronic printing, biotechnology and microelectronics [11]. In aqueous solutions, surfactant molecules self-assemble spontaneously into distinct aggregates [365]. Depending on different parameters, such as temperature, concentration and the chain length of amphiphile, various phase morphologies, such as micellar, hexagonal, lamellar, inverse micellar, inverse hexagonal and gyroid phases have been observed experimentally [366]. Although the complexity of amphiphilic systems may preclude a theoretical treatment, it is ideally suited to be studied by molecular simulation. Computer modelling of the structural and mesophase behaviour of the surfactant systems does however present formidable challenges, as the size of the expected aggregates, the relatively slow dynamics and the delicate interplay between energetic and entropic effects contribute to the requirement of a large computational effort.

The detailed atomistic modelling of a surfactant, while technically possible, only allows for the study of rather small system and time scales, which may be orders of magnitude away from those required to study self-assembly. For this type of systems, mesoscale methods in general and coarse-graining in particular have been proven useful as they help to overcome some of the time and length scale limitations.

In this chapter, we will describe the development of a CG model for non-ionic alkyl polyoxyethylene glycol surfactants, known as $C_iE_j$. Possibly the earliest simulation of coarse-grained amphiphilic molecules was performed by Smit *et al.* [131]. The authors studied the self-aggregation of non-ionic surfactants into micelles by using only three types of particles, labeled as head, tail and water. All constituent beads interacted via the Lennard-Jones potential with the same energy parameter $\epsilon$ and size parameter $\sigma$, differentiating from each other only in the cutoff radius $R_{ij}^c$, thereby resulting in the different strength of interaction. Despite the simplicity of that model, the authors observed not only self-assembly with micellar size distribution, but also dynamic process of entering of a single surfactant into micelles, a single surfactant leaving micelles, the fusion of two micelles, and the slow break-down of a micelle. Goetz and Lipowski [367] studied lipid membranes using a similar model, where lipid consisted of a hydrophilic head, hydrophobic tail and the solvent LJ beads. Depending on the concentration, spontaneous self-assembly into

micelles or fluid bilayers was observed. The authors also studied the effect of the size parameter and the tail flexibility of lipids on the structure and lateral stress distribution in the bilayer. Similarly, Fujiwara *et al.* [368] investigated self-assembly of single surfactants into micelles and the effect of the hydrophobic attraction and hydrophilic repulsion on the micellar shape. By changing the interaction parameters, the authors observed three different micellar morphologies: discs, cylindrical and spherical micelles. The above models are commonly labeled "toy models". This derogatory name is misleading, because these models take into account correct relative energies amongst the beads and the correct space-filling profiles. In this way, the appropriate balance between enthalpy and entropy required for a qualitative description of self-assembly is achieved. The works by Crane *et al.* [133] or Glotzer and co-workers [369] illustrate the variety of lyotropic phase behaviour that can be obtained by using these "toy models". However, a more robust link to the chemical nature of the components is needed in order to account for the quantitative description.

Several approaches have been applied to derive the CG force fields for the complex molecules, including liquid crystals or biological molecules. Within the bottom-up approach, the common techniques of the Iterative Boltzmann Inversion or force matching have been extensively used in this research area. The IBI technique is a most employed method in the study of polymers [370]. Liquid crystals have also been studied by the IBI method and the transition from the isotropic to the liquid crystalline phase could be qualitatively reproduced [371, 372]. The force matching has been applied to investigate complex fluids such as the ionic liquids [128], lipid bilayers [125, 373, 374], peptides [129], and carbohydrates [126, 127]. Within the top-down approach, the semi-empirical models have been widely used for a number of different systems. By using the group contribution approach, the Martini force field [78] is able to describe a variety of surfactants and phospholipids. Despite the limited set of parameters and the minimalistic approach of the model, the Martini force field is able to reproduce several key structural and thermodynamic properties, thereby becoming one of the mostly used force fields in the study of biological systems. Among a large number of examples, Lee and Pastor *et al.* [147] studied the self-assembly of aqueous lipid mixtures (dipalmitoylglycerophosphocholine (DPPC) with polyethylene glycol-grafted dipalmitoylglycerophosphoethanolamine (DPPE)), thereby detecting aggregation into liposomes, bicelles, and micelles in qualita-

tive agreement with experiment data for the aggregate sizes. Sangwai *et al.* [148] observed a self-assembly of surfactant systems (cetyltrimethylammonium chloride (CTAC)) to various micellar structures such as spheres, rods, or lamellae and their shape transition driven by addition of various salts. Another successful CG force field for surfactants [85, 86, 161], phospholipids [84, 375], peptides [88], and ionic liquids [376] has been developed by Klein and co-workers, using the thermodynamic macroscopic data. Based on the Mie (9-6) or Mie (12-4) intermolecular potential, the force fields were parameterised to reproduce the experimental surface tensions and liquid densities at ambient conditions, while the unlike interaction parameters were adjusted to reproduce the solvation free energies. The spontaneous self-assembly into micelles, hexagonal and lamellar bulk structures can be predicted by the Klein model in agreement with experimental observations at the corresponding concentration and thermodynamic conditions for those phases. Alongside with density and tension, various different macroscopic properties can be used in the parameterisation of a CG force field. For instance, in the ELBA (ELectrostatic-BAsed) [162] CG model for lipids, the parameters were obtained by adjusting to the experimental volume and area per lipid, and the average segmental tail order parameter of fluid DMPC. Recently, Vishnyakov *et al.* [163] developed a CG model for the non-ionic surfactants in the dissipative particle dynamics (DPD) simulations by fitting to the infinite dilution activity coefficients of the corresponding binary mixtures. For an extensive overview on the "top-down" coarse grained models along with the "bottom-up" models in biomolecular systems, the reader is referred to the recent reviews by Marrink *et al.* [7], Noid [93] and Brini *et al.* [92].

Notwithstanding, the inconvenience of those semi-empirical CG models is given through the fact that the search for the optimal force field parameters, on the basis of relating the output of simulations to target macroscopic properties, is performed by trial and error [82]. The procedure requires a significant computational effort but still does not guarantee an optimal solution. On the other hand, by using an analytical model, such as an equation of state, a wide parameter space can be explored, thereby yielding an optimal parameter set. This provides an advantage in terms of both, the parameter optimisation speed and the robustness of the model. Several thermodynamic properties could be determined using a theoretical approach in the study of biological systems. One of the examples includes the work by Chen *et al.* [377], where the authors proposed to calculate the critical micelle concentration (CMC) for aqueous $C_iE_j$ solutions using an

NRTL-like segment-based local composition model [378]. Based on Chen's work, Li *et al.* used a segment-based UNIQUAC (s-UNIQUAC) model and the original version of the SAFT EoS [44] to estimate the CMC of non-ionic [379] and ionic [380] surfactant solutions. Flores *et al.* used the UNIFAC model [381] to predict the CMC of non-ionic surfactants in aqueous [382] and non-aqueous solutions [383] using the phase separation approach, where the micellar phase is approximated as a second liquid phase resulting from the liquid-liquid equilibrium between the solvent and the surfactant. Sassen *et al.* [200] used the UNIQUAC model [384] to predict the phase behaviour in the ternary system $C_7E_5 + H_2O + C_{12}H_{26}$. Rudolph *et al.* [385] used the Peng-Robinson EoS and the UNIQUAC model to predict the cloud point curves for binary water-non-ionic surfactant system.

Although the theoretical models have an advantage in computational efficiency, they cannot describe the complex structures of the amphiphilic solutions. The molecular simulations provide a wealth of information not accessible through the theory, such as aggregation numbers and mesophase formation, structural, interfacial and transport properties, etc. Combining both approaches would yield the merged assets such as a more efficient sampling of the free energy landscape in the parameter estimation procedure alongside with information about the structure and the dynamics of the system. Unfortunately, the $g^E$-models such as NRTL, UNIQUAC, and UNIFAC are described in terms of the segment and area parameters, and therefore cannot be related to intermolecular potentials required in the molecular simulations. By contrast, the SAFT EoS uses a well-described molecular basis. Several versions of the SAFT EoS have already been employed to describe the surfactant phase behaviour. For instance, Garcia-Lisbona *et al.* [386, 387] used the SAFT-HS approach to predict the liquid-liquid equilibria of binary alkyl polyoxyethylene glycol surfactant mixtures. Clark *et al.* [388] employed the SAFT-VR EoS to study the phase behaviour of long chain polyoxyethylene glycol mixtures. However, in these cases no attempt was made to translate the EoS parameters into a molecular simulation force field. The novelty of the approach presented here, is that we use the analytical SAFT-$\gamma$ Mie EoS to derive the set of intermolecular parameters, which can be subsequently used as a direct input in molecular simulation. Encouraged by the success of previous publications by Avendaño *et al.* [22, 66, 67] and the quality of prediction of the chain molecules and binary mixtures in Chapters 3 and 5, we apply the SAFT-$\gamma$ Mie approach to study complex amphiphilic solutions.

## 6.3 Methods and models

### 6.3.1 Simulation details

We use Molecular Dynamics simulations in isobaric-isothermal ($NPT$) ensemble, where the number of particles $N$, temperature $T$, and the pressure $P$ are kept constant. The system temperature is maintained using a Nosé-Hoover thermostat [188, 189] with the coupling time constant being 1.0 ps, whereas the pressure is maintained by means of a Parrinello-Rahman barostat with a relaxation constant of 10.0 ps. The cutoff radius is chosen to be 2.0 nm, in consistency with the Model 2 for water. The equation of motion is integrated by using the leap-frog algorithm [18] with a time step size of 0.01 ps. The simulations are carried out with the Gromacs package, version 4.5.5 [190]. A detailed simulation set-up is given for the low and the high surfactant concentrations in Sections 6.4.1 and 6.4.5, respectively.

### 6.3.2 SAFT-$\gamma$ molecular CG force field development

In this chapter, we present a general CG alkyl polyoxyethylene glycol surfactant model, containing alkyl, ethoxy and the end groups. Since the surfactants comprise the functional groups of a different chemical nature, it is reasonable to describe the surfactant molecule within the group-contribution approach.

Hence, the molecule is split into three type of functional groups according to their chemical nature, $CM$, $EO$ and $OA$, where $CM$ is the alkyl group and is parameterised to reproduce pure component properties of $n-$alkanes (Section 3.4.1). $EO$ is the ethoxy-group, parameterised to reproduce the VLE properties of ethers (Section 3.4.3), whereas the $OA$ group is the surfactant end-group, parameterised to reproduce properties of the triethylene glycol (Section 3.4.4). This breakdown is not unique, however it resembles the grouping found in other CG models [85, 149]. The chosen mapping of the $C_{10}E_4$ surfactant is presented in Figure 3.6. Since one of the key properties in the surfactant mixtures is interfacial tension, the water is described by Model 2 (Section 4.4.4), a CG parameterisation fit to the saturated liquid density and surface tension. The model is based on the Mie (8-6) potential with temperature dependent size and energy parameters, $\sigma/\text{Å} = \sigma(T) = -6.455 \times 10^{-9}T^3 + 9.1 \times 10^{-6}T^2 - 4.291 \times 10^{-3}T + 3.543$ and $(\epsilon/k_B)/K = (\epsilon/k_B)(T) = -4.806 \times 10^{-4}T^2 + 0.6107T + 165.9$. Apart from the like in-

teractions, the parameter matrix involves the estimation of the unlike interactions. The unlike interactions between the surfactant CG beads are derived from the parameterisation of heteronuclear models and adjusting the unlike energy parameters to the VLE pure component data (Section 3.4.5). An exception is the unlike interaction between the *OA* and *EO* CG beads, which are adjusted the liquid-liquid equilibrium data, in order to ensure the highly non-ideal phase behaviour of the mixture. The interaction parameters between the surfactant CG beads and water are derived in the following section, based on the modified Lorentz-Berthelot combining rules given by Eqs. 2.57, 2.59, and 2.60. The unlike energy parameter $\epsilon_{ij}$ is adjusted directly to the experimental data, whereas $\sigma_{ij}$ and $\lambda_{r,ij}$ are fixed according to the combining rule (Eqs. 2.57 and 2.59) unless indicated otherwise.

### 6.3.3 Unlike interactions between water and alkanes

Alkanes and water phase separate into two liquid phases at ambient conditions. One could use SAFT EoS to fit to the mutual solubilities, however, we found that an optimal parameter set in both phases is not universal. Since one of our main requirements for the force field is its robustness and transferability, we are interested in a force field that is able to reproduce a correct phase behaviour at both, low concentration limit and high concentration limit. One of the key properties in the study of phase morphologies and structural properties of surfactants is the interfacial tension and it has been used by several authors as a descriptor in the development of the CG models for surfactants [85, 389]. Following this approach, the parameter estimation is carried out to reproduce the interfacial tension of hexane-water system at $T = 298.15$ K using MD simulations. The optimal parameter is found for $k_{ij} = 0.3$, which appears to be transferable for calculation of interfacial tensions of mixtures of water with longer alkanes (Table 6.1).

**Table 6.1:** Comparison of the experimental interfacial tensions [390] in $[mN \times m^{-1}]$ at $T = 298.15$ K with those calculated by MD simulation using $k_{ij} = 0.3$ for *n*-hexane, *n*-nonane, *n*-dodecane.

| Component | $C_6H_{14}$ | $C_9H_{20}$ | $C_{12}H_{26}$ |
|---|---|---|---|
| calc. ($k_{ij} = 0.3$) | 49.11 | 51.95 | 52.19 |
| expt. | 50.29 | 51.57 | 52.44 |

### 6.3.4 Unlike interactions between water and polar groups

Since there is a complete miscibility between water and substances containing exclusively polar beads, unlike interaction parameters between water and polar groups are fitted to the excess enthalpies of the mixture. Triethylene glycol-water system at $T = 298.15$ K is used to optimise parameters between $OA$ and $W$ groups. It was found that $\lambda_{r,ij}$ has an impact on the width of the curve, whereas $\epsilon_{ij}$ affects its well-depth. The optimal $OA$-$W$ unlike interaction parameter, $\lambda_{r,ij}$=17, $\epsilon_{ij}/k_B$=497 K, can reproduce the excess enthalpy of the mixture in a good agreement with the experimental data. (Figure 6.1).
The excess enthalpy of mixture of 2,5,8,11-tetraoxadodecane ($C_8H_{18}O_4$) and water at $T =$ 298.15 K is used to parameterise the unlike interactions between $EO$ and $W$ groups. The parameter set that provides the lowest deviation from the experimental data is $\lambda_{r,ij}$=11.94, $\epsilon_{ij}/k_B$=435 K. The quality of representation can be seen in Figure 6.1 (bottom).
All intermolecular interaction parameters are summarised in Table C.2; the kind of the experimental data that has been used to obtain these parameters is given in Table C.1 in Appendix.

#### Hydrophilic-lipophilic balance

One of the key properties that controls the phase boundaries of the surfactant-water mixture is the hydrophilic-lipophilic balance of the system. Even a slight unbalance in the hydrophilic-lipophilic interactions is enough to shift the phase boundary so that the phase morphologies that are observed experimentally at the corresponding concentration and thermodynamic conditions cannot be reproduced. In this study, we observed that this hydrophilic-lipophilic balance is extremely sensitive to the unlike energy interactions between the solvent and the surfactant groups, so that a small change in the parameter value can cause significant changes in the aggregation number and the density of the micelle, as well as a substantial shift in the phase boundaries. For instance, if the interactions between the water and the alkyl groups are chosen too weak, all surfactants self-assemble into one large aggregate, thereby forming a second phase. On the other hand, if these interactions are too strong, a bilayer structure is not stable at high concentrations; also the aggregation number and density of micelles are fallaciously low.
A rough estimate of the hydrophilic-lipophilic balance of the system can be achieved by

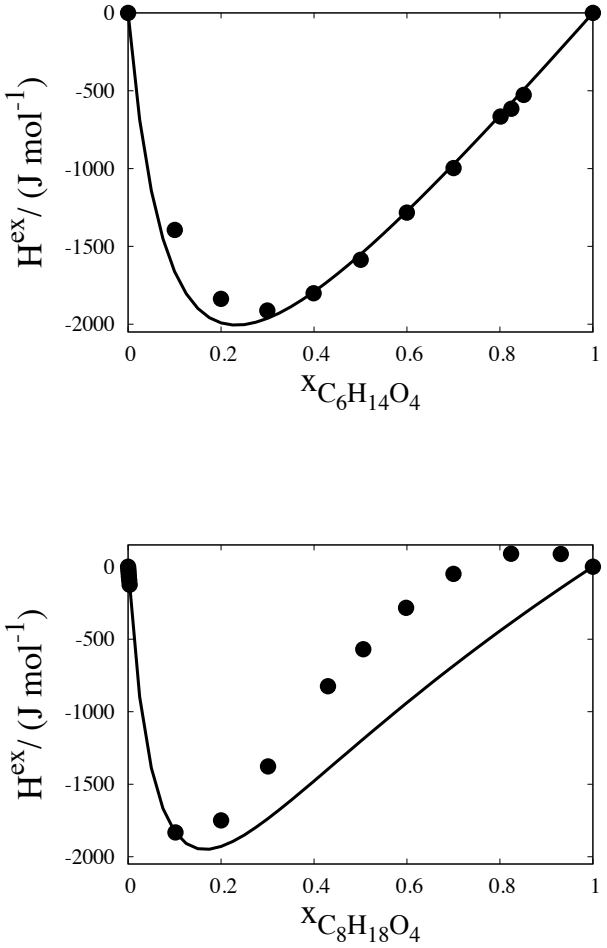



**Figure 6.1:** The excess enthalpy of the triethylene glycol-water (top) and $C_8H_{18}O_4$-water mixture (bottom) at $T = 298.15K$ and $P = 1MPa$. The filled circles denote the experimental data [196] and the lines are the prediction with SAFT-$\gamma$ Mie EoS. The continuous lines are calculations with SAFT-$\gamma$ Mie EoS based on the parameter sets $\lambda_{r,ij}$=17, $\epsilon_{ij}/k_B$=497K in (top) and $\lambda_{r,ij}$=11.94, $\epsilon_{ij}/k_B$=435 K in (bottom), respectively.

estimating the internal energies between moieties, which can be related to the van der Waals integrated energy

$$\alpha = \int_{\sigma}^{\infty} u(r) r^2 dr, \tag{6.1}$$

which for the Mie fluid takes the form

$$\alpha = \sigma^3 \epsilon \mathcal{C} \left( \frac{1}{\lambda_a - 3} - \frac{1}{\lambda_r - 3} \right), \tag{6.2}$$

with $\mathcal{C}$ given by Eq. 2.18. The term $\alpha$ can be used to compare the strength of the molecular interactions between the beads, and is obtained directly from the corresponding intermolecular potential parameters.

The hydrophilic-lipophilic balance can be estimated as a sum over all interactions in the system, $\alpha^{total}$. Our SAFT-$\gamma$ Mie CG force field results in $\alpha^{total} = 1.14 \times 10^{-2}\ Jm^3$. For the studied $C_{10}E_4 + H_2O$ system, it was found that the experimentally observed phases, such as the lamellar phase at the high concentration limit and the dynamic micellar equilibrium at the low concentration limit can be reproduced only if the value of $\alpha^{total}$ is in a very narrow range between $(1.13 - 1.16) \times 10^{-2}\ Jm^3$.

In Figure 6.2, the der Waals energies of our CG model are compared to commonly used Martini [78] and Klein models [85]. Our CG model shows similar results to the Klein model. This finding can be ascribed to the fact that both force fields are parameterised to reproduce the density and interfacial tension and are based on a similar CG mapping (3-to-1). The Martini model, which is based on the (4-to-1) CG mapping and is parameterised to reproduce properties as the free energy of hydration, free energy of vaporisation and the partitioning free energy, differs significantly from both other models.

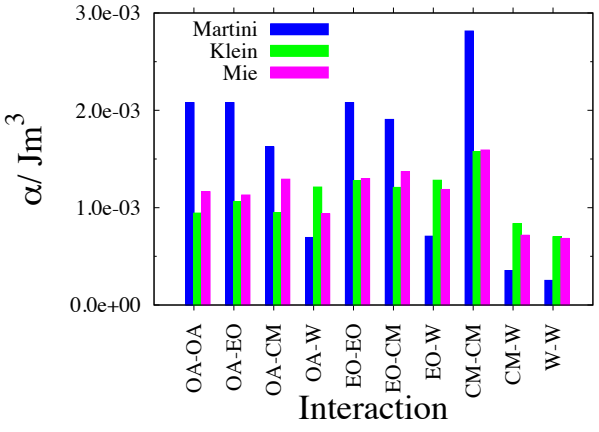



**Figure 6.2:** Comparison of the van der Waals energy values of Martini, Klein, and CG models developed in this work.

## 6.4 Results and discussion

### 6.4.1 Morphology of the micelles as a function of the surfactant concentration

The molecular simulations of the micellar solutions at the low surfactant concentrations are performed at $T = 298.15$ K and $P = 1$ MPa in $NPT$ ensemble. The simulation cell contains overall 100000 molecules of the surfactant and water and the standard periodic boundary conditions [18] are applied. One common approach to study micelle formation is to perform MD simulation starting from a random configuration of fully dissolved surfactants in water. At a very low concentration, after the self-assembly of free surfactants into smaller aggregates, the number of collisions between these aggregates becomes limited. The micelle formation slows down with the height of nucleation barrier, which can be order of magnitude of $15k_BT$ [391] so that an equilibrium might not be reached. In this study, molecular simulations are performed by starting from a phase separated system, where all surfactant molecules are assembled into one aggregate and placed in a cubic box with the side length of ca. 15.0 nm. This approach of studying a micelle formation based on the instability of a large aggregate has already been implemented in the literature [392, 393]. The advantage is that for large aggregation numbers, the micelle shape fluctuations occur spontaneously and frequently, so that a neck-formation followed by a break down of an aggregate occurs in a short time. The fission pathway of $C_{10}E_4$ micelle is illustrated in Figure 6.3. The break down into separate micelles occurs after approximately 7 ns.

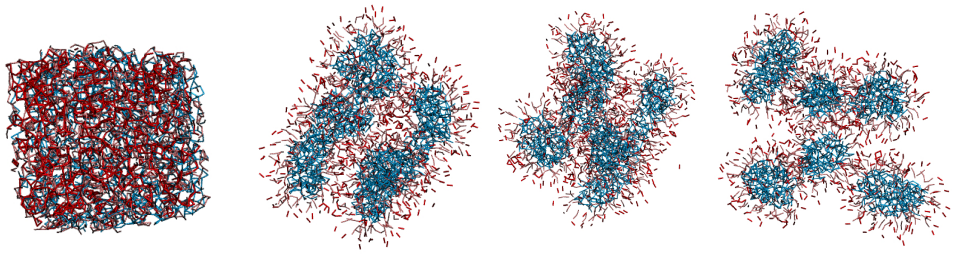


**Figure 6.3:** Simulated $C_{10}E_4$ micelle formation. The snapshots are (left to right, top to bottom) at the beginning of the process (0 ns), after 0.6 ns, after 1.7 ns, and after about 14 ns.

After the equilibrium is established, the formed clusters are sampled along their trajectories for further 25 ns and analysed by means of the cluster cutoff method. In this method,

two or more surfactant monomers are identified as belonging to the same aggregate if any of their respective alkyl beads are found within a cutoff distance to one another. The cluster cutoff distance is fixed at the minimum of the Mie potential of two alkyl beads ($\sigma\,(\frac{n}{m})^{1/(n-m)}$, where $n$=15, $m$=6). In order to characterise the aggregate morphologies, we compute a number of properties, such as the radius of gyration, moments of inertia, shape anisotropy, aggregation numbers, and the density profiles within an aggregate.

The $C_{10}E_4$ micelles are analysed at six different concentrations between $x_{C_{10}E_4} = 2.8\times10^{-3}$ to $1.7\times10^{-2}$ (Table 6.2). The shape of the micelles is obtained from the principal moments of inertia $I_i/I_j$ and the relative shape anisotropy $\kappa^2$, defined by

$$\kappa^2 = \frac{b^2 + \frac{3}{4}c^2}{R_g^4}, \tag{6.3}$$

where the squared radius of gyration $R_g^2$, asphericity $b$ and acylindricity $c$ are given as

$$R_g^2 = I_1^2 + I_2^2 + I_3^2 \tag{6.4}$$

$$b = I_1^2 - \frac{1}{2}(I_2^2 + I_3^2) \tag{6.5}$$

$$c = I_2^2 - I_3^2. \tag{6.6}$$

For spherical micelles, $I_1 \approx I_2 \approx I_3$ and hence $I_1/I_2 \approx 1$, $I_2/I_3 \approx 1$ and $\kappa^2 \approx 0$ , whereas for rod-like micelles $I_1 << I_2 \approx I_3$ and thus $I_1/I_2 \approx 0$ and $I_2/I_3 \approx 1$. As the micellar shape becomes more non-spherical, $\kappa^2$ deviates stronger from zero.

MD simulations reveal that at very low surfactant concentrations ($x_{C_{10}E_4} < 5.2 \times 10^{-3}$), monodisperse spherical micelles with an average aggregation number of about 87 surfactants are found in the solution. As the concentration increases, rod-like micelles with average aggregation number of 132 surfactants are formed, whereas spherical micelles are still present in the mixture, indicating a degree of polydispersity (Figure 6.4).

The relative shape anisotropy $\kappa^2$ as a function of the aggregation number is given in Figure 6.5. The minimal deviation from sphericity is found for the aggregation numbers between 40 and 100 surfactants. For aggregates of less than 20 surfactants, $\kappa^2$ is larger, indicating the rod-like shape. This result is consistent with findings by LeBard [394] for $C_7E_6$ surfactant, indicating that the so called pre-micelles exhibit an elongated shape

**Table 6.2:** Number of surfactant and water molecules with corresponding molar and weight fraction in the modelled systems.

| $N_{H_2O}$ | $N_{C_{10}E_4}$ | $x_{C_{10}E_4}$ / % | $w_{C_{10}E_4}$ / % |
|---|---|---|---|
| | Low concentrations | | |
| 97784 | 277 | $2.8 \times 10^{-3}$ | 0.05 |
| 99650 | 350 | $3.5 \times 10^{-3}$ | 0.06 |
| 96000 | 500 | $5.2 \times 10^{-3}$ | 0.09 |
| 94400 | 700 | $7.4 \times 10^{-3}$ | 0.12 |
| 92000 | 1000 | $1.1 \times 10^{-2}$ | 0.17 |
| 88000 | 1500 | $1.7 \times 10^{-2}$ | 0.24 |
| | High concentrations | | |
| 18071 | 1902 | 0.05 | 0.49 |
| 28491 | 1819 | 0.06 | 0.54 |
| 26863 | 2022 | 0.07 | 0.58 |
| 24690 | 2294 | 0.09 | 0.63 |
| 22785 | 2532 | 0.10 | 0.67 |
| 21102 | 2742 | 0.12 | 0.71 |
| 19503 | 2942 | 0.13 | 0.74 |
| 16671 | 2942 | 0.15 | 0.77 |
| 14364 | 2942 | 0.17 | 0.79 |
| 12542 | 2942 | 0.19 | 0.81 |
| 6865 | 2942 | 0.30 | 0.89 |
| 1471 | 2942 | 0.67 | 0.97 |

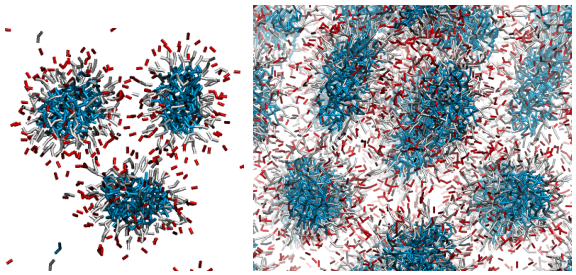


**Figure 6.4:** Snapshots of the spherical micelles present at concentration $x = 3.5 \times 10^{-3}$ with $< N_{agg} >= 87$ (left) and polydisperse micelles present at concentration $x = 1.1 \times 10^{-2}$ with $< N_{agg} >= 85$ for spherical micelle and rod-like micelle with $< N_{agg} >= 132$ (right).

and a less dense core. A steep increase in the shape anisotropy and hence the transition from the spherical to the rod-like micelles is observed in the range between 100 and 150

surfactants. A similar result was presented by Velinova *et al.* [149], reporting the sphere-to-rod transition in the range between 80 and 100 molecules for the $C_{12}E_5$ surfactant. For the aggregation number above 150, the value of $\kappa^2$ continues to increase, thereby implying a formation of the worm-like micelles. The sphere-to-rod transition is not an actual phase transition, since the rod-like aggregates are present in a high concentrated solution alongside with spherical micelles due to polydispersity. The radius of gyration is directly related to the aggregation number and shape of the micelle. As the relative shape anisotropy, it undergoes a sharp increase in the range between 100 and 150 surfactants (Figure 6.5 bottom).

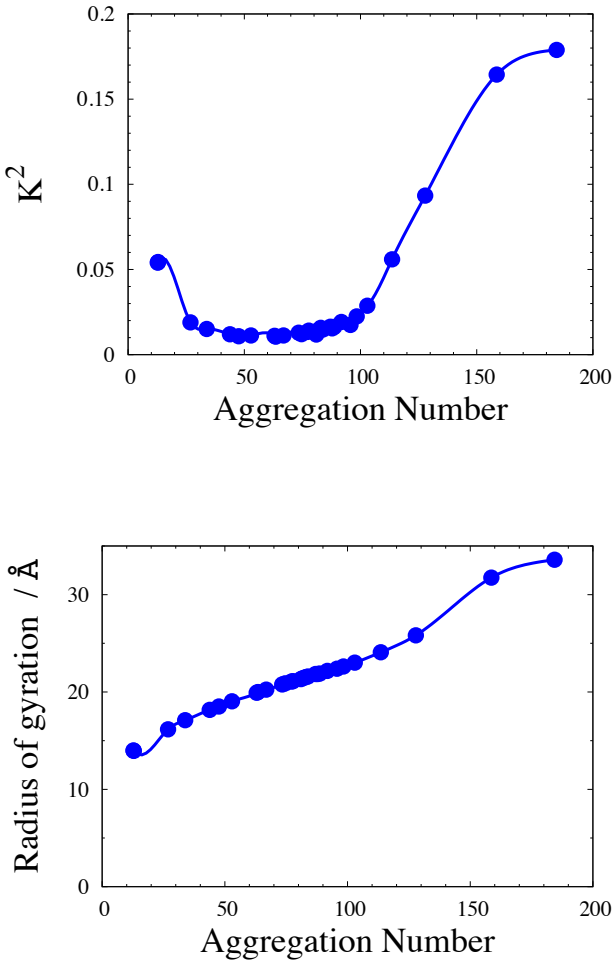



**Figure 6.5:** Relative shape anisotropy $\kappa^2$ and radius of gyration as functions of the aggregation number of the $C_{10}E_4$ micelles, summarised over different concentrations and averaged over time.

### Density profiles within the micelle

The interior of the micelle is characterised by the density profiles of the surfactant beads. A characteristic profile of a spherical $C_{10}E_4$ micelle at the concentration $x = 3.5 \times 10^{-3}$ is given in Figure 6.6. The profiles of the alkyl, ethoxy and the end groups are found to follow Gaussian distribution functions of the distance from the centre of mass of the corresponding micelle.

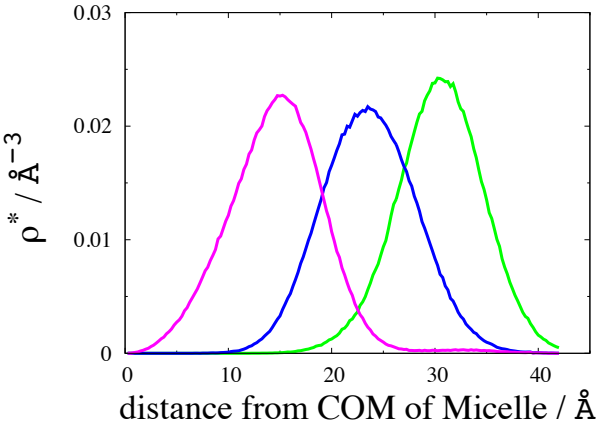



**Figure 6.6:** Normalised density profiles for *OA* (green line), *EO* (blue line) and *CM* (pink line) beads within a spherical $C_{10}E_4$ micelle.

## 6.4.2 Surfactant chain length effects

In this section, the simulation of aqueous solutions of surfactants with different alkyl chain length ($C_7E_4$, $C_{10}E_4$, $C_{13}E_4$) and ethoxy chain length ($C_{10}E_5$, $C_{10}E_6$, $C_{10}E_7$, $C_{10}E_8$) are conducted at three different concentrations above the experimental CMC ($x_{C_iE_j} = 3.5 \times 10^{-3}$, $5.2 \times 10^{-3}$ and $7.4 \times 10^{-3}$) for each of the systems.

### End-to-end distance

The end-to-end distance of a surfactant molecule within the micellar aggregate is shown in Figure 6.7. For all studied systems, the distance is found to be a Gaussian distribution, with the peak shifting to higher values for the longer surfactant chains. Since the alkyl bead and ethoxy bead are of a similar size, the increase in the alkyl chain length (Figure 6.7 top) and increase in the ethoxy chain length (Figure 6.7 bottom) has nearly the same

constant effect on end-to-end distance. With an increasing chain length, the distribution becomes broader, indicating a larger number of conformations of a longer molecule.

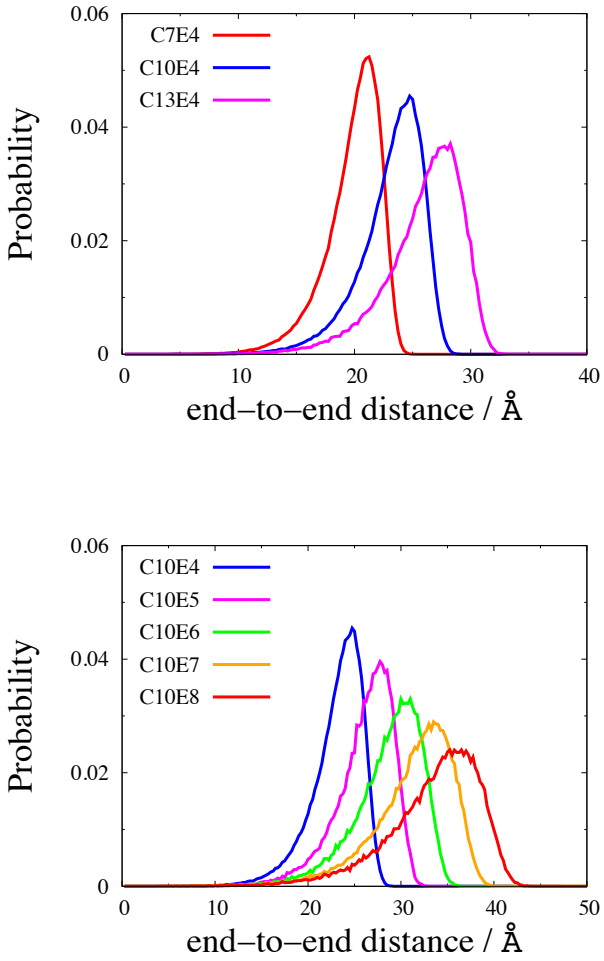



**Figure 6.7:** Probability distribution of end-to-end distances for $C_iE_j$ surfactants of different alkyl chain length, represented by $C_iE_4$ and of different ethoxy chain length, represented by $C_{10}E_j$ surfactants.

### Aggregation numbers and sizes

In order to evaluate the chain length effects on the average aggregation number, we compare the surfactants of different alkyl- and ethoxy chain length at a low concentration ($x_{C_iE_j} = 3.5 \times 10^{-3}$). The CG simulation results for the different surfactant architectures, which are given as weighted averages from the aggregate distribution functions together with the corresponding standard deviations, are illustrated in Figure 6.8. For the same

number of ethoxy groups, the increase in the alkyl chain length has a significant impact on the aggregation number. The smallest micellar aggregation numbers (between 25-53 molecules) are observed for the surfactant with the shortest alkyl chain, $C_7E_4$. For $C_{10}E_4$ system, the micelles shift to higher aggregation numbers, ranging between 75 to 96 molecules. The largest aggregates are found for $C_{13}E_4$ system with aggregation number between 165 and 182.

The change of the aggregation number with the ethoxy chain length is far less pronounced. Only a slight shift to lower aggregation numbers can be identified with increasing ethoxy chain length. This behaviour can be interpreted as follows: the core of the micelle is established through the hydrophobic tails, whereas the hydrophilic heads act as an extension of the chain, thereby having only a minor contribution on the aggregation size. The fact that the aggregation size increases significantly with the alkyl chain length and decreases slightly with the ethoxy chain length is well established through theoretical and experimental findings [13, 366]. For surfactants with 4 to 8 ethyl groups, most of spherical aggregates are found within the range from 62 to 98 molecules. Yet, some smaller aggregates, or "pre-micelles", seem to appear with increasing ethoxy chain length.

A direct comparison of the aggregation number with experimental data is difficult as very different results are reported in the literature [395–400] ascribed to different measured techniques and the data analysing methodologies. Experimental aggregation numbers at the room temperature are also displayed in Figure 6.8 in order to validate the MD simulation results. Identifying the trends of the experimental averaged aggregation number with changing chain length seems to be problematic due to the scatter in the data. Notwithstanding, the MD results seem to be within the reported experimental range. For $C_{10}E_6$ and $C_{10}E_8$ micelles, the calculated results match the experimental values very closely.

The information about the aggregation size is given through the radius of gyration, $R_g$ (Eq. 6.4). The radius of gyration is related to both, the aggregation number of a cluster and the length of the surfactant molecule (*i.e.* end-to-end distance, see Figure 6.7). Weighted averages of radii of gyration as a function of the alkyl and ethoxy chain length are presented in Figure 6.9. As expected, a steep increase in $R_g$ is observed for the increasing alkyl chain length, whereas the increase in ethoxy chain length has only a minor effect.

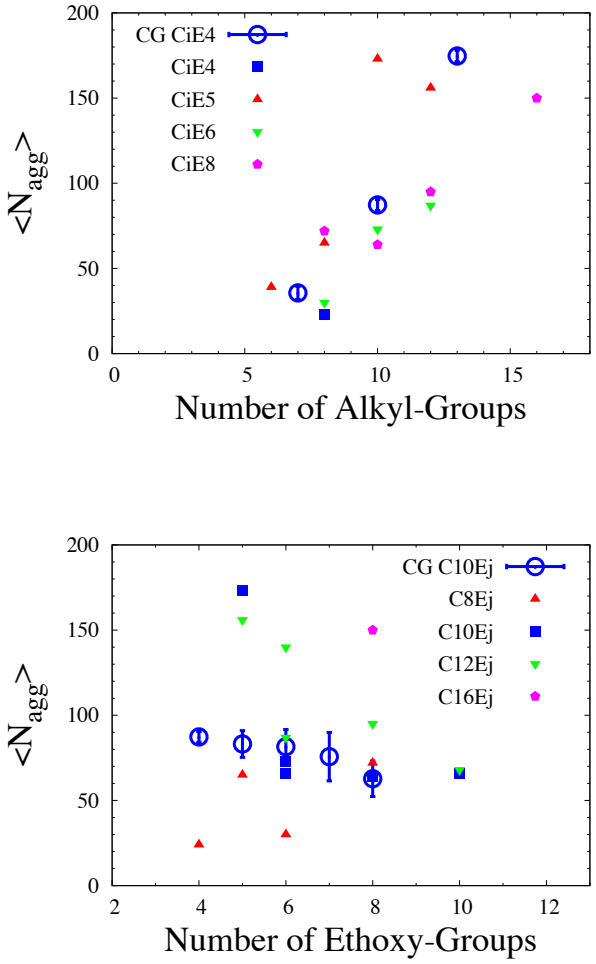



**Figure 6.8:** Average aggregation numbers as a function of the alkyl and ethoxy groups. Experimental data [395–400] is given as filled symbols; the CG simulation results are represented through open circles.

### Critical Micelle Concentration

The critical micelle concentration is one of the main descriptors of the surfactant. It is defined as a concentration above which single surfactant molecules self assemble into micellar clusters driven by hydrophobicity of the tails and hydrophilic nature of their heads [13]. The driving force consists of two competitive contributions from the negative free energy change from transferring the hydrophobic core in the spherical micelles and positive surface free energy. Therefore, CMC is directly related to the free energy of the system driving self-assembly, which determines the aggregation properties. In this manner,

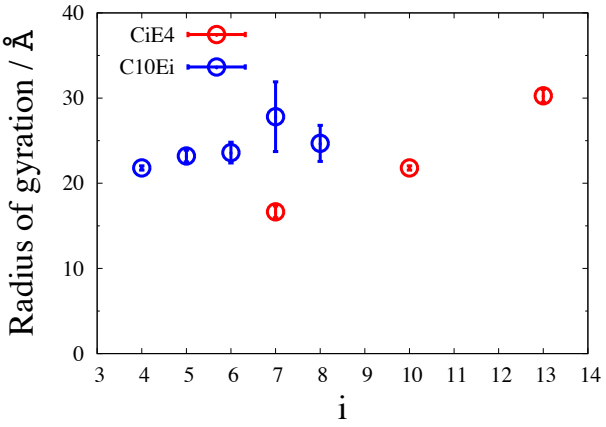



**Figure 6.9:** Radius of gyration as a function of alkyl and ethoxy chain length of various $C_iE_j$ calculated by MD simulation.

developing models that are able to reproduce quantitatively the structural properties of aggregates along with the CMC remains a challenge [401].

For the non-ionic surfactants, the CMC values are quite low. Hence using a direct method in the molecular simulation is not feasible as a a very large system would be required to predict the CMC. In this work, we compute CMC using an indirect method on the basis of the assumption that above the CMC, the number of free monomer molecules in the solution remains constant [13]. In this manner, the CMC analysis is performed in the liquid bulk phase and at the surfactant concentration well above the actual CMC value. The number of the free monomer molecules is counted at every time step and subsequently averaged over the entire simulation time. Free monomers are those whose alkyl beads are not within the cutoff distance with alkyl beads of any other surfactant molecules. The calculated CMC result is an average over all studied concentrations.

It is well established [366] that the increasing length of the alkyl chain and therefore the hydrophobic character of the surfactant causes an exponential decrease in CMC. On the contrary, the increase in the number of hydrophilic groups contributes to a higher solubility and therefore an increase in CMC. However, the effect is considerably smaller compared to the one caused by the increase of the hydrophobic tail [366]. The CMCs are calculated for different surfactant architectures and compared to the experimental data, which were summarised in Ref. [402]. Indeed, the largest number of free monomers, and

therefore the highest CMC, is observed for the system with the shortest alkyl chain, $C_7E_4$ (Figure 6.10). As the length of the alkyl chain increases, the CMC decreases exponentially, which is well reproduced by molecular simulations. No increase in CMC is observed with increasing ethoxy length (Figure 6.10 bottom), however the MD results are very close to the experimental findings. The trends are consistent with the trends for the aggregation numbers (Figure 6.8), indicating that a driving force of the system is captured correctly. Considering that the CMC results are true prediction of the model, the agreement with experimental data is remarkable.

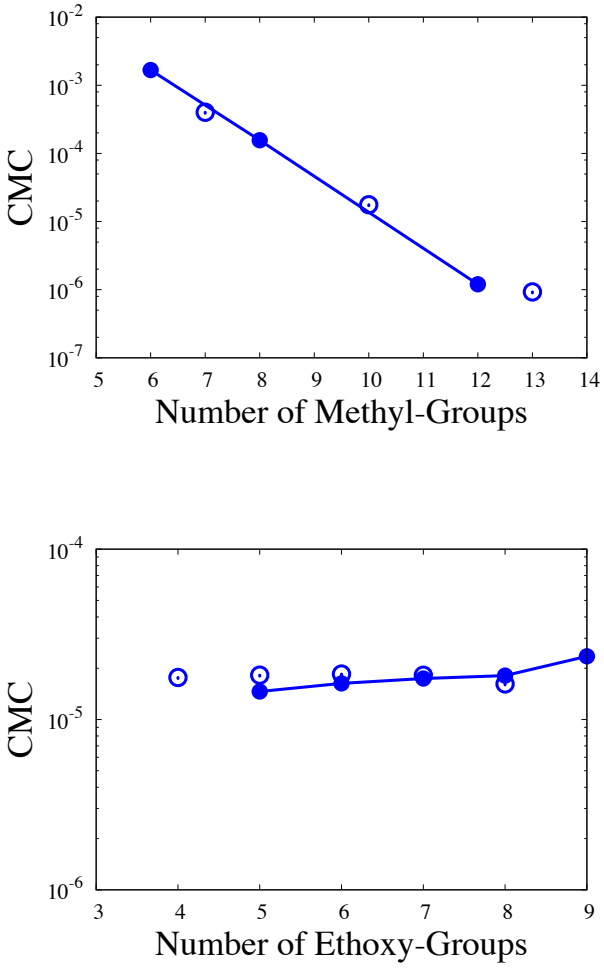



**Figure 6.10:** Critical micelle concentration as a function of the alkyl chain length within the $C_iE_4$ surfactant family (top), and as a function of the ethoxy chain length within $C_{10}E_i$ (bottom). Filled symbols denote the experimental data, summarised in Ref. [402] and open symbols are the MD simulation results. Line is a guide to the eye.

### 6.4.3 Equilibrium surface tension and surface excess

Close related to the CMC is a property known as an effectiveness of adsorption, $\Gamma_{max}$ [403], which is defined as a surface excess concentration at which the maximum adsorption is attained. Effectiveness is used as a measure of a maximum drop of the surface tension $\gamma_{min}$ which a surfactant can achieve. A great deal of experimental data is available in the literature indicating the importance of this property. Fortunately, it can also be accessed via molecular simulation, which might significantly facilitate the surfactant design by focusing on the most promising species before conducting experiments.

Due to the system size limitations mentioned above, the molecular simulation at the concentrations lower than $\Gamma_{max}$ using a direct method is not feasible. However, in the dilute regime, the concentration of the free surfactant monomers in the bulk solution is negligible as compared to the accumulation at the surface. Therefore, based on the Gibbs isotherm

$$\Gamma = -\frac{1}{RT}\frac{d\gamma}{d\ln x}, \tag{6.7}$$

the decrease in surface tension can be calculated as a function of the surface excess concentration, *i.e.* the focus is on the interface itself rather than the bulk phase. This approach is described in greater detail in Ref. [311]. As can be seen from Figure 6.11, the decrease in surface tension as a function of the surface excess as well as $\Gamma_{max}$ of the $C_{10}E_4$ surfactant system are in a good agreement with experimental data. This result is, again, a pure prediction of the model.

Another important characteristic of the surfactant is an area per molecule, which is the area used by a single surfactant molecule at the surface at the CMC. According to the experimental findings, the area per molecule depends nearly linearly on the length of the head group and is independent on the length of the tail group of the surfactants [404]. The experimental data reported in the literature for the $C_iE_4$ family, ranges from 0.42 to 0.54 nm$^2$ [404–406]. From the MD simulation run at $\Gamma_{max} = 3.32 \times 10^{-6}\text{mol} \times \text{m}^{-2}$, the area per molecule is equal 0.51 nm$^2$.

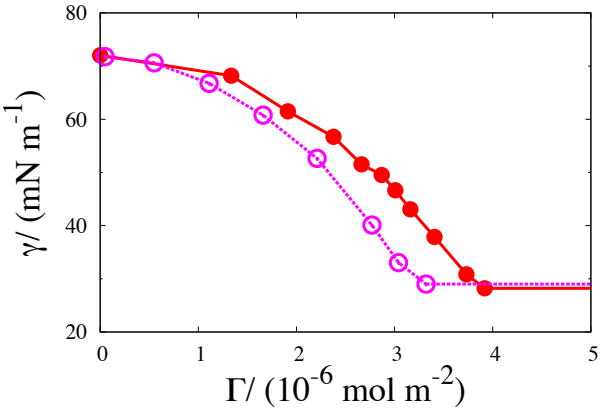



**Figure 6.11:** Surface tension as a function of the surface excess concentration for $C_{10}E_4$. The MD calculated results (open circles) are compared to the experimental data [407] (filled circles).

### 6.4.4 Micelle kinetics

#### Comparison of the micellisation theory with simulation results

According to Becker-Döring theory [408, 409], the growth of micelles occurs in terms of stepwise addition of a single molecule to the cluster of the same molecules, similar to a step-growth polymerisation. This simple kinetic model was used by Marrink [410] and Jorge [411] to validate their simulation results and to describe the growth of the micelles at the early stage of micelle formations. It was assumed that the self-assembly takes place in three stages: fast aggregation of monomers to small oligomers, collision of oligomers to form larger aggregates and a slower stage involving collisions between larger micelles. A simple kinetic model was used, where the rate constant was assumed to be size-independent and was estimated from the self-diffusion coefficient and the collision radius of the surfactants. This model was shown to correctly describe the first two stages of the micelle formation. The aggregation process, to the first order, is given by the following reaction:

$$[M_i] + [M_j] \xrightarrow{k_{i,j}^{Agg}} [M_{i+j}] \tag{6.8}$$

where $[M_i]$ is the concentration of the clusters with $i$ surfactant molecules. The aggregation rate is given as

$$\frac{d[M_i]}{dt} = \sum_{j=1}^{i/2} k_{i-j,j}[M_{i-j}][M_j] - [M_i]\sum_{j=1} k_{ij}[M_j] - k_{i,i}[M_i]^2 \tag{6.9}$$

The first term on the right hand side represents the formation of clusters of size $i$ due to the collision of two smaller clusters, while the second and the third term describe the disappearance of clusters of size $i$ due to the collision with other clusters. Assuming a size-independent rate constant, the differential equation 6.9 with the initial conditions $[M_1] = M_0$ and $[M_{i>1}] = 0$ at $t = 0$, where $M_0$ is the initial surfactant concentration, can be solved analytically:

$$[M_n] = M_0 \left(\frac{t}{\tau}\right)^{n-1} \left(1 + \frac{t}{\tau}\right)^{-n-1}, \tag{6.10}$$

where $\tau = \frac{2}{kM_0}$. For the monomers, equation 6.10 is given as

$$[M_1] = M_0 \left(1 + \frac{t}{\tau}\right)^{-2}. \tag{6.11}$$

The total concentration of clusters can be expressed as

$$[M_{tot}] = \sum_{i=1}^{\infty}[M_i] = M_0 \left(1 + \frac{t}{\tau}\right)^{-1}. \tag{6.12}$$

In the diffusion-controlled regime, the rate constant $k_{i,j}$ can be expressed as a function of the diffusion coefficient $D_i$ and the collision radius $R_i$

$$k_{i,j} = 4\pi(D_i + D_j)(R_i + R_j). \tag{6.13}$$

If a size-independent aggregation constant is assumed, Eq. 6.13 becomes:

$$k_{i,j} = 16\pi DR. \tag{6.14}$$

Therefore, the characteristic aggregation time $\tau$ can be expressed as

$$\tau = \frac{L^3}{8\pi DRN_0}, \tag{6.15}$$

where the initial concentration $M_0 = \frac{N_0}{L^3}$, with $N_0$ is the initial number of surfactant molecules and $L$ is the length of the box, respectively.

We use the described kinetic model to validate our CG simulation results. The Molecular Dynamics simulation is carried out in the canonical ensemble, where the number of particles $N$, temperature $T$ and volume $V$ are kept constant. In the cubic simulation box with the side length of 10.5 nm, 175 surfactant molecules are randomly distributed between 34825 water molecules. Figure 6.12 displays the evolution in time of the self-assembly process. The number of monomers is given by the blue line, whereas the total number of clusters is given in red. The total number of clusters takes into account all clusters, including monomers and oligomers. The decay in the number of clusters is fast on the early stage of the simulation and reaches a saturation level as simulation proceeds. After $5 \times 10^3$ ps, the number of monomers has essentially decreased to zero, appearing only statistically in correspondence with the CMC value. The total number of clusters has decayed to approximately 6 clusters. According to the previous simulations, where the micelle formation was observed starting from one large aggregate, the average aggregation number of clusters at the same concentration was estimated to approximately 70 surfactants per cluster, which corresponds to 2-3 clusters at given initial number of surfactant molecules. It is assumed, that the equilibrium has not been established yet and a further decrease in the numbers of clusters is to expect. Since the last stage of the aggregation process is the slowest one and is associated to the number of collisions between smaller micelles, the energy barrier is very high and a very long simulation time is required to reach equilibrium.

The comparison with the theoretical model requires an estimate of the characteristic aggregation time (Eq. 6.15) and thereby the diffusion coefficient $D$ and the collision radius $R$, with both values can be obtained directly from the MD simulation. In order to estimate the characteristic aggregation time, computation of the collision radius and the diffusion coefficient is required. The collision radius $R$ is approximated through the end-to-end distance of the surfactant, which is estimated as 2.4 nm (Figure 6.7). The average diffusion coefficient is estimated from the slope of the mean-square displacement as $24.02 \times 10^{-11} m^2 s^{-1}$. One has to consider that the diffusion coefficient of the given water model is overestimated by approximately factor 3 (7.4 instead of the experimental value $2.3 \times 10^{-9} m^2 s^{-1}$ at T=298 K, see Table 4.4). Assuming that the diffusion in a diluted surfactant solution is dominated by the diffusion of water, the first approximation is to use the same factor to scale the diffusion coefficients of the coarse-grained surfactant models

to the "real" surfactant molecules. In this manner, the diffusion coefficient of $C_{10}E_4$ is reduced to $7.47 \times 10^{-11} m^2 s^{-1}$. Note that this is a crude approximation, averaging over the fast surfactant monomers at the early stage of micelle formation and the slower surfactants as a part of micelles at the later stage of the micelle formation.

The computed collision radius and the average diffusion coefficient yield the characteristic aggregation time $\tau = 146\ ps$. This value can be used to compare the simulation results with the kinetic model, which is visualised in Figure 6.12. Despite its simplicity, the kinetic model appears to be able to describe the aggregation process for approximately the first 1000 ps of simulation time. A good agreement between the kinetic model and simulation results suggests that at the early stages, the self-aggregation takes place in the diffusion-controlled regime and is driven by the collision of the monomers and small oligomers.

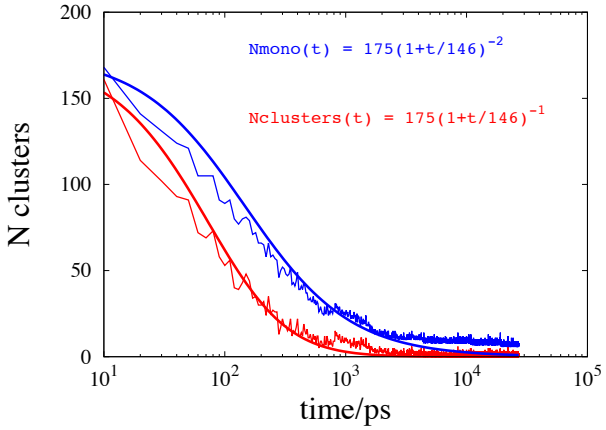



**Figure 6.12:** Evolution in time of the self-assembly of $C_{10}E_4$ surfactants. Comparison between simulation results (thin lines) and kinetic theory predictions (smooth thick lines). Blue lines denote the number of monomers, whereas red lines denote the total number of clusters in the simulation box.

### Mobility rate as a function of the surfactant chain length

The rate constant can be estimated in an alternative way, in correspondence to the surfactant mobility, analog to the "Hamming distance" approach. The mobility can be defined as a frequency of how often a molecule changes cluster, averaged over all molecules during the entire simulation time. Since this value is not absolute it is difficult to compare

to the experimental values. But the "mobility" can be compared within the surfactant family, depending on the chain length of the surfactants. The computation is conducted in the equilibrated micellar solution and is presented in Figure 6.13. In agreement with the results obtained from the cluster analyses (Section 6.4.2), the clusters of the smallest size and the largest CMC value exhibit the highest mobility, as it is the case for $C_7E_4$. With increasing alkyl chain length, the exponential decrease in in the mobility has been observed, as it was also the case for the CMC (c.f. Figure 6.10). The impact of the ethoxy group on the mobility is minor, in line with the observation of the aggregation numbers and CMC.

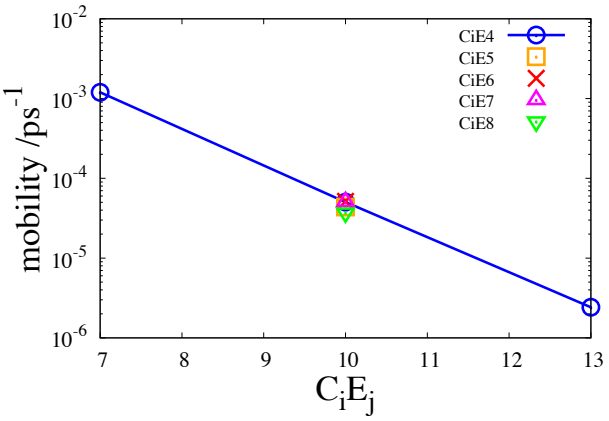



**Figure 6.13:** Mobility (rate constants) in the micellar solution of $C_{10}E_4$ in dependence of the chain length.

### 6.4.5 Isotropic-lamellar phase transition

One of the main challenges for the CG model is an ability to accurately reproduce the phase boundaries and the mesostructures of amphiphiles. After having studied the low concentration limit, we should now test our CG model across a wider range of concentrations. For this purpose, we use a simulation cell containing 43039 particles. A random initial configuration is generated in the orthorhombic box of size 13 nm. Semiisotropic pressure coupling is applied, where the $x$ and $y$ direction are scaled isotropically and $z$ direction is scaled independently. At the surfactant concentration $w_{C_{10}E_4} = 0.74$, a lamellar phase is formed after 200 ns. The pathway of the bilayer formation is illustrated in Figure

6.14. It appears that the bilayer is built through the elongation of the rod-like aggregates.

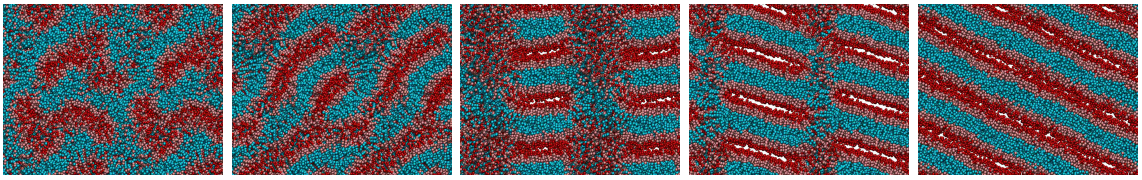


**Figure 6.14:** Simulated bilayer formation at the surfactant concentration $C_{10}E_4$= 0.74. The snapshots are at taken at 1 ns, 20 ns, 50 ns, 150 ns, and 190 ns (from left to right)

Molecular trajectories are sampled in the subsequent 15 ns equilibration run. The stability of the lamellar phase is studied in the concentration range $w_{C_{10}E_4} = 0.49 - 0.97$ (Table 6.2). For this, at the concentrations between 0.49 and 0.74, the total number of particles is kept constant, and the concentration is altered by substituting surfactant through the water particles. For concentrations between 0.74 and 0.97, the number of surfactant molecules is kept constant, and the number of water molecules is reduced according to the concentration studied.

The lamellar phase can be characterised by the second rank orientational order parameter $S_2$, calculated from the Saupe order tensor [412]

$$Q = \frac{1}{2}(3\langle \hat{u}_i \hat{u}_i \rangle - 1), \tag{6.16}$$

where $\hat{u}_i$ is the unit orientation vector of the $i$th surfactant molecule. The order parameter $S_2$ is the largest eigenvalue of the matrix, the corresponding eigenvector $\hat{n}$ is called the director, as it points in the average direction of the molecules. Order parameter is measured as the width of the distribution of $\hat{u}_i$ around $\hat{n}$, and ranges from zero for randomly oriented to one for perfectly aligned molecules. The unit orientation vector $\hat{u}_i$ of a flexible surfactant molecule can be calculated as a director over the vectors which connect each surfactant bead with the centre of mass of the surfactant.

Typical values for the order parameter for the amphiphilic bilayer computed via molecular simulation are in the order of 0.3 to 0.5 [141, 413]. The order parameter depicts the ranges of the lamellar phase enabling the comparison to the experimental phase diagram [414]. Figure 6.15 represents the second-rank orientational order parameter $S_2$ as a function of surfactant concentration in weight per cent. The formation of the ordered lamellar phase is characterised by the high order parameter, which is observed from the molecular

simulations in the concentration range between 0.74 and 0.81, in a reasonable agreement with experimental concentrations between 0.6 and 0.83. The transition from isotropic to nematic phase is found in the range between $w_{C_{10}E_4}$ =0.58 and 0.74 and from nematic to isotropic phase between $w_{C_{10}E_4}$ =0.81 and 0.88. The observed continuous increase in the order parameter may be interpreted as a second order phase transition. However, $^2$H NMR measurements indicated the phase coexistence between the isotropic and the lamellar phases [414]. MD simulations might not capture this coexistence due to the finite size effects.

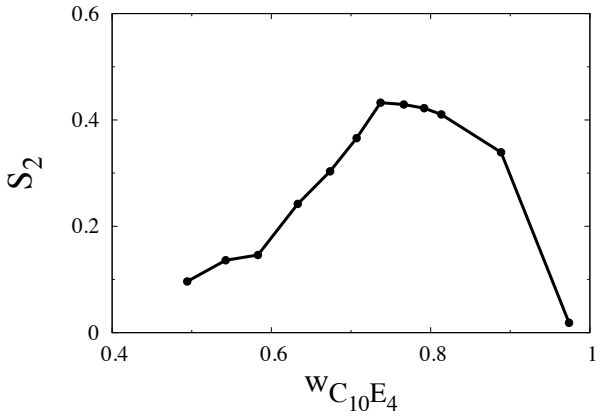



**Figure 6.15:** Orientational order parameter $S_2$ as a function of concentration in an aqueous $C_{10}E_4$ system. The increase in the order parameter suggests that the self-assembly into the ordered phase.

The arrangement of molecules in layers, *i.e.* layer thickness and spacing, can be characterised by means of the distribution function parallel to the director which is defined as

$$g_{||}(r_{||}) = \frac{1}{N}\langle \sum_i \sum_{j \neq i} \delta(r_{||} - \hat{n} r_{ij}) \rangle \tag{6.17}$$

In Figure 6.16, the periodicity of $g_{||}(r_{||})$ is shown. The bilayer thickness, calculated at the concentration $w_{C_{10}E_4} = 0.74$ as a distance between two minima in the distribution of the alkyl groups, to 46 Å, which is in a good agreement with experimental result 50±5Å [415]. Figure 6.17 illustrates the experimental phase diagram for $C_{10}E_4$ [414]. Snapshots from MD simulation at $w_{C_{10}E_4} = 0.06, 0.17, 0.74, 0.97$ indicate the spherical micelles and the rod-

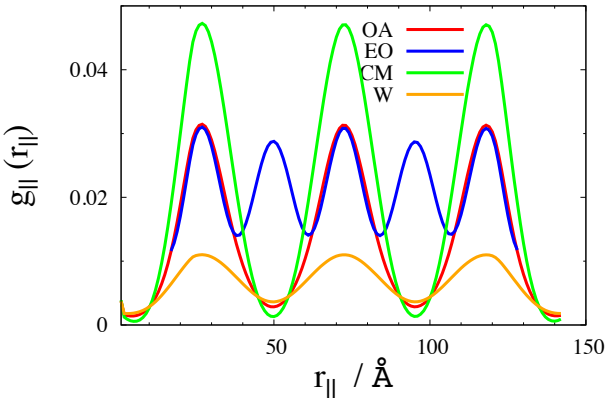



**Figure 6.16:** Beads distribution along the director $g_{||}(r_{||})$ for $C_{10}E_4$-water system at $w_{C_{10}E_4} = 0.74$. The red line denotes the distribution of head group beads, the blue line the ethoxy beads, the green line the alkyl beads and the yellow line water.

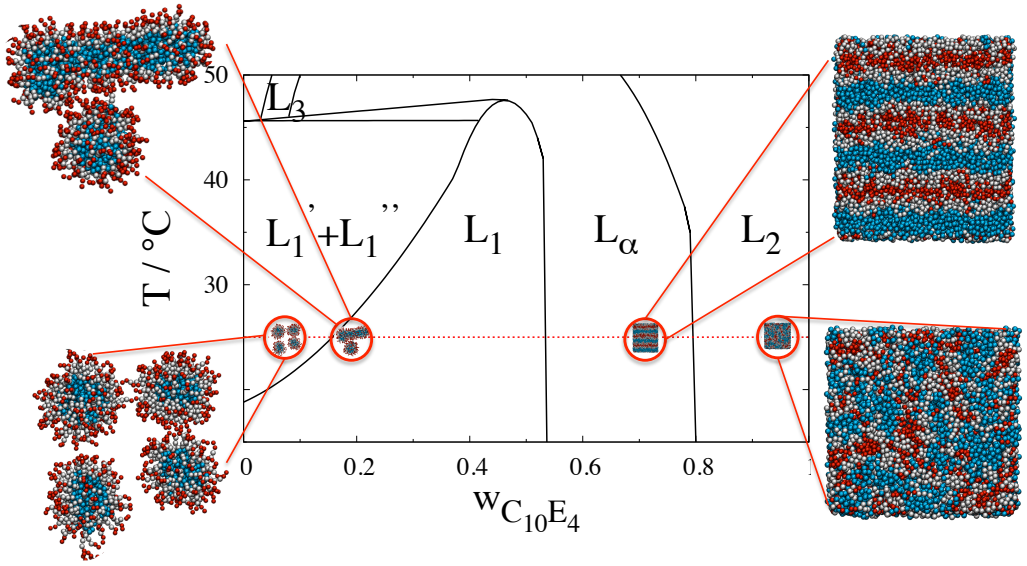



**Figure 6.17:** Experimental phase diagram for $C_{10}E_4$ [414]. Snapshots from MD simulation at $w_{C_{10}E_4} = 0.05, 0.17, 0.74, 0.97$ indicate the spherical micelles and rod-like micelles in the micellar phase $L_1$, lamellar phase $L_\alpha$, and isotropic surfactant rich phase $L_2$.

like micelles in the micellar phase $L_1$, lamellar phase $L_\alpha$, and isotropic phase of inversed micelles $L_2$, in agreement with experimental observations [414]. Hence, our CG force field is able to reproduce key structural properties and to capture the phase boundaries correctly.

## 6.5 Conclusions

In this chapter, we have presented the coarse-grained force field for non-ionic surfactant solutions using the SAFT-$\gamma$ Mie approach.

The force field parameters for the CG surfactant model was developed in Chapter 3. Water is represented by Model 2 (Section 4.4.4), which can accurately reproduce the liquid density and the surface tension. The unlike interactions between water and alkyl beads are adjusted to the interfacial tension of the alkane-water mixture, whereas the interactions between water and polar surfactant groups are parameterised to reproduce the excess enthalpy of mixture.

The derived CG force field is used as a direct input in molecular simulation to study the complex phase morhoplogies along with the interfacial properties of the surfactant mixture. The force field is tested at different concentrations of the $C_{10}E_4$-water system, revealing a formation of the spherical micelles with the aggregation number around 87 surfactants at a low concentrations. The transition from the spherical to the rod-like micelles occurs in the range between 100 and 150 surfactants per aggregate. As the concentration increases, both, the spherical and the rod-like micelles are found in the solution, indicating polydisperisty. The transferability of the force field is verified by analysing the aggregation properties of surfactants of different architectures. The aggregation number is found to increase significantly with the length of the alkyl chain, whereas the length of the ethoxy chain causes a minor decrease in the aggregation number. The critical micelle concentration is estimated using the indirect method, where the number of free monomers is calculated at the concentration above the actual CMC value. The CG force field is able to reproduce an exponential decrease in the CMC with the increasing alkyl chain length and a nearly constant values of CMC with changing ethoxy chain length, in a good agreement with experimental data. The robustness of the model is additionally confirmed through accurate predictions of the surface tensions, surface excess properties, and the area per molecule. At high surfactant concentrations, bilayer is formed, displaying the bilayer thickness of 46 Å, which is within the range of the experimental error, 50±5Å [415]. Furthermore, the ranges of the phase boundaries are reproduced close to the experimental findings. In this manner, our force field is able to predict the properties that were not considered in the parameterisation procedure and is therefore robust and transferable.

# Chapter 7

# Concluding remarks

A systematic coarse-graining methodology was employed throughout the work described in this thesis to derive accurate molecular force fields for the complex fluids. The effective intermolecular interaction parameters were obtained within a top-down framework, by considering the optimal description of a large number of macroscopic thermodynamic states. An efficient parameterisation of the intermolecular model was achieved by using a rigorous analytical equation of state, SAFT-$\gamma$, based on the Mie intermolecular potential. Due to a flexible and versatile form of the Mie potential, an accurate representation of the target physical properties could be achieved for the systems considered. The force fields derived with the aid of the equation of state were used as a direct input in Molecular Dynamics simulation in order to investigate the structural, interfacial, and dynamical properties of fluids.

The studied systems embrace the chain molecules, such as $n-$alkanes, $n-$perfluoroalkanes, $n-$ethers, semi-fluorinated alkanes, and non-ionic surfactants. It was found that including an intramolecular potential in the CG models significantly improved the representation of the thermophysical properties. The vapour-liquid equilibrium behaviour including the saturation density and the vapour pressure were utilised in the parameter estimation procedure, and a quantitative agreement was obtained between the molecular simulation results and the experimental data for the target properties. It is particularly gratifying that the predicted vapour-liquid interfacial tension could also be accurately reproduced. The development of an accurate CG model of water represented a particular challenge, since water is well-known for its complex and anomalous behaviour due to the pronounced

dipolar and the hydrogen-bonding interactions that characterise the system. Using a simple isotropic single-site CG model based on the spherically-symmetric potential leads to a number of the transferability and representability issues. It has been shown that an isotropic potential cannot capture all of the properties of water simultaneously. Two different CG water models based on the one-to-one mapping were developed and implemented depending on the systems and properties of interest: for the first, the target properties were the fluid-phase equilibria, namely the vapour pressure and saturated liquid density; for the second the target property were the vapour-liquid interfacial properties, namely the surface tension and liquid density. A two-to-one CG model was also developed at a higher level of coarse-graining for use in large scale molecular simulation.

The models of the pure components were then employed to study properties of mixtures. In the majority cases, one adjustable parameter for the energetic interactions of the various beads of different type was sufficient to describe the unlike interactions between components. The studied systems included the binary mixtures of carbon dioxide with water, water with $n-$alkanes, $n-$alkanes with carbon dioxide, and the ternary mixture comprising all three components. A good overall description of the vapour-liquid, liquid-liquid, and supercritical behaviour could be obtained for the binary mixture of carbon dioxide with water. The phenomenon of the barotropic density inversion was reproduced in a reasonable agreement with experimental data. A comparison with the more detailed atomistic force fields revealed an equally good and frequently even better performance of our simple CG models. In the mixture with extreme immiscibility, such as $n-$alkane and water, an accurate description of the liquid-liquid could be achieved. The triphase line and the azeotropic points were also found in a very good agreement with experimental results. A transition from type II to type III behaviour in the classification of Scott and van Konynenburg was observed for the binary mixture of carbon dioxide with $n-$alkanes of increasing chain length, in line with the experimental findings. Finally, the equation of state and the molecular simulation could accurately predict the behaviour of the ternary mixture, described only by the binary interaction parameters.

A key goal of the research presented in this thesis was, however, dedicated to the modelling of the aqueous alkyl polyoxyethylene glycol non-ionic surfactant mixtures. The phase boundaries between the various fluid and liquid crystalline state and the aggregation properties of these aqueous surfactants are very sensitive to the delicate interplay of

the intermolecular interactions between the surfactants and the solvent so that obtaining the appropriate balance presented a significant challenge. Our simple CG models of aqueous surfactants provided one with the capability to reproduce the complex lyotropic phase behaviour exhibited by the system at ambient conditions. A transition from spherical to rod-like micelles was observed with increasing surfactant concentrations. By studying the aggregation properties of surfactants with different architectures, a strong dependence on the length of the alkyl chain and a weak dependence on the length of the ethoxy chain was detected, in accordance with experimental measurements. The critical micelle concentration of the non-ionic surfactant solutions is commonly quite low, hence an indirect method was applied in order to estimate the CMC and its dependence on the surfactant chain length, thereby providing a quantitative agreement with experimental data. An accurate predictions of the area per molecule at the interface, surface excess properties, and bilayer thickness verified the robustness and transferability of the developed force fields.

## 7.1 Future work and potential applications

Despite the encouraging results presented in this thesis there is certainly room for further improvement. In particular, a development of a more robust and rigorous CG model for water is still highly desirable. One route would be to develop a model that directly accounts for association. This could be achieved by incorporating several association sites, as it is commonly implemented with the water model developed within the SAFT framework based on the SW potential [416]. However in the case of Molecular Dynamics simulations, a continuous potential is a near-essential requirement (as discontinuous forces lead to severe complications in the solutions of the equations of motion), hence appropriately deep association sites based on a continuous Mie potential would be an obvious choice. The computational demand of models with explicit association sites is expected to be significantly higher compared to the spherically symmetric models developed in this thesis and will therefore be less efficient for the large-scale simulations. An alternative treatment of the directionality of associating interactions would be the consideration of two-site CG model, formed from beads characterised by very different energetic / repulsive interaction. Such a model could be used to account for the dipolar-like directional interactions present though this crude geometric representation of the association may still be too simplistic

in order to describe the structure of real tetrahedral water.

In case of the aqueous surfactant mixtures, the intermolecular model parameters between the surfactants and water molecules could be further improved by a careful re-assessment of the target properties. For instance, the current parameter set for the alkyl group-water interactions is developed by considering the interfacial tension as the target property. An important driving force of self-assembly in such systems is, however, the solubility of hydrocarbons in water. Obtaining an accurate estimate is, however, challenging, since the solubility is quite low, e.g. corresponding to a mole fraction for $n-$hexane of $10^{-6}$. In order to study the solubility of hydrocarbons directly with conventional MD simulations one would require very large systems and long times. An alternative technique could be applied, such as for example free energy integration or Widom insertion. An accurate knowledge of the solubility of hydrocarbons would provide further insight into the mechanism of micelle formation. Additionally, this would help one to develop a more transferable parameter set applicable for different conditions.

In our current work, the phase diagram of the surfactant mixtures was studied as a function of different concentrations at ambient conditions. It is well-known that the temperature strongly affects the phase morphologies in such systems [417]. Unfortunately, preliminary MD simulations at different temperature have indicated that our current set developed to describe the system at room temperature do not allow one to reproduce the experimentally observed phase morphologies at other temperatures. Due to the complex interactions of surfactants with water and a delicate interplay between energetic and entropic effects, the intermolecular parameters cannot be used in a transferable manner and have to be re-derived at different temperatures. This is perhaps not that surprising as the effect of the dominant hydrogen-bonding interactions between the water and surfactant head groups is known to be highly temperature dependent. As the temperature increases, the hydrogen bonding interactions tend to become less prevalent due to the breaking of the bonds, so that the micelles tend to minimise their contact with water by building larger aggregates. Entropic effects become more pronounced at higher temperatures, so that at higher surfactant concentrations, the bilayer is destabilised until its ultimate disappearance at $\sim$ 330 K [415]. In order to accurately capture this behaviour with our effective potentials, the interaction between the ethoxy groups and water needs to be made weaker and that between the alkyl groups and water stronger compared to the corresponding

values at ambient conditions. The new parameter set could then be tested against the phase morphologies, aggregation properties and the critical micelle concentrations, which undergo a minimum at $\sim$ 320 K, according to the experimental measurements [418]. By having a reliable parameter set which accounts for the temperature dependent behaviour of the system, one would be able to predict the global phase diagrams for the entire family of the non-ionic surfactants.

Our newly developed SAFT-$\gamma$ Mie CG force fields can be used for a variety of different applications. The current model of surfactants can be extended to study the molecules with different architectures. For instance, branched, or so-called T-shaped surfactants could be explored, in order to investigate the effect of a different molecular geometry on the phase morphologies, adsorption, and wetting behaviour. Experimentally [419], a T-shaped topology is found to favour the formation of bilayers and to have a critical micelle concentration which is up to two orders of magnitude lower than the CMC of their linear counterparts. A question arises whether the parameters derived for linear molecules can be used to model the corresponding branched components, or whether a new refined force field needs to be developed in order to reproduce the correct behaviour of isomers. Similarly, gemini surfactants, which are composed of two monomeric surfactants connected by a spacer, are widespread in nature and have attracted substantial attention due to their unique behaviour [420]. The properties differ from the linear surfactants owing to their high surface activity, lower CMC, increased viscosity, and shear-thickening behaviour. The aggregation and interfacial properties of gemini surfactants strongly depend on the nature and length of the spacer [421]. In this manner, the CG models could be implemented to provide guidance in the screening process and to suggest the most suitable candidates for the applications of interest.

The modelling of charged systems, such as the ionic surfactants, gives rise to a particular challenge. The explicit treatment of the Coulombic interactions significantly increases the computational demand, thereby counteracting the main advantage of coarse-graining techniques. Truncating or effectively averaging of the Coulombic interactions is usually not satisfactory [422]. Ongoing research in the MSE group on the ionic surfactant systems aims to provide a suitable compromise for the treatment of the Coulombic interactions.

The modelling of the surfactant mixtures can be extended to the many-component systems, for instance by introducing $n-$alkanes of the intermediate chain length to study

the phase behaviour of microemulsions. Microemulsions exhibit a rich lyotropic phase behaviour, depending on the temperature and overall concentration. The experimental data for microemulsions is usually presented as ternary phase diagram at given temperature and pressure, or alternatively, as a pseudo-binary phase diagram (usually referred to as a fish diagram) at a given pressure and a fixed ratio of water and the $n-$alkane. The microemulsions are commonly classified with the nomenclature of Winsor [423], thereby distinguishing between oil-in-water (o/w) ($\underline{2}\phi$) microemulsions, water-in-oil (w/o) ($\bar{2}\phi$) microemulsions, bicontinuous middle phase microemulsions ($3\phi$), and the microemulsions which are not in equilibrium with either oil or water ($1\phi$). The phase boundaries are very sensitive to the chain length of the surfactants and $n-$alkanes. The force fields described and developed in this thesis could be used directly to study the structural properties of the microemulsions for surfactants and alkanes of different chain length at various concentrations and ambient conditions. Based on the description of the corresponding binary systems, we are hopeful that this would provide a reasonable agreement with the experimental data for the microemulsion systems, since the interaction between the surfactant and water have been shown to be accurate and robust (Chapter 6), and the interactions between the surfactant head groups and the alkanes can be treated as the interactions with the alkyl groups of the surfactant. Extending the modelling of a two-component to a three-component system by implementing the binary interactions has already been shown to provide an accurate description for the $n-$alkane-water-carbon dioxide system (Chapter 5). However, in order to reproduce the fish diagram with respect to the temperature and to capture the transition from $\underline{2}\phi$ over $3\phi$ to $\bar{2}\phi$ phases at low surfactant concentrations, further adjustment of the parameters may be required due to the aforementioned issue about the temperature dependence of the interactions for systems with prevalent hydrogen bonding. Systems comprising mixed surfactants can already be explored directly with our current force fields in terms of the effect of combining systems with different surfactant chain length on the phase morphologies. The modelling of the mixed surfactant systems can serve as a guide for the experimental formulation of mixtures with tailored properties. In addition to structural properties and phase morphologies, the transport properties such as self and mutual diffusion and viscosity could be studied in greater depth. Equilibrium methods, such as the Green-Kubo integration or the Einstein relation, or non-equilibrium methods, which measure the response of the system to external perturbations, could both

be applied to examine the transport properties. Due to the integrating out the fast degrees of freedom with the coarse-graining technique leads to an over-estimated diffusion coefficient of Model 2 for water (Chapter 4), and hence is also expected in the aqueous surfactant mixtures. A scaling factor can be introduced in order to account for this discrepancy. Qualitatively, the diffusion is expected to decrease and viscosity to increase with increasing surfactant concentration. This change in the transport properties is attributed to the different degrees of molecular aggregation and to the various phase transitions.

Coarse-grained models of surfactants are particularly useful in study of the wetting properties of aqueous droplets on the solid hydrophobic substrates, and of the spreading mechanism, which since its discovery in 1960s still remains not fully understood [424]. A molecular understanding of wetting and superspreading phenomena finds numerous applications in the development of effective herbicides, active pharmaceutical ingredients for respiratory diseases, and coating materials. The shape, the polarity, the concentrations of surfactant and the nature of the surface strongly affect the degree of wettability; understanding the mechanism and the dynamics of spreading at the molecular level by using CG models, even in a qualitative manner, could help to understand how to control the wetting properties and to select the most suitable substrate-surfactant combinations.

A coarse-graining framework can be applied to investigate new systems of scientific or industrial interest. For instance, the asphaltenes, which are found in crude oil, cause severe production problems, such as fouling in the reservoir and production pipe-line. This behaviour is attributed to the formation of highly viscous agglomerates upon the precipitation of asphaltene. Since the composition of crude oil cannot be characterised explicitly, coarse-grained models are well suitable for this type of applications. By grouping molecules of a similar molecular weight and chemical composition into the simplified coarse-grained pseudo-components, one can mimic the average behaviour of the complex polydisperse mixtures, thereby significantly reducing the complexity of modelling. The models can provide an insight into the global phase diagrams, thereby identifying the precipitation boundaries of asphaltene containing crudes and could also be used to study the liquid crystalline ordering of asphaltene and its effect on the viscosity of the agglomerates. In a similar way, biofuel mixtures could be studied by introducing the pseudo-component CG models. Hence, the ratios between the biofuels and the conventional diesel could be optimised to obtain a right balance between the maximum concentration of biofuel and a

reasonably low viscosity.

This list of possible future applications of the coarse-graining methodology presented in this thesis is far from being exhaustive but gives a flavour of the great potential of the approach.

# Appendices

# Appendix A

# SAFT-$\gamma$ Mie equation of state

In this section, the key relations of the SAFT-$\gamma$ Mie equation of state are shortly summarised. More details are given in the original SAFT-$\gamma$ Mie papers [22, 24, 55]. For the applications in Molecular Dynamics simulations presented in this work, an abridged version of the original SAFT-$\gamma$ Mie equation of state is used, where the conventional association term is not included in the Helmholtz free energy description. The elongated molecules are treated as chains formed from tangentially bonded segments. The Helmholtz free energy is given as a perturbation expansion of an ideal, monomer, and chain term:

$$\frac{A}{Nk_BT} = \frac{A^{Ideal}}{Nk_BT} + \frac{A^{Mono}}{Nk_BT} + \frac{A^{Chain}}{Nk_BT}, \tag{A.1}$$

which are discussed in the next sections.

## A.1 Ideal contribution

The ideal-gas contribution incorporates the contributions to the kinetic energy (including translational, rotational, and vibrational modes of motion) and is given as

$$\frac{A^{Ideal}}{Nk_BT} = \left( \sum_{i=1}^{N_C} x_i \ln(\rho_i \Lambda_i^3) \right) - 1, \tag{A.2}$$

where $x_i$ is the mole fraction of component $i$, $\rho_i = N_i/V$ is the number density, $N_i$ is the number of component $i$ in the mixture, and $\Lambda_i^3$ is the de Broglie thermal volume. The summation is performed for the number of components $N_C$ in the mixture.

## A.2 Monomer contribution

The interactions between groups $k$ and $l$ are represented by the Mie (generalised Lennard-Jones) potential

$$u_{kl}(r_{kl}) = \epsilon_{kl}\mathcal{C}_{kl}\left[\left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{r,kl}} - \left(\frac{\sigma_{kl}}{r_{kl}}\right)^{\lambda_{a,kl}}\right], \tag{A.3}$$

where

$$\mathcal{C}_{kl}(\lambda_{a,kl}, \lambda_{r,kl}) = \left(\frac{\lambda_{r,kl}}{\lambda_{r,kl} - \lambda_{a,kl}}\right)\left(\frac{\lambda_{r,kl}}{\lambda_{a,kl}}\right)^{\frac{\lambda_{a,kl}}{\lambda_{r,kl}-\lambda_{a,kl}}}.$$

The monomer term is obtained from a high-temperature perturbation expansion of the Helmholtz free energy up to the third-order [24, 55]:

$$\frac{A^{Mono}}{Nk_BT} = \frac{A_0}{Nk_BT} + \frac{A_1}{Nk_BT} + \frac{A_2}{Nk_BT} + \frac{A_3}{Nk_BT}. \tag{A.4}$$

Each term of the free energy contribution per molecule can be obtained as

$$\frac{A_q}{Nk_BT} = \left(\frac{1}{k_BT}\right)^q \left(\sum_{i=1}^{N_C} x_i \sum_{k=1}^{N_G} \nu_{ki}\nu_k^*\right) a_q \tag{A.5}$$

for $q = 0, 1, 2, 3$. The summation is performed over all components $i$ and all groups $k$. Here, $N_G$ is the number of groups present, $\nu_{ki}$ is the number of occurrences of a group of type $k$ of component $i$, and $\nu_k^*$ is the number of identical spherical segments forming the group $k$. The perturbation terms $a_q$ are obtained by summing the pairwise contributions $a_{q,kl}$ between groups $k$ and $l$ over all groups:

$$a_q = \sum_{k=1}^{N_G}\sum_{l=1}^{N_G} x_{s,k}x_{s,l}a_{q,kl}, \tag{A.6}$$

where $x_{s,k}$ and $x_{s,l}$ are the overall fractions of segments of groups $k$ and $l$ in the mixture, given as

$$x_{s,k} = \frac{\sum_{i=1}^{N_C} x_i\nu_{ki}\nu_k^*}{\sum_{i=1}^{N_C} x_i \sum_{l=1}^{N_G} \nu_{li}\nu_l^*} \tag{A.7}$$

Each free energy contribution $a_q$ can now be described in more detail.

### A.2.1 Repulsive reference fluid

The repulsive reference fluid $a_0$ is described in terms of the Helmholtz free energy of a hard-sphere fluid $a^{HS}$ with a corresponding Barker-Henderson [178] diameter $d_{kk}$,

$$d_{kk} = \int_0^{\sigma_{kk}} \left[1 - \exp\left(-\frac{u_{kk}^{Mie}(r_{kk})}{k_B T}\right)\right], \tag{A.8}$$

which is obtained by means of Gauss-Laguerre quadrature [425]. The free energy of the reference hard-sphere mixtures is obtained using the expression of Boublik [426] and Mansoori *et al.* [427] as

$$a^{HS} = \frac{6}{\pi\rho_s}\left[\left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0\right)\ln(1-\zeta_3) + 3\frac{\zeta_1\zeta_2}{1-\zeta_3} + \frac{\zeta_2^3}{\zeta_3(1-\zeta_3^2)}\right] \tag{A.9}$$

where the total segment number density is given by

$$\rho_s = \rho \sum_{i=1}^{N_C} x_i \sum_{k=1}^{N_G} \nu_{ki} \tag{A.10}$$

and the moment densities are given by

$$\zeta_m = \frac{\pi\rho_s}{6} \sum_{k=1}^{N_G} x_{s,k} d_{kk}^m \tag{A.11}$$

for $m = 0, 1, 2, 3$. The effective hard-sphere diameter $d_{kk}$ is obtained from Eq. A.8 and the fraction of segments $x_{s,k}$ of a group $k$ from Eq. A.7. After substituting Eq. A.7 into Eq. A.11 and expressing Eq. A.9 as a function of $\rho$ using Eq. A.10, one obtains the usual form of the Helmholtz free energy for the hard-sphere mixture [426, 427]:

$$\frac{A^{HS}}{N k_B T} = \frac{6}{\pi\rho}\left[\left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0\right)\ln(1-\zeta_3) + 3\frac{\zeta_1\zeta_2}{1-\zeta_3} + \frac{\zeta_2^3}{\zeta_3(1-\zeta_3^2)}\right]. \tag{A.12}$$

### A.2.2 Mean-attractive energy

The first-order term $a_{1,kl}$ is usually referred to as the mean-attractive energy and can be expressed as

$$a_{1,kl} = 2\pi\rho_s \int_\sigma^\infty g_d^{HS}(r) u_{kl}^{Mie}(r) r^2 dr, \tag{A.13}$$

where $g_d^{HS}(r)$ is the radial distribution function of the reference hard-sphere fluid. In order to obtain an analytical expression rather than using numerical quadrature technique to

evaluate Eq. A.13, the limits of integration can be referred to the hard-sphere diameter $d$ by decomposing Eq. A.13 into two contributions:

$$I_{1A,kl} = 2\pi\rho_s \int_d^\infty g_d^{HS}(r) u_{kl}^{Mie}(r) r^2 dr \tag{A.14}$$

and

$$I_{1B,kl} = -2\pi\rho_s \int_d^\sigma g_d^{HS}(r) u_{kl}^{Mie}(r) r^2 dr \tag{A.15}$$

The integral $I_{1A,kl}$ can be evaluated by using the expressions of the first-order term for the hard-core Sutherland fluid of diameter $d_{kl}$

$$a_{1,kl}^s = 2\pi\rho_s\epsilon_{kl}d_{kl}^3 \int_1^\infty g_d^{HS}(xd) \left(-\frac{1}{x^{\lambda_{kl}}}\right) x^2 dx, \tag{A.16}$$

which is a function of exponent $\lambda$ and segment number density $\rho_s$, an $x$ is given as $x = r/d$ (and accordingly $x_0 = \sigma/d$); $g_d^{HS}(xd)$ is evaluated beforehand by solving the Ornstein-Zernike integral equation with the reference hypernetted chain (RHNC) closure for different values of the exponent $\lambda$. Instead of solving the integral in Eq. A.16, the mean-value theorem (MVT) [24, 53] is applied in order to obtain an analytical expression,

$$\begin{aligned} a_{1,kl}^s(\rho_s, \lambda_{kl}) &= 2\pi\rho_s\epsilon_{kl}d_{kl}^3 g_d^{HS}(\xi) \int_1^\infty \left(-\frac{1}{x^{\lambda_{kl}}}\right) x^2 dx \\ &= 2\pi\rho_s\epsilon_{kl}d_{kl}^3 g_d^{HS}(\xi) \left(\frac{1}{\lambda_{kl}-3}\right), \end{aligned} \tag{A.17}$$

so that $g_d^{HS}$ is expressed as a function of a distance $\xi$ which satisfies the MVT equality. Hence, $g_d^{HS}(\xi)$ is mapped to the contact RDF $g_d^{HS}(d)$ evaluated at an effective packing fraction $\zeta_{x,eff}$, employing a dependence which is consistent with the Carnahan and Starling expression [34] (Eq. A.24)

$$a_{1,kl}^s(\rho_s, \lambda_{kl}) = -2\pi\rho_s \left(\frac{\epsilon_{kl}d_{kl}^3}{\lambda_{kl}-3}\right) \frac{1-\zeta_{x,eff}/2}{(1-\zeta_{x,eff})^3}. \tag{A.18}$$

The effective packing fraction $\zeta_{eff}$ has been parameterised for the range $5 < \lambda \leq 100$ and correlated as [24]

$$\zeta_{eff,kl} = c_{1,kl}\zeta_x + c_{2,kl}\zeta_x^2 + c_{3,kl}\zeta_x^3 + c_{4,kl}\zeta_x^4, \tag{A.19}$$

where $\zeta_x$ is the packing fraction of a hypothetical pure fluid expressed in van der Waals

one-fluid form,

$$\zeta_x = \frac{\pi \rho_s}{6} \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} x_{s,k} x_{s,l} d_{kl}^3, \tag{A.20}$$

and the coefficients $c_i$ are functions of $\lambda$ and are given as following:

$$\begin{pmatrix} c_{1,kl} \\ c_{2,kl} \\ c_{3,kl} \\ c_{4,kl} \end{pmatrix} = \begin{pmatrix} 0.81096 & 1.7888 & -37.578 & 92.284 \\ 1.0205 & -19.341 & 151.26 & -463.50 \\ -1.9057 & 22.845 & -228.14 & 973.92 \\ 1.0885 & -6.1962 & 106.98 & -677.64 \end{pmatrix} \times \begin{pmatrix} 1 \\ \lambda_{kl} \\ \lambda_{kl}^2 \\ \lambda_{kl}^3 \end{pmatrix}. \tag{A.21}$$

Once $a_{1,kl}^s(\rho_s, \lambda_{kl})$ is known, the integral $I_{1A}$ (Eq. A.14) can be evaluated analytically as

$$I_{1A,kl} = \mathcal{C}_{kl} \epsilon_{kl} \left( x_0^{\lambda_{a,kl}} a_{1,kl}^s(\rho_s, \lambda_{a,kl}) - x_0^{\lambda_{r,kl}} a_{1,kl}^s(\rho_s, \lambda_{r,kl}) \right). \tag{A.22}$$

The second integral $I_{1B}$ in Eq. A.15 embrace a short integration range between $\sigma$ and $d$. It has been shown by Nezbeda and Iglesias-Silva [428] that in this case $g_d^{HS}(r)$ can be approximated as a first-order Taylor expansion of the contact value of $g^{HS}(r)$ and its first derivative with respect to separation $r$, so that

$$g_d^{HS}(xd) \simeq g_d^{HS}(d) + (x-1) \left( \frac{dg_d^{HS}(xd)}{dx} \right)_{x=1}. \tag{A.23}$$

The Carnahan and Starling relation is used to evaluate $g_d^{HS}(d)$ and the Percus-Yevick approximation is used to obtain an expression for the derivative at contact:

$$g_d^{HS}(d) = \frac{1 - \zeta_x/2}{(1-\zeta_x)^3} \tag{A.24}$$

$$\left( \frac{dg_d^{HS}(xd)}{dx} \right)_{x=1} = -\frac{9\zeta_x(1+\zeta_x)}{2(1-\zeta_x)^3}. \tag{A.25}$$

By inserting Eqs. A.24 and A.25 into Eq. A.15, we obtain the following analytical expression for the integral:

$$I_{1B} = \mathcal{C} \left( x_0^{\lambda_a} B_{kl}(\rho_s, \lambda_{a,kl}) - x_0^{\lambda_r} B_{kl}(\rho_s, \lambda_{r,kl}) \right), \tag{A.26}$$

where

$$B_{kl}(\rho_s, \lambda_{kl}) = 2\pi \rho_s d_{kl}^3 \epsilon_{kl} \left( \frac{1 - \zeta_x/2}{(1-\zeta_x)^3} I(\lambda_{kl}) - \frac{9\zeta_x(1+\zeta_x)}{2(1-\zeta_x)^3} J(\lambda_{kl}) \right). \tag{A.27}$$

The integrals $I(\lambda_{kl})$ and $J(\lambda_{kl})$ are evaluated over the distance $x$ and are functions of $\lambda_{kl}$:

$$I(\lambda_{kl}) = \int_1^{x_{0,kl}} \frac{x^2}{x^{\lambda_{kl}}} dx = \frac{1 - x_{0,kl}^{(3-\lambda_{kl})}}{(\lambda_{kl} - 3)} \tag{A.28}$$

and

$$J(\lambda_{kl}) = \int_1^{x_{0,kl}} \frac{x^3 - x^2}{x^{\lambda_{kl}}} dx = \frac{1 - x_{0,kl}^{(4-\lambda_{kl})}(\lambda_{kl} - 3) + x_{0,kl}^{(3-\lambda_{kl})}(\lambda_{kl} - 4)}{(\lambda_{kl} - 3)(\lambda_{kl} - 4)}. \tag{A.29}$$

After further algebra, the mean-attractive energy term $a_1$ from Eq. A.13 can be expressed in the following compact form:

$$a_{1,kl} = \mathcal{C}_{kl} \left[ x_{0,kl}^{\lambda_{a,kl}} (a_{1,kl}^s(\rho_s, \lambda_{a,kl}) + B_{kl}(\rho_s, \lambda_{a,kl})) - x_{0,kl}^{\lambda_{r,kl}} (a_{1,kl}^s(\rho_s, \lambda_{r,kl}) + B_{kl}(\rho_s, \lambda_{r,kl})) \right]. \tag{A.30}$$

### A.2.3 Fluctuation term

The second-order perturbation (fluctuation) term of the current SAFT-$\gamma$ Mie [24] version of the EoS is based on the modified compressibility approximation (MCA) proposed by Zhang [429], which can be expressed as

$$a_{2,kl} = -\pi \rho_s K^{HS} (1 + \chi_{kl}) \int_\sigma^\infty g_d^{HS}(r) [u_{kl}^{Mie}(r)]^2 r^2 dr. \tag{A.31}$$

As for the $a_{1,kl}$ term (Eq. A.13), $a_{2,kl}$ can be expressed in terms of the hard-sphere diameter for the lower bounds of integration by decomposing $a_{2,kl}$ into two separate contributions $I_{2A,kl}$ and $I_{2B,kl}$:

$$I_{2A,kl} = -\pi \rho_s K^{HS} (1 + \chi_{kl}) \int_d^\infty g_d^{HS}(r) [u_{kl}^{Mie}(r)]^2 r^2 dr \tag{A.32}$$

$$I_{2B,kl} = \pi \rho_s K^{HS} (1 + \chi_{kl}) \int_d^\sigma g_d^{HS}(r) [u_{kl}^{Mie}(r)]^2 r^2 dr \tag{A.33}$$

Employing Eqs. A.18 and A.27 with Eq. A.31, we obtain

$$\begin{aligned} a_{2,kl} = \frac{1}{2} K^{HS} (1 + \chi_{kl}) \epsilon_{kl} \mathcal{C}_{kl}^2 \Big[ & x_{0,kl}^{2\lambda_{a,kl}} (a_{1,kl}^s(\rho_s, 2\lambda_{a,kl}) + B_{kl}(\rho_s, 2\lambda_{a,kl})) \\ & - 2 x_{0,kl}^{(\lambda_{a,kl} + \lambda_{r,kl})} (a_{1,kl}^s(\rho_s, \lambda_{a,kl} + \lambda_{r,kl}) + B_{kl}(\rho_s, \lambda_{a,kl} + \lambda_{r,kl})) \\ & + x_{0,kl}^{2\lambda_{r,kl}} (a_{1,kl}^s(\rho_s, 2\lambda_{r,kl}) + B_{kl}(\rho_s, 2\lambda_{r,kl})) \Big], \end{aligned} \tag{A.34}$$

where $K^{HS}$ is the isothermal compressibility of the effective van der Waals one-fluid characterised at a packing fraction $\zeta_x$ (Eq. A.20), obtained from the appropriate density derivative of the Carnahan and Starling equation:

$$K^{HS} = k_B T \left( \frac{\partial \rho_s}{\partial P^{HS}} \right)_T = \frac{(1-\zeta_x)^4}{1 + 4\zeta_x + 4\zeta_x^2 - 4\zeta_x^3 + \zeta_x^4}. \tag{A.35}$$

The empirical correction factor $\chi_{kl}$ is established by correlating the computer simulation values of the fluctuation contribution with following expression:

$$\chi_{kl} = f_1(\alpha_{kl})\zeta_x^* + f_2(\alpha_{kl})(\zeta_x^*)^5 + f_3(\alpha_{kl})(\zeta_x^*)^8. \tag{A.36}$$

The functions $f_i (i = 1, 2, 3)$, which are specified in Eq. A.40, depend on both the attractive and the repulsive exponents through a dimensionless van der Waals attractive constant $\alpha$ which is defined as

$$\alpha_{kl} = \frac{1}{\epsilon_{kl}\sigma_{kl}^3} \int_\sigma^\infty u_{kl}^{Mie}(r) r^2 dr = \mathcal{C}_{kl} \left( \frac{1}{\lambda_{a,kl} - 3} - \frac{1}{\lambda_{r,kl} - 3} \right). \tag{A.37}$$

The packing fraction $\zeta_x^*$ is defined in a similar way to Eq. A.20, but now by referring to $\sigma_{kl}$:

$$\zeta_x^* = \frac{\pi \rho_s}{6} \sum_{k=1}^{N_G} \sum_{l=1}^{N_G} x_{s,k} x_{s,l} \sigma_{kl}^3. \tag{A.38}$$

### A.2.4 The third-order term

The third-order term $a_3$ is introduced in order to improve the representation of the fluid properties near the critical point and is obtained from fitting to vapour-liquid equilibria and critical densities for different Mie fluids. By correlating to appropriate computer simulation data, following expression for $a_3$ is obtained :

$$a_{3,kl} = -\epsilon_{kl}^3 f_4(\alpha_{kl}) \zeta_x^* \exp[f_5(\alpha_{kl})\zeta_x^* + f_6(\alpha_{kl})(\zeta_x^*)^2], \tag{A.39}$$

where the functions $f_i$ $(i = 4, 5, 6)$ are defined as

$$f_m(\alpha_{kl}) = \frac{\sum_{n=0}^{3} \phi_{m,n} \alpha_{kl}^n}{1 + \sum_{n=0}^{3} \sum_{n=4}^{6} \phi_{m,n} \alpha_{kl}^{n-3}} \tag{A.40}$$

and coefficients $\phi_{m,n}$ are given in Table A.1.

**Table A.1:** Coefficients $\phi_{m,n}$ for Eq. A.40

| n | $\phi_{1,n}$ | $\phi_{2,n}$ | $\phi_{3,n}$ | $\phi_{4,n}$ | $\phi_{5,n}$ | $\phi_{6,n}$ | $\phi_{7,n}$ |
|---|---|---|---|---|---|---|---|
| 0 | 7.5365557 | -359.44 | 1550.9 | -1.19932 | -1911.28 | 9236.9 | 10 |
| 1 | -37.60463 | 1825.6 | -5070.1 | 9.063632 | 21390.175 | -129430 | 10 |
| 2 | 71.745953 | -3168 | 6534.6 | -17.9482 | -51320.7 | 357230 | 0.57 |
| 3 | -46.83552 | 1884.2 | -3288.7 | 11.34027 | 37064.54 | -315530 | -6.7 |
| 4 | -2.467982 | -0.82376 | -2.7171 | 20.52142 | 1103.742 | 1390.2 | -8 |
| 5 | -0.50272 | -3.1935 | 2.0883 | -56.6377 | -3264.61 | -4518.2 | |
| 6 | 8.0956883 | 3.709 | 0 | 40.53683 | 2556.181 | 4241.6 | |

## A.3 Chain contribution

The SAFT formulation for the heteronuclear chains is based on an effective molecular representation of the contact value of the radial distribution function. For this, the effective molecular parameters $\bar{\sigma_{ii}}$, $\bar{d_{ii}}$, $\bar{\epsilon_{ii}}$, and $\bar{\lambda_{ii}}$ for each species $i$ are introduced. The molecular size and energy parameters do not depend on the composition of mixture, but are expressed in terms of the molecular fraction $z_{ki}$ of a group $k$ in the molecule $i$:

$$z_{ki} = \frac{\nu_{ki}\nu_k^*}{\sum_{i=1}^{N_G} \nu_{ki}\nu_k^*}. \tag{A.41}$$

The molecular parameters $\bar{\sigma_{ii}}$, $\bar{d_{ii}}$, $\bar{\epsilon_{ii}}$, and $\bar{\lambda_{ii}}$ are obtained as:

$$\bar{\sigma_{ii}}^3 = \sum_{k=1}^{N_G}\sum_{l=1}^{N_G} z_{ki}z_{li}\sigma_{kl}^3, \tag{A.42}$$

$$\bar{d_{ii}}^3 = \sum_{k=1}^{N_G}\sum_{l=1}^{N_G} z_{ki}z_{li}d_{kl}^3, \tag{A.43}$$

$$\bar{\epsilon_{ii}}^3 = \sum_{k=1}^{N_G}\sum_{l=1}^{N_G} z_{ki}z_{li}\epsilon_{kl}^3, \tag{A.44}$$

and

$$\bar{\lambda_{ii}}^3 = \sum_{k=1}^{N_G}\sum_{l=1}^{N_G} z_{ki}z_{li}\lambda_{kl}^3. \tag{A.45}$$

The chain term can be expressed as a function of molecular parameters:

$$\frac{A^{Chain}}{Nk_BT} = -\sum_{i=1}^{N_C} x_i \sum_{k=1}^{N_G}(\nu_{ki}\nu_k^* - 1)\ln g_{ii}^{Mie}(\bar{\sigma}, \zeta_x) \tag{A.46}$$

where $g_{ii}^{Mie}(\bar{\sigma}, \zeta_x)$ is the one-fluid radial distribution function of the effective Mie reference (characterised by Eqs. A.42, A.43, A.44, and A.45) at a packing fraction $\zeta_x$ evaluated at the contact value $\bar{\sigma_{ii}}$. An accurate representation of $g_{ii}^{Mie}(\bar{\sigma}, \zeta_x)$ is therefore crucial for a reliable description of the properties of chain fluids. The exponential expansion to the second-order has been proven convenient and rapidly convergent, avoiding the possibility of negative values of the pair correlation function. It is given as

$$g_{ii}^{Mie}(\bar{\sigma_{ii}}, \zeta_x) = g_d(\bar{\sigma_{ii}}) \exp\left(\frac{\beta\bar{\epsilon_{ii}} g_1(\bar{\sigma_{ii}})}{g_d^{HS}(\bar{\sigma_{ii}})} + \frac{(\beta\bar{\epsilon_{ii}})^2 g_2(\bar{\sigma_{ii}})}{g_d^{HS}(\bar{\sigma_{ii}})}\right) \tag{A.47}$$

The contact values of the contributions to the pair distribution function are approximated by the corresponding values at the effective Barker-Henderson diameter:

$$g_0(\bar{\sigma_{ii}}) \approx g_d^{HS}(\bar{d_{ii}}) \tag{A.48}$$

$$g_1(\bar{\sigma_{ii}}) \approx g_1(\bar{d_{ii}}) \tag{A.49}$$

$$g_2(\bar{\sigma_{ii}}) \approx g_2(\bar{d_{ii}}) \tag{A.50}$$

The term $g_d^{HS}(\bar{\sigma_{ii}})$ corresponds to pair distribution function of the hard-sphere fluid of diameter $\bar{d_{ii}}$ evaluated at $\bar{\sigma_{ii}}$. It is obtained based on the expression proposed by Boublik [430]:

$$g_d^{HS}(\bar{\sigma_{ii}}) = \exp(k_0 + k_1 \bar{x_{0,ii}} + k_2 \bar{x_{0,ii}}^2 + k_3 \bar{x_{0,ii}}^3) \tag{A.51}$$

where $\bar{x_{0,ii}} = \bar{\sigma_{0,ii}}/\bar{d_{0,ii}}$, coefficients $k_m$ are functions of the vdW one-fluid packing fraction $\zeta_x$ (Eq. A.20).

The first- and second-order terms $g_1(\bar{d_{ii}})$ and $g_2(\bar{d_{ii}})$ are obtained by equating the pressures from the virial and the thermodynamic routes [53].The second-order term requires an additional simplification, obtained by using the configurational energy and the MCA approximation [429] along with a correction term $\bar{\chi_{ii}}$ [24]. The following expressions are eventually obtained for the first- and the second-order contribution to radial distribution

function at the contact value:

$$g_1(\bar{\sigma_{ii}}) \approx g_1(\bar{d_{ii}}) = \frac{1}{2\pi\bar{\epsilon_{ii}}\bar{d_{ii}}^3}\left[3\frac{\partial \bar{a_{1,ii}}}{\partial\rho_s} - \bar{\mathcal{C}_{ii}}\bar{\lambda_{a,ii}}\bar{x_{0,ii}}^{\bar{\lambda_{a,ii}}}\frac{\bar{a_{1,ii}}^s(\rho_s,\bar{\lambda_{a,ii}}) + B_{ii}(\rho_s,\bar{\lambda_{a,ii}})}{\rho_s}\right.$$
$$\left. + \bar{\mathcal{C}_{ii}}\bar{\lambda_{r,ii}}x_0^{\bar{\lambda_{r,ii}}}\frac{\bar{a_{1,ii}}^s(\rho_s,\bar{\lambda_{r,ii}}) + B_{ii}(\rho_s,\bar{\lambda_{r,ii}})}{\rho_s}\right], \tag{A.52}$$

$$g_2^{MCA}(\bar{\sigma_{ii}}) \approx g_2(\bar{d_{ii}}) = \frac{1}{2\pi\bar{\epsilon_{ii}}^2\bar{d_{ii}}^3}\left[3\frac{\partial}{\partial\rho_s}\left(\frac{\bar{a_{2,ii}}}{1+\bar{\chi_{ii}}}\right)\right.$$
$$- \bar{\epsilon_{ii}}K^{HS}\bar{\mathcal{C}_{ii}^2}\bar{\lambda_{r,ii}}\bar{x_{0,ii}}^{2\bar{\lambda_{r,ii}}}\frac{\bar{a_{1,ii}}^s(\rho_s,2\bar{\lambda_{r,ii}}) + B(\rho_s,2\lambda_{r,ii})}{\rho_s}$$
$$+ \bar{\epsilon_{ii}}K^{HS}\bar{\mathcal{C}_{ii}^2}(\bar{\lambda_{r,ii}}+\bar{\lambda_{a,ii}})\bar{x_{0,ii}}^{(\bar{\lambda_{r,ii}}+\bar{\lambda_{a,ii}})}\frac{\bar{a_{1,ii}}^s(\rho_s,\bar{\lambda_{r,ii}}+\bar{\lambda_{a,ii}}) + \bar{B_{ii}}(\rho_s,\bar{\lambda_{r,ii}}+\bar{\lambda_{a,ii}}}{\rho_s}$$
$$\left. - \bar{\epsilon_{ii}}K^{HS}\bar{\mathcal{C}_{ii}^2}\bar{\lambda_{a,ii}}\bar{x_{0,ii}}^{2\bar{\lambda_{a,ii}}}\frac{\bar{a_{1,ii}}^s(\rho_s,2\bar{\lambda_{a,ii}}) + \bar{B_{ii}}(\rho_s,2\bar{\lambda_{a,ii}})}{\rho_s}\right]. \tag{A.53}$$

# Appendix B

# Integrating the equations of motion: Verlet and leap-frog algorithms

In the Verlet algorithm, the position of molecule $i$ at time $t+\delta t$ is related to the previous position at time $t$ by a Taylor series about $r$:

$$\mathbf{r_i}(t+\delta t) = \mathbf{r_i}(t) + \delta t \frac{\mathbf{dr_i}(t)}{dt} + \frac{1}{2!}(\delta t)^2 \frac{\mathbf{dr_i}^2(t)}{dt^2} + ... = \mathbf{r_i}(t) + \delta t \mathbf{v_i}(t) + \frac{1}{2}(\delta t)^2 \mathbf{a_i}(t) + ... \quad \text{(B.1)}$$

Analogously, for the time $t-\delta t$, the Taylor expansion yields

$$\mathbf{r_i}(t-\delta t) = \mathbf{r_i}(t) - \delta t \mathbf{v_i}(t) + \frac{1}{2}(\delta t)^2 \mathbf{a_i}(t) - ... \quad \text{(B.2)}$$

The summation of Eqs. B.1 and B.2 and substitution $\mathbf{a_i} = \frac{\mathbf{F_i}}{m}$ gives the position of the particle only as a function of its two previous positions and the force:

$$\mathbf{r_i}(t+\delta t) = 2\mathbf{r_i}(t) - \mathbf{r_i}(t-\delta t) + \frac{\mathbf{F_i}(t)}{m}(\delta t)^2 + ... \quad \text{(B.3)}$$

Computing the positions with the Verlet algorithm does not require a knowledge of the particle velocity. The velocities are nonetheless useful for estimating the kinetic (and hence the total) energy and can be calculated separately as

$$\mathbf{v_i}(t) = \frac{\mathbf{r_i}(t+\delta t) - \mathbf{r_i}(t-\delta t)}{2\delta t}. \quad \text{(B.4)}$$

However, not handling velocities explicitly can introduce some numerical deficiencies, and therefore, alternative algorithms have been proposed. In our current work, we use the

leap-frog algorithm [431] for the integration of the equation of motion. In the leap-frog algorithm, the new position is calculated as a function of the previous position and the mid-step velocity:

$$\mathbf{r_i}(t+\delta t) = \mathbf{r}(t) + \delta t \mathbf{v_i}(t+\frac{1}{2}\delta t). \tag{B.5}$$

The mid-step velocity is obtained as

$$\mathbf{v_i}(t+\frac{1}{2}\delta t) = \mathbf{v_i}(t-\frac{1}{2}\delta t) + \frac{\mathbf{F_i}(t)}{m}\delta t \tag{B.6}$$

In order to evaluate the total energy ($\mathcal{H} = \mathcal{K} + \mathcal{V}$) and all other properties at time $t$, the velocity at time $t$ is required. It is calculated as

$$\mathbf{v_i}(t) = \frac{1}{2}(\mathbf{v_i}(t+\frac{1}{2}\delta t) + \mathbf{v_i}(t-\frac{1}{2}\delta t)). \tag{B.7}$$

The trajectories, determined in the leap-frog algorithm with Eq. B.5 are identical to the Verlet algorithm trajectories obtained by Eq. B.3.

# Appendix C

# Force Fields developed in this work: the experimental data used and the corresponding parameter sets

**Table C.1:** Experimental data used in the parameter estimation procedure.

| Interactions | Type of data used | Reference |
|---|---|---|
| $CM - CM$ | VLE of $n$-alkanes between the triple point upto 90 % of the critical point, selected fitting points are within step of 20 K | NIST [195] |
| $FF - FF$ | VLE of $n$-perfluoroalkanes between the triple point up to 90 % of the critical point, | NIST [195] |
| $EO - EO$ | VLE of ethers between the triple point upto 90 % of the critical point, selected fitting points are within step of 20 K | NIST [195] |
| $OA - OA$ | VLE of triethylene glycol between the triple point upto 90 % of the critical point selected fitting points are within step of 20 K | NIST [195] |
| $CM - FF$ | LLE of $C_6F_{14} + C_6H_{14}$ mixture between 272 and 296 K | DETHERM [196] |
| $CF - CF$ | liquid density and vapour pressure for SFA | Refs. [198, 199] |
| $CF - FF$ | liquid density and vapour pressure for SFA | Refs. [198, 199] |
| $EO - CM$ | VLE of butyl methyl ether ($C_5H_{12}O$), 1-methoxyheptane ($C_8H_{18}O$), 1-(2-methoxyethoxy)butane ($C_7H_{16}O_2$), between the triple point upto 90 % of the critical point | NIST [195] |
| $OA - CM$ | VLE of 2-butoxyethanol ($C_6H_{14}O_2$), 2-(heptyloxy)ethanol ($C_9H_{20}O_2$) between the triple point upto 90 % of the critical point | NIST [195] |
| $OA - EO$ | Liquid-liquid equilibrium of $C_7E_5$ surfactant in $n$-dodecane between 272-278.5 K | DETHERM [196] |
| $W - W$ (M1) | Saturation density and vapour pressure, fit to each temperature | NIST [195] |
| $W - W$ (M2) | Saturation density and surface tension, fit to each temperature | NIST [195] |
| $CO_2 - W$ (M1) | Pressure-composition isotherm at $T = 473$ K | Ref. [323] |
| $CM - W$ (M1) | Pressure-composition isotherm of water-hexane mixture at $T = 473.2$ K | Refs. [359–361] |
| $CO_2 - CM$ | Pressure-composition isotherm of carbon dioxide-hexane mixture at $T = 313$ K | DETHERM [196] |
| $CM - W$ (M2) | Interfacial tension of $n$-hexane-water interface at $T = 298.15$ K | Ref. [390] |
| $EO - W$ (M2) | Excess enthalpy of 1-methoxy-2-[2-(2-methoxy-ethoxy]-ethane ($C_8H_{18}O_4$) - water mixture at $T = 298.15$ K | DETHERM [196] |
| $OA - W$ (M2) | Excess enthalpy of triethylene glycol-water mixture at $T = 298.15$ K | DETHERM [196] |

**Table C.2:** CG intermolecular interaction parameters obtained with SAFT-$\gamma$ Mie EoS ($M_1$, $M_2$ refers to Models 1 and 2, respectively, the indices $W_1$ and $W_2$ indicate level of coarse-graining)

| Compound | Groups | $\sigma$/Å | $\epsilon/k_B/K$ | $\lambda_r$ | $\lambda_a$ | $k_{ij}$ |
|---|---|---|---|---|---|---|
| | | Like-like | | | | |
| $CO_2$ | $(CO_2)$ | 3.741 | 353.55 | 23 | 6.66 | |
| water at $T$ =298 K | | | | | | |
| $H_2O$ | $W_1$ (M1) | 3.0089 | 496.09 | 8 | 6 | |
| $H_2O$ | $W_1$ (M2) | 2.9016 | 305.21 | 8 | 6 | |
| $(H_2O)_2$ | $W_2$ | 3.7467 | 400 | 8 | 6 | |
| triethylene glycol | | | | | | |
| $C_6H_{14}O_4$ | $(OA)4$ | 3.7244 | 461.11 | 19 | 6 | |
| ethers | | | | | | |
| $C_4H_{10}O_2$ | $(EO)2$ | 4.2410 | 409.26 | 19 | 6 | |
| $C_6H_{14}O_3$ | $(EO)3$ | 4.1527 | 405.19 | 19 | 6 | |
| $C_8H_{18}O_4$ | $(EO)4$ | 4.0882 | 399.25 | 19 | 6 | |
| $C_{10}H_{22}O_5$ | $(EO)5$ | 4.0613 | 396.90 | 19 | 6 | |
| $n-$alkanes | | | | | | |
| $C_5H_{12}$ | $(CM)2$ | 4.2449 | 310.78 | 15 | 6 | |
| $C_6H_{14}$ | $(CM)2$ | 4.5089 | 342.00 | 15 | 6 | |
| $C_7H_{16}$ | $(CM)2$ | 4.5736 | 380.87 | 15 | 6 | |
| $C_8H_{18}$ | $(CM)3$ | 4.2412 | 321.65 | 15 | 6 | |
| $C_9H_{20}$ | $(CM)3$ | 4.4212 | 344.83 | 15 | 6 | |
| $C_{10}H_{22}$ | $(CM)3$ | 4.4908 | 363.99 | 15 | 6 | |
| $C_{11}H_{24}$ | $(CM)4$ | 4.2212 | 328.18 | 15 | 6 | |
| $C_{12}H_{26}$ | $(CM)4$ | 4.3635 | 344.42 | 15 | 6 | |
| $C_{13}H_{28}$ | $(CM)4$ | 4.4680 | 354.25 | 15 | 6 | |
| $C_{14}H_{30}$ | $(CM)5$ | 4.2332 | 329.81 | 15 | 6 | |
| $C_{15}H_{32}$ | $(CM)5$ | 4.3286 | 344.85 | 15 | 6 | |
| $n-$perfluoroalkanes | | | | | | |
| $C_6F_{14}$ | $(FF)3$ | 4.3838 | 240.32 | 13 | 6 | |
| $C_8F_{18}$ | $(FF)4$ | 4.2789 | 238.63 | 13 | 6 | |
| $C_{10}F_{22}$ | $(FF)5$ | 4.2095 | 230.91 | 13 | 6 | |
| $-C_2H_4-C_2F_4-$ | $(CF)$ | 5.3087 | 522.33 | 13.95 | 6 | |
| | | Unlike interactions for SFA | | | | |
| $F_6H_8$ | $(CM)2+(CF)$ | | 315.37 | 14.46 | 6 | 0.2457 |
| $F_6H_8$ | $(CF)+(FF)2$ | | 219.23 | 13.46 | 6 | 0.3681 |
| $F_6H_8$ | $(CM)2+(FF)2$ | | 265 | 13.95 | 6 | 0.0755 |
| | | Unlike interactions for surfactant | | | | |
| $C_{10}E_4$ | $(EO)3+(CM)3$ | | 352.00 | 16.86 | 6 | 0.0461 |
| $C_{10}E_4$ | $(OA)2+(CM)3$ | | 375.50 | 16.86 | 6 | 0.0489 |
| $C_{10}E_4$ | $(OA)2+(EO)3$ | | 392.00 | 19 | 6 | 0.0811 |
| | | Unlike interactions for $CO_2$-alkanes | | | | |
| $CO_2+C_{15}H_{32}$ | $(CO_2)+(CM)5$ | | 318.69 | 18.5 | 6.31 | 0.08 |
| | | Unlike interactions with water at T=298 K | | | | |
| $CO_2+H_2O$ | $(CO_2)+(W)$ (M1) | | 440.23 | 13 | 6.31 | -0.07 |
| $C_{12}H_{26}+H_2O$ | $(CM)4+(W)$ (M1) | | 251.27 | 10.75 | 6 | 0.36 |
| $C_6H_{14}O_4+H_2O$ | $(OA)4+(W_1)$ (M2) | | 497 | 17 | 6 | -0.3561 |
| $C_8H_{18}O_4$ | $(EO)4+(W_1)$ (M2) | | 435 | 11.94 | 6 | -0.3037 |
| $C_{12}H_{26}+H_2O$ | $(CM)4+(W_1)$ (M2) | | 213.31 | 10.75 | 6 | 0.3 |
| $C_6H_{14}O_4+(H_2O)_2$ | $(OA)4+(W_2)$ | | 492 | 11.94 | 6 | -0.1457 |
| $C_8H_{18}O_4+(H_2O)_2$ | $(EO)4+(W_2)$ | | 480 | 11.94 | 6 | -0.2077 |
| $C_{12}H_{26}+(H_2O)_2$ | $(CM)4+(W_2)$ | | 250 | 10.75 | 6 | 0.3205 |

**Table C.3:** Parameter set for Model 1, Mie (8-6) which reproduces saturation liquid density and vapour pressure.

| T/K | $\sigma$/ Å | $\epsilon/k_B$/K |
|---|---|---|
| 283 | 3.0119 | 501.19 |
| 293 | 3.0085 | 497.84 |
| 303 | 3.0060 | 494.55 |
| 313 | 3.0041 | 491.30 |
| 323 | 3.0028 | 488.11 |
| 333 | 3.0020 | 484.97 |
| 343 | 3.0015 | 481.87 |
| 353 | 3.0014 | 478.82 |
| 363 | 3.0016 | 475.81 |
| 373 | 3.0021 | 472.83 |
| 393 | 3.0039 | 466.97 |
| 413 | 3.0065 | 461.20 |
| 433 | 3.0099 | 455.51 |
| 453 | 3.0141 | 449.85 |
| 473 | 3.0192 | 444.21 |
| 493 | 3.0251 | 438.55 |
| 513 | 3.0321 | 432.82 |
| 533 | 3.0403 | 426.99 |
| 553 | 3.0503 | 420.98 |
| 573 | 3.0627 | 414.68 |
| 593 | 3.0790 | 407.90 |
| 613 | 3.1028 | 400.21 |

**Table C.4:** Parameter set for Model 2, Mie (8-6) which reproduces liquid density and surface tension.

| $T$/K | $\sigma$/ Å | $\epsilon/k_B$/K |
|---|---|---|
| 293 | 2.9055 | 304.28 |
| 298 | 2.9016 | 305.21 |
| 313 | 2.8938 | 309.01 |
| 343 | 2.8811 | 318.84 |
| 373 | 2.8737 | 326.85 |
| 393 | 2.8721 | 332.18 |
| 433 | 2.8673 | 340.25 |
| 463 | 2.8660 | 345.43 |
| 493 | 2.8666 | 350.25 |

# Appendix D

# Representation of the VLE properties of the chain molecules by means of the SAFT-$\gamma$ Mie EoS

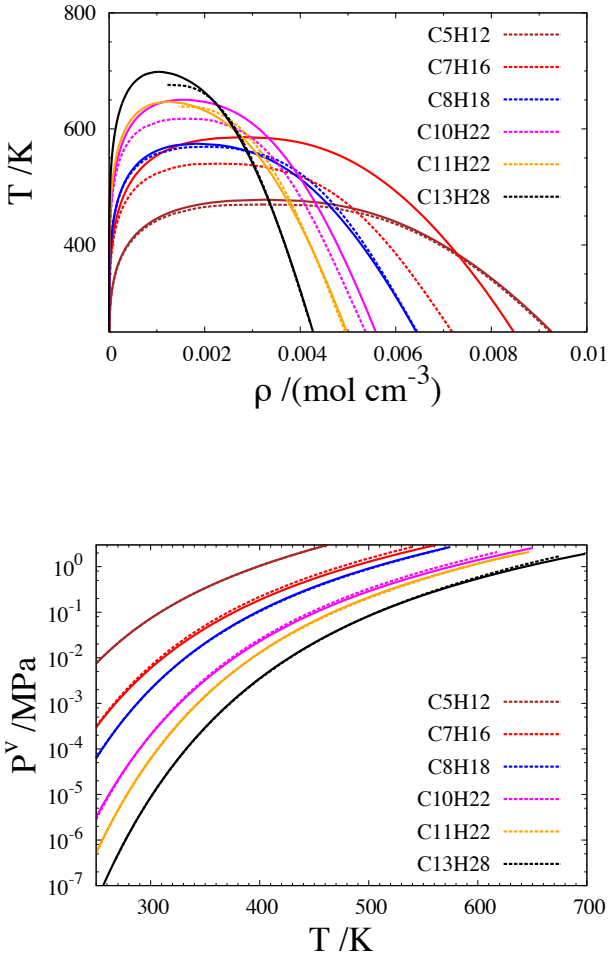



**Figure D.1:** Saturation densities and pressures for homologous series of alkanes non-divisible by three ($C_5H_{12}$-$C_{14}H_{30}$). Dashed line denotes the smoothed experimental data from NIST [195]; continuous lines are the results from SAFT-$\gamma$ Mie EoS using the parameters from Table 3.4.

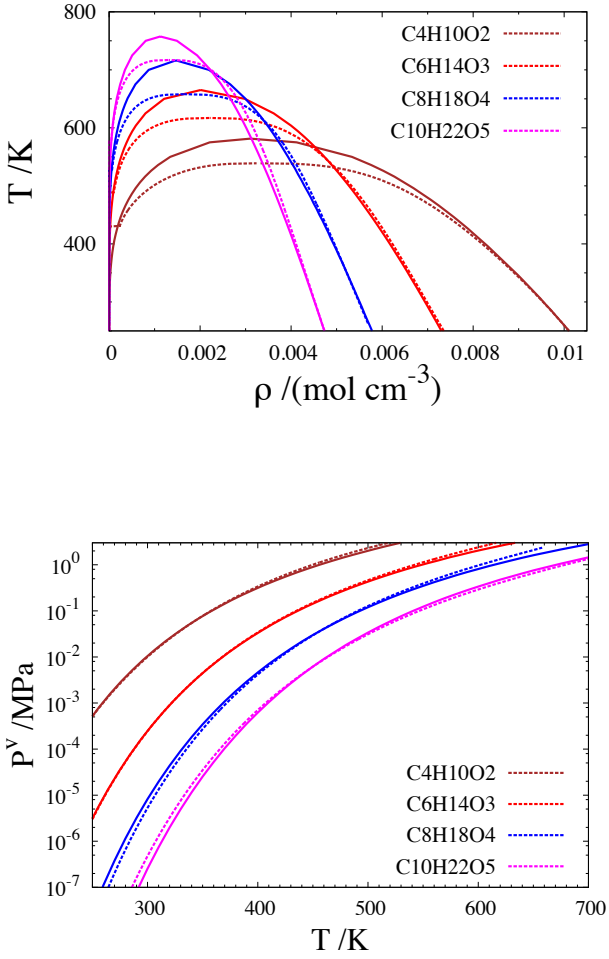



**Figure D.2:** Saturation densities and pressures for ethers ($C_4H_{10}O_2$ to $C_{10}H_{26}O_5$). Dashed line denotes the smoothed experimental data from NIST [195]; continuous lines are the results from SAFT-$\gamma$ Mie EoS using the parameters from Table 3.6.

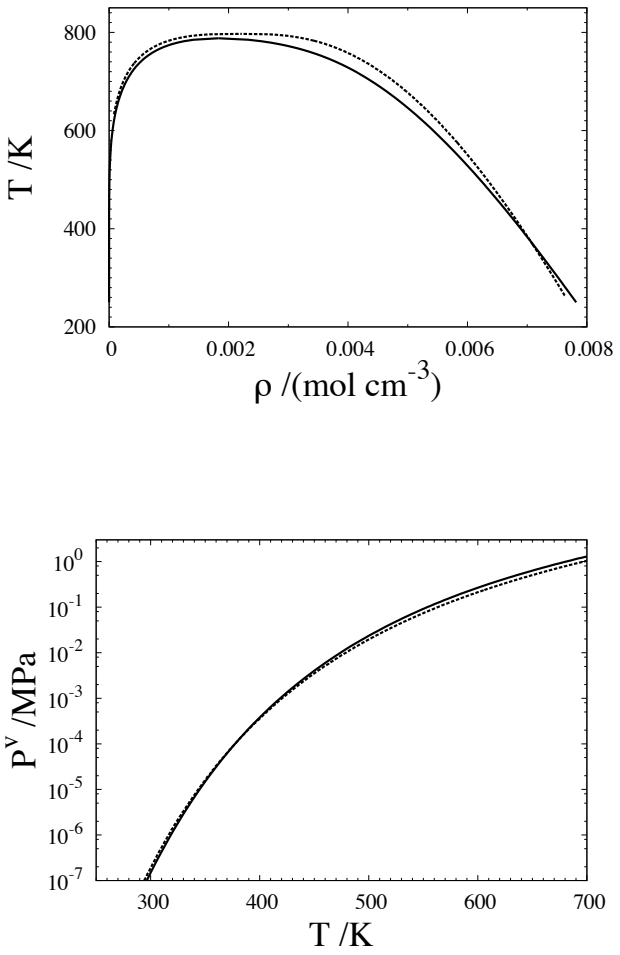



**Figure D.3:** Saturation density and pressure for triethyleneglycol. Dashed line denotes the smoothed experimental data from NIST [195]; continuous line is the result from SAFT-$\gamma$ Mie EoS based on the Mie (19-6) with $\sigma$=3.7244 Åand $\epsilon/k_B$=461.11 K.

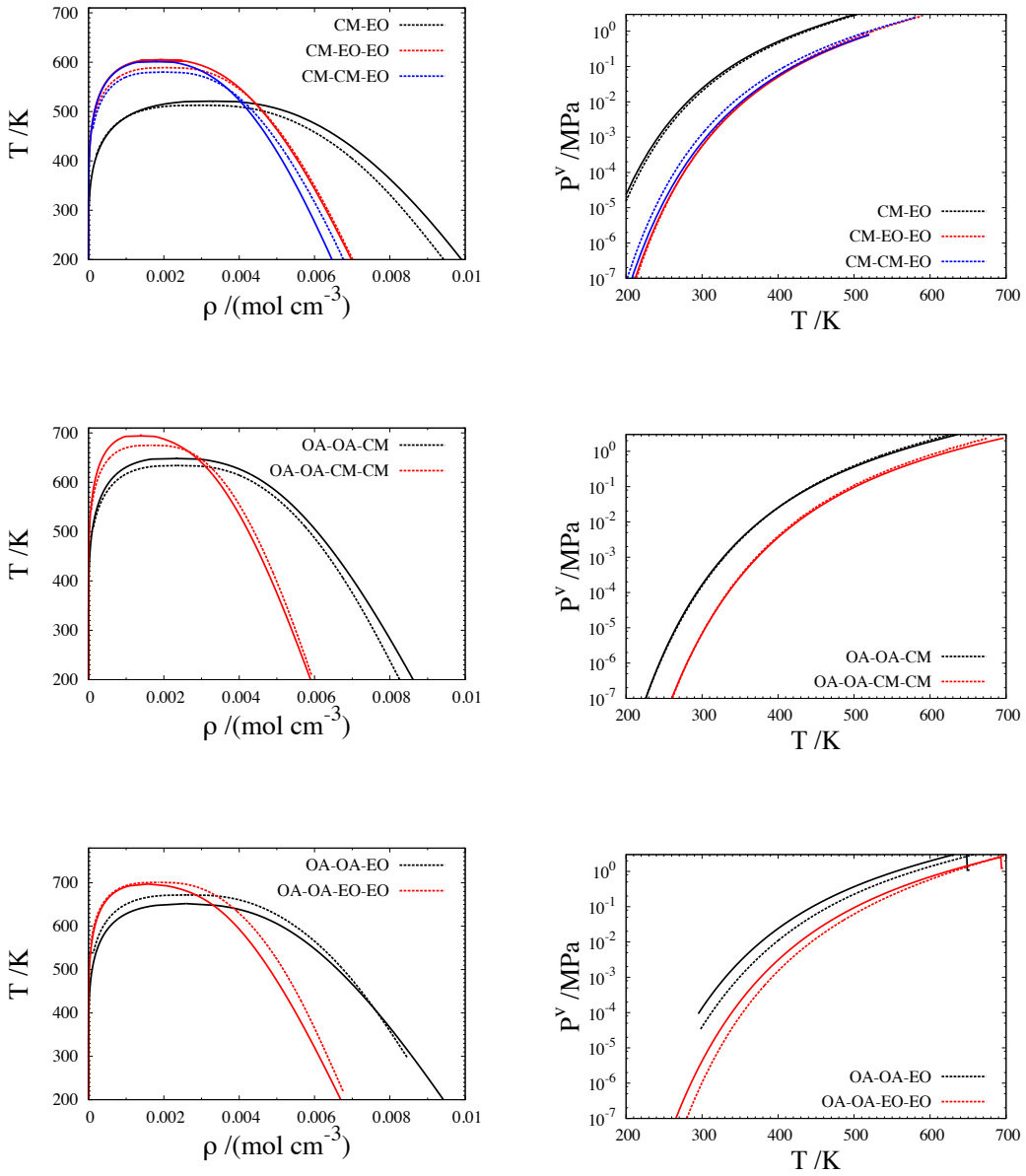



**Figure D.4:** Liquid densities (on the left) and vapour pressures (on the right) of surfactants of different chain lengths. Dashed line denotes the smoothed experimental data from NIST [195]; continuous lines are the results from SAFT-$\gamma$ Mie EoS using the parameters from Table 3.8. The upper line displays molecules containing $CM$ and $EO$ groups, such as butyl methyl ether ($CM-EO$, black line), 1-(2-methoxyethoxy)butane ($CM-EO-EO$, red line), and 2-methoxyheptane ($CM-CM-EO$, blue line). The middle line displays molecules containing $OA$ and $CM$ groups, such as 2-butoxyethanol ($OA-OA-CM$, black line) and 2-(heptyloxy)ethanol ($OA-OA-CM-CM$, red line). The lower line displays molecules containing $OA$ and $EO$ groups, such as 2-(2-methoxyethoxy)ethanol ($OA-OA-EO$, black line) and 2-(2-(2-methoxyethoxy)ethoxy)ethanol ($OA-OA-EO-EO$, red line).

# Appendix E

# Representation of properties of water by means of the molecular simulations

**Table E.1:** The vapour-liquid coexistence properties including the saturated liquid $\rho^L/(\mathrm{kgm}^{-3})$ and vapour $\rho^V/(\mathrm{kgm}^{-3})$ densities, and vapour pressure $P^v/(\mathrm{MPa})$ for the Mie (8-6) Model 1 of water at $T$=393 K as a function of the cutoff radius $R_c/$ Å obtained via MD simulations.

| $R_c$ | $\rho^L$ | $\rho^V$ | $P^v$ |
|---|---|---|---|
| 15 | 936 | 1.36 | 0.243 |
| 20 | 944 | 1.16 | 0.214 |
| 25 | 947 | 1.13 | 0.200 |
| 30 | 949 | 1.05 | 0.175 |
| 35 | 949 | 1.02 | 0.176 |

**Table E.2:** The vapour-liquid coexistence properties including the saturated liquid $\rho^L/(\mathrm{kgm}^{-3})$ and vapour $\rho^V/(\mathrm{kgm}^{-3})$ densities, vapour pressure $P^v/(\mathrm{MPa})$, enthalpy of vaporisation $\Delta H^v/(\mathrm{kJmol}^{-1})$, and vapour-liquid interfacial tension $\gamma/(\mathrm{mNm}^{-1})$ for the Mie (8-6) Model 1 of water at different state points $T/\mathrm{K}$ obtained via MD simulations using the cutoff radius $R_c = 30$ Å.

| T | $\rho^L$ | $\rho^V$ | $P^v$ | $\gamma$ | $\Delta H^v$ |
|---|---|---|---|---|---|
| 343 | 994 | 0.20 | 0.032 | 158 | 35.5 |
| 363 | 977 | 0.40 | 0.067 | 146 | 34.6 |
| 393 | 949 | 1.11 | 0.198 | 131 | 33.2 |
| 413 | 928 | 2.10 | 0.393 | 120 | 32.3 |
| 433 | 906 | 3.14 | 0.610 | 108 | 31.4 |
| 453 | 883 | 4.89 | 0.997 | 98 | 30.3 |
| 493 | 832 | 10.61 | 2.245 | 78 | 28.3 |
| 533 | 774 | 20.79 | 4.521 | 58 | 25.8 |
| 573 | 704 | 36.68 | 8.000 | 39 | 22.9 |
| 613 | 611 | 64.25 | 13.391 | 21 | 20.3 |

**Table E.3:** The vapour-liquid coexistence properties including the saturated liquid $\rho^L/(\mathrm{kgm}^{-3})$ and vapour $\rho^V/(\mathrm{kgm}^{-3})$ densities, vapour pressure $P^v/(\mathrm{MPa})$, enthalpy of vaporisation $\Delta H^v/(\mathrm{kJmol}^{-1})$, and vapour-liquid interfacial tension $\gamma/(\mathrm{mNm}^{-1})$ for the Mie (8-6) Model 2 of water at different state points $T/\mathrm{K}$ obtained via MD simulations using the cutoff radius $R_c = 20$ Å.

| T | $\rho^L$ | $\rho^V$ | $P^v$ | $\gamma$ | $\Delta H^v$ |
|---|---|---|---|---|---|
| 293 | 999 | 4.02 | 0.527 | 73 | 20.6 |
| 298 | 998 | 4.28 | 0.536 | 72 | 20.6 |
| 313 | 992 | 5.91 | 0.820 | 68 | 20.6 |
| 343 | 979 | 9.24 | 1.512 | 64 | 21.4 |
| 373 | 960 | 14.06 | 2.235 | 59 | 20.7 |
| 393 | 943 | 17.92 | 2.953 | 55 | 20.6 |
| 433 | 907 | 28.73 | 5.004 | 46 | 20.0 |
| 463 | 875 | 39.65 | 7.110 | 39 | 19.4 |
| 493 | 840 | 54.52 | 9.901 | 33 | 18.5 |

# Appendix F

# Representation of properties of binary mixtures comprising water, carbon dioxide, and $n-$alkanes

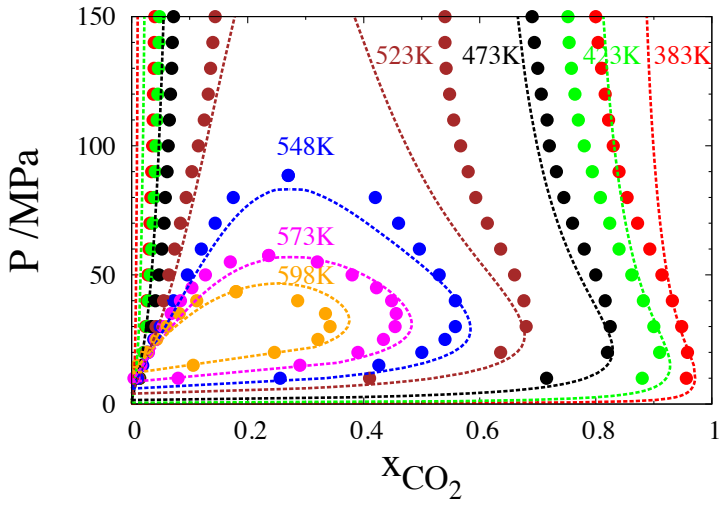



**Figure F.1:** Pressure-composition ($Px$) phase diagram for carbon dioxide+water mixture at isotherms corresponding to $T$ =383, 423, 473, 523, 548, 573, and 598 K. The filled circles correspond to experimental fluid-liquid equilibrium data [323]. The dashed lines correspond to the predictions obtained by SAFT-$\gamma$ with $k_{ij} = -0.07$, the open circles are the MD simulation results for our SAFT-$\gamma$ CG Mie force field

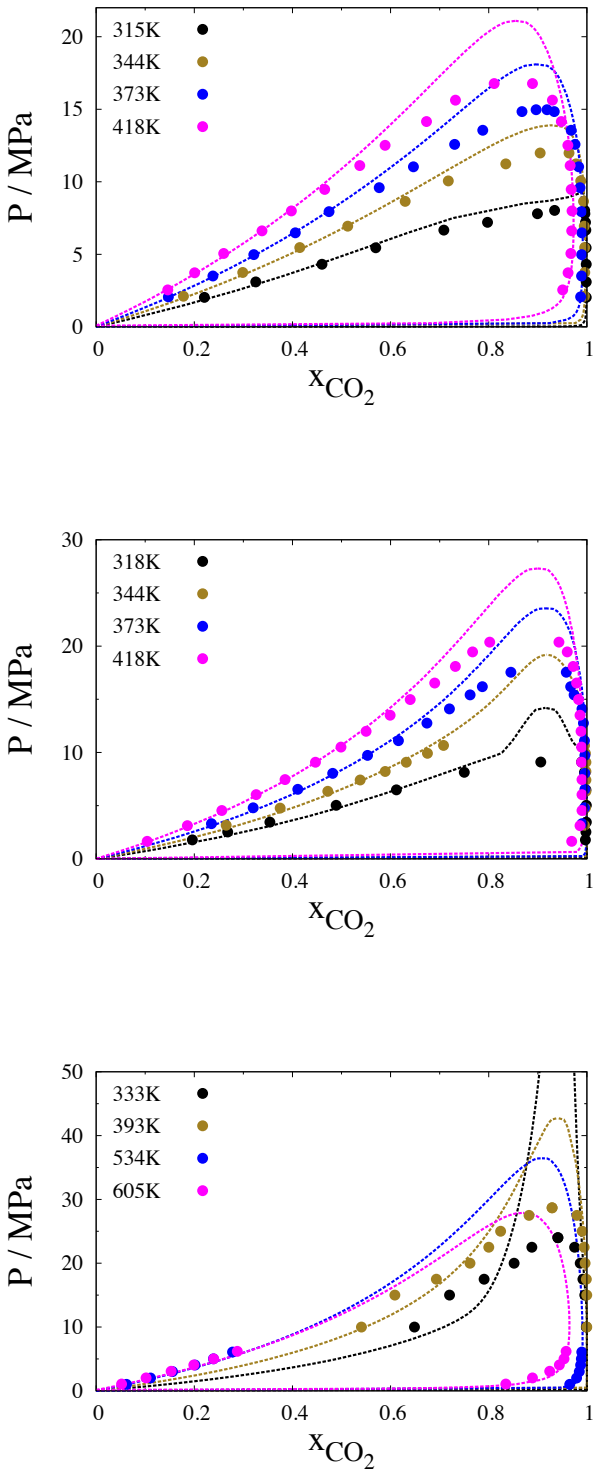



**Figure F.2:** Pressure-composition ($Px$) phase diagram for mixtures of carbon dioxide with $n-$nonane, $n-$dodecane, and $n-$octadecane at different isotherms. Legend as in Figure 5.19.

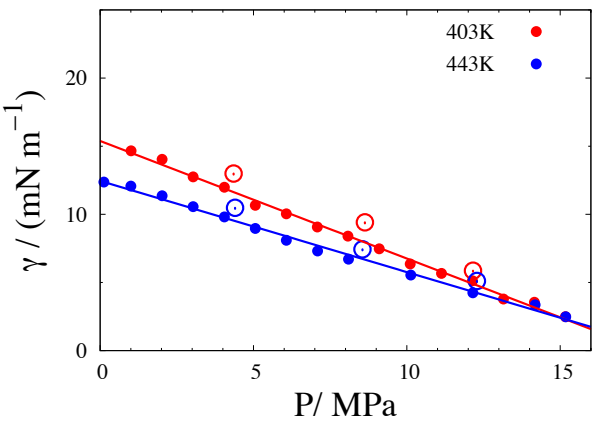



**Figure F.3:** Interfacial tension as a function of pressure in carbon dioxide+ $n-$dodecane system at $T = 403$ and 443 K. The filled circles are the experimental data from Ref. [432]. The solid straight line was added as a guide to the eye. The open circles are the MD simulation results for SAFT-$\gamma$ CG Mie force field.

# Appendix G

# Aggregation number distributions of the low concentration non-ionic surfactant solutions

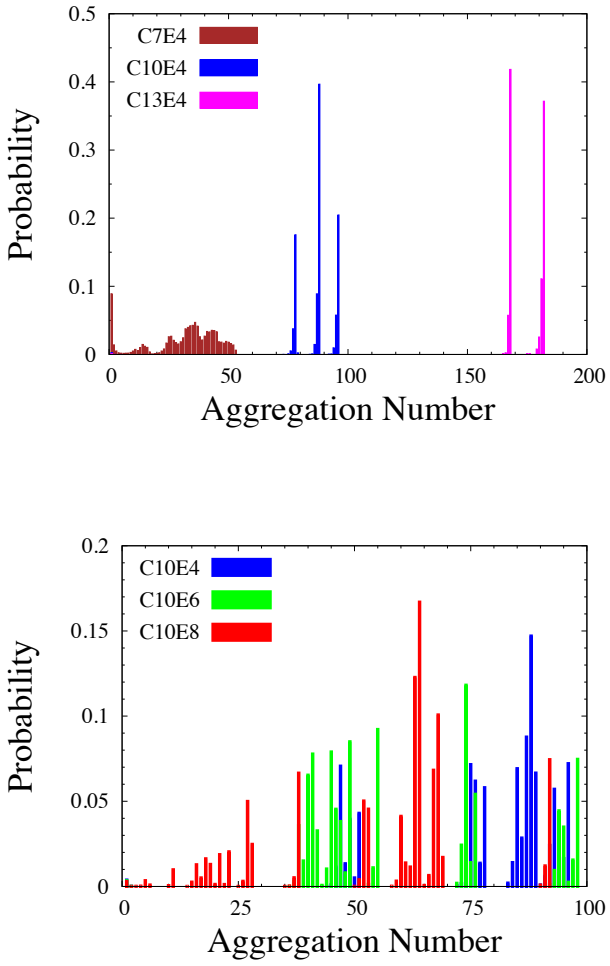



**Figure G.1:** Aggregation number distribution at the surfactant concentrations $x = 3.5 \times 10^{-3}$ for different alkyl and ethoxy chain length calculated in MD simulation.

# References


[1] P. Paricaud, A. Galindo, and G. Jackson. Recent advances in the use of the saft approach in describing electrolytes, interfaces, liquid crystals and polymers. *Fluid Phase Equilibria*, 194:87–96, 2002.

[2] K. E. Gubbins and J. D. Moore. Molecular modeling of matter: Impact and prospects in engineering. *Industrial and Engineering Chemistry Research*, 49(7):3026–3046, 2010.

[3] P. Ungerer, B. Tavitian, and A. Boutin. *Applications of Molecular Simulation in the Oil and Gas Industry: Monte Carlo Methods*. IFP Publications. Editions Technip, 2005.

[4] P. Ungerer, C. Nieto-Draghi, V. Lachet, A. Wender, A. Di Lella, A. Boutin, B. Rousseau, and A. H. Fuchs. Molecular simulation applied to fluid properties in the oil and gas industry. *Molecular Simulation*, 33(4-5):287–304, 2007.

[5] G. E. Moore. Cramming more components onto integrated circuits (Reprinted from Electronics, pg 114-117, April 19, 1965). *Proceedings of the IEEE*, 86(1):82–85, 1998.

[6] W. Eckhardt and A. Heinecke. SuperMUC boosts the largest molecular dynamics simulation by 4X in number of particles. *inSiDe*, 11(1):1, 2013.

[7] H. I. Ingolfsson, C. A. Lopez, J. J. Uusitalo, D. H. Jong, S. M. Gopal, X. Periole, and S. J. Marrink. The power of coarse graining in biomolecular simulations. *WIREs Computational Molecular Science*, 2013.

[8] A. A. Louis. Beware of density dependent pair potentials. *Journal of Physics-Condensed Matter*, 14(40):9187–9206, 2002.

[9] D. Peng and D. B. Robinson. New 2-constant equation of state. *Industrial and Engineering Chemistry Fundamentals*, 15(1):59–64, 1976.

[10] G. Soave. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chemical Engineering Science*, 27(6):1197, 1972.

[11] D. Myers. *Surfactant Science and Technology*. Wiley, 2005.

[12] J.M. Prausnitz, R.N. Lichtenthaler, and E.G. de Azevedo. *Molecular thermodynamics of fluid-phase equilibria*. Prentice-Hall international series in the physical and chemical engineering sciences. Prentice-Hall, 1986.

[13] J. N. Israelachvili. *Intermolecular and Surface Forces*. Intermolecular and Surface Forces. Academic Press, 3, illustrated edition, 2010.

[14] T. Kihara. *Intermolecular forces*. John Wiley Sons Canada, Limited, 1978.

[15] A. R. Leach. *Molecular Modelling: Principles and Applications*. Pearson Education. Prentice Hall, 2001.

[16] T.M.K. Reed and K.E. Gubbins. *Applied Statistical Mechanics: Thermodynamic and Transport Properties of Fluids*. McGraw-Hill Chemical Engineering Series. McGraw-Hill, 1973.

[17] G. M. Kontogeorgis and G. K. Folas. *Thermodynamic Models for Industrial Applications: From Classical and Advanced Mixing Rules to Association Theories*. John Wiley and Sons Ltd, West Sussex, United Kingdom, 2010.

[18] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, Oxford, 1987.

[19] J. E. Lennard-Jones. On the determination of molecular fields I. From the variation of the viscosity of a gas with temperature. *Proc. R. Soc. Lond. A*, 106(738):441–462, 1924.

[20] J. E. Lennard-Jones. On the determination of molecular fields II. From the equation of state of a gas. *Proc. R. Soc. Lond. A*, 106(738):463–477, 1924.

[21] G. Mie. Zur kinetischen Theorie der einatomigen Körper. *Annalen der Physik*, 11:657–697, 1903.

[22] C. Avendaño, T. Lafitte, A. Galindo, C. S. Adjiman, G. Jackson, and E. A. Müller. SAFT-$\gamma$ force field for the simulation of molecular fluids. 1. A single-site coarse grained model of carbon dioxide. *Journal of Physical Chemistry B*, 115(38):11154–11169, 2011.

[23] T. Lafitte, D. Bessieres, M. M. Pineiro, and J. L. Daridon. Simultaneous estimation of phase behavior and second-derivative properties using the statistical associating fluid theory with variable range approach. *Journal of Chemical Physics*, 124(2):024509, 2006.

[24] T. Lafitte, A. Apostolakou, C. Avendaño, A. Galindo, C. S. Adjiman, E. A. Müller, and G. Jackson. Accurate statistical associating fluid theory for chain molecules formed from Mie segments. *The Journal of Chemical Physics*, 139:154504, 2013.

[25] A. J. Haslam. *An introduction to statistical thermodynamics and its applications in equation of state modelling of fluids*. Department of Chemical Engineering, Imperial College London, 2013.

[26] E. Clapeyron. Puissance motrice de la chaleur. *Journal de l'Ecole Royale Polytechnique*, Tome XIV:153–190, 1834.

[27] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, Sausalito, CA, USA, 2000.

[28] J. D. van der Waals. *Over de Continuiteit van den Gas- en Vloeistoftoestand*. PhD thesis, Leiden, 1873.

[29] M. Kac, G. E. Uhlenbeck, and P. C. Hemmer. On the van der Waals theory of the vapor liquid equilibrium. i. discussion of a one dimensional model. *Journal of Mathematical Physics*, 4(2):216–228, 1963.

[30] G. E. Uhlenbeck, P. C. Hemmer, and M. Kac. On the van der waals theory of the vaporâĂŘliquid equilibrium. ii. discussion of the distribution functions. *Journal of Mathematical Physics*, 4(2):229–247, 1963.

[31] P. C. Hemmer, M. Kac, and G. E. Uhlenbeck. On the van der waals theory of the vaporâĂŘliquid equilibrium. iii. discussion of the critical region. *Journal of Mathematical Physics*, 5(1):60–74, 1964.

[32] R. L. Scott and P. H. Van Konynenburg. Static properties of solutions. Van der Waals and related models for hydrocarbon mixtures. *Discuss. Faraday Soc.*, 49(87), 1970.

[33] P. H. van Konynenburg and R. L. Scott. Critical lines and phase equilibria in binary van der Waals mixtures. *Philos. Trans. R. Soc. London*, 298(495), 1980.

[34] N. F. Carnahan and K. E. Starling. Equation of state for nonattracting rigid spheres. *Journal of Chemical Physics*, 51:635, 1969.

[35] N. F. Carnahan and K. E. Starling. Intermolecular repulsions and equation of state for fluids. *AIChE Journal*, 18(6):1184, 1972.

[36] O. Redlich and J. N. S. Kwong. On the thermodynamics of solutions. *Chem. Rev.*, 44(1):233–244, 1949.

[37] S. Beret and J. M. Prausnitz. Perturbed hard-chain theory - equation of state for fluids containing small or large molecules. *AIChE Journal*, 21(6):1123–1132, 1975.

[38] J.V. Sengers, R.F. Kayser, C.J. Peters, and H.J. White. *Equations of State for Fluids and Fluid Mixtures*, volume 5 of *Experimental Thermodynamics*. Elsevier Science Limited, 2000.

[39] M. S. Wertheim. Fluids with highly directional attractive forces. 1. Statistical thermodynamics. *Journal of Statistical Physics*, 35(1-2):19–34, 1984.

[40] M. S. Wertheim. Fluids with highly directional attractive forces. 2. Thermodynamic perturbation-theory and integral-equations. *Journal of Statistical Physics*, 35(1-2):35–47, 1984.

[41] M. S. Wertheim. Fluids with highly directional attractive forces. 3. Multiple attraction sites. *Journal of Statistical Physics*, 42(3-4):459–476, 1986.

[42] M. S. Wertheim. Fluids with highly directional attractive forces. 4. Equilibrium polymerization. *Journal of Statistical Physics*, 42(3-4):477–492, 1986.

[43] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz. SAFT: Equation-of-state solution model for associating fluids. *Fluid Phase Equilibria*, 52:31–38, 1989.

[44] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz. New reference equation of state for associating liquids. *Industrial and Engineering Chemistry Research*, 29(8):1709–1721, 1990.

[45] E. A. Müller and K. E. Gubbins. Molecular-based equations of state for associating fluids: A review of SAFT and related approaches. *Industrial and Engineering Chemistry Research*, 40(10):2193–2211, 2001.

[46] I. G. Economou. Statistical associating fluid theory: A successful model for the calculation of thermodynamic and phase equilibrium properties of complex fluid mixtures. *Industrial and Engineering Chemistry Research*, 41(5):953–962, 2002.

[47] S. P. Tan, H. Adidharma, and M. Radosz. Recent advances and applications of Statistical Associating Fluid Theory. *Industrial and Engineering Chemistry Research*, 47(21):8063–8082, 2008.

[48] C. McCabe and A. Galindo. *Applied Thermodynamics of Fluids*. Royal Society of Chemistry, 2010.

[49] J. Gross and G. Sadowski. Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules. *Industrial and Engineering Chemistry Research*, 40(4):1244–1260, 2001.

[50] T. Kraska and K. E. Gubbins. Phase equilibria calculations with a modified SAFT equation of state. 1. Pure alkanes, alkanols, and water. *Industrial and Engineering Chemistry Research*, 35(12):4727–4737, 1996.

[51] T. Kraska and K. E. Gubbins. Phase equilibria calculations with a modified SAFT equation of state. 2. Binary mixtures of n-alkanes, 1-alkanols, and water. *Industrial and Engineering Chemistry Research*, 35(12):4738–4746, 1996.

[52] F. J. Blas and L. F. Vega. Critical behavior and partial miscibility phenomena in binary mixtures of hydrocarbons by the statistical associating fluid theory. *Journal of Chemical Physics*, 109(17):7405–7413, 1998.

[53] A. Gil-Villegas, A. Galindo, P. J. Whitehead, S. J. Mills, G. Jackson, and A. N. Burgess. Statistical associating fluid theory for chain molecules with attractive potentials of variable range. *Journal of Chemical Physics*, 106(10):4168–4186, 1997.

[54] A. Galindo, L. A. Davies, A. Gil-Villegas, and G. Jackson. The thermodynamics of mixtures and the corresponding mixing rules in the SAFT-VR approach for potentials of variable range. *Molecular Physics*, 93(2):241–252, 1998.

[55] V. Papaioannou, T. Lafitte, C. Avendano, C. S. Adjiman, G. Jackson, E. A. Müller, and A. Galindo. Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments. *The Journal of Chemical Physics*, 140:054107, 2014.

[56] L. Constantinou and R. Gani. New group-contribution method for estimating properties of pure compounds. *AIChE Journal*, 40(10):1697–1710, 1994.

[57] A. Fredenslund, R. L. Jones, and J. M. Prausnitz. Group-contribution estimation of activity-coefficients in nonideal liquid-mixtures. *AIChE Journal*, 21(6):1086–1099, 1975.

[58] A. Lymperiadis, C. S. Adjiman, A. Galindo, and G. Jackson. A group contribution method for associating chain molecules based on the statistical associating fluid theory (SAFT-$\gamma$). *Journal of Chemical Physics*, 127(23), 2007.

[59] A. Lymperiadis, C. S. Adjiman, G. Jackson, and A. Galindo. A generalisation of the SAFT-$\gamma$ group contribution method for groups comprising multiple spherical segments. *Fluid Phase Equilibria*, 274(1-2):85–104, 2008.

[60] J.S. Rowlinson and F.L. Swinton. *Liquids and liquid mixtures.* Butterworths monographs in chemistry and chemical engineering. Butterworth Scientific, 1982.

[61] A. J. Haslam, A. Galindo, and G. Jackson. Prediction of binary intermolecular potential parameters for use in modelling fluid mixtures. *Fluid Phase Equilibria*, 266(1-2):105–128, 2008.

[62] S. T. Cui, H. D. Cochran, and P. T. Cummings. Vapor-liquid phase coexistence of alkane carbon dioxide and perfluoroalkane carbon dioxide mixtures. *Journal of Physical Chemistry B*, 103(21):4485–4491, 1999.

[63] J. J. Potoff, J. R. Errington, and A. Z. Panagiotopoulos. Molecular simulation of phase equilibria for mixtures of polar and non-polar components. *Molecular Physics*, 97(10):1073–1083, 1999.

[64] L. Zhang and J. I. Siepmann. Pressure dependence of the vapor-liquid-liquid phase behavior in ternary mixtures consisting of n-alkanes, n-perfluoroalkanes, and carbon dioxide. *Journal of Physical Chemistry B*, 109(7):2911–2919, 2005.

[65] J. Vrabec, J. Stoll, and H. Hasse. Molecular models of unlike interactions in fluid mixtures. *Molecular Simulation*, 31(4):215–221, 2005.

[66] T. Lafitte, C. Avendaño, V. Papaioannou, A. Galindo, C. S. Adjiman, G. Jackson, and E. A. Müller. SAFT-$\gamma$ force field for the simulation of molecular fluids: 3. Coarse-grained models of benzene and hetero-group models of n-decylbenzene. *Molecular Physics*, 110(11-12):1189–1203, 2012.

[67] C. Avendaño, T. Lafitte, C. S. Adjiman, A. Galindo, E. A. Müller, and G. Jackson. SAFT-$\gamma$ force field for the simulation of molecular fluids: 2. Coarse-grained models of greenhouse gases, refrigerants, and long alkanes. *Journal of Physical Chemistry B*, 117(9):2717–2733, 2013.

[68] E. A. Müller and G. Jackson. Force field parameters from the SAFT-$\gamma$ equation of state for use in coarse-grained molecular simulations. *Annual Review of Chemical and Biomolecular Engineering*, 5:405–427, 2014.

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

[70] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. N. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[71] N. Metropolis. The Beginning of the Monte Carlo Method. *Los Alamos Science*, 12:125–130, 1987.

[72] L. Verlet. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review*, 159:98–103, 1967.

[73] W. L. Jorgensen, J. D. Madura, and C. J. Swenson. Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of the American Chemical Society*, 106(22):6638–6646, 1984.

[74] M. G. Martin and J. I. Siepmann. Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *Journal of Physical Chemistry B*, 102(14):2569–2577, 1998.

[75] S. K. Nath, F. A. Escobedo, and J. J. de Pablo. On the simulation of vapor-liquid equilibria for alkanes. *Journal of Chemical Physics*, 108(23):9905–9911, 1998.

[76] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79(2):926–935, 1983.

[77] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[78] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries. The MARTINI force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.

[79] J. K. Johnson, J. A. Zollweg, and K. E. Gubbins. The Lennard-Jones equation of state revisited. *Molecular Physics*, 78(3):591–618, 1993.

[80] J. J. Potoff and D. A. Bernard Brunel. Mie potentials for phase equilibria calculations: Application to alkanes and perfluoroalkanes. *Journal of Physical Chemistry B*, 113(44):14725–14731, 2009.

[81] S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein. A coarse grain model for n-alkanes parameterized from surface tension data. *Journal of Chemical Physics*, 119(14):7043–7049, 2003.

[82] X. B. He, W. Shinoda, R. DeVane, and M. L. Klein. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Molecular Physics*, 108(15):2007–2020, 2010.

[83] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, P. B. Moore, and M. L. Klein. Simulations of phospholipids using a coarse grain model. *Journal of Physical Chemistry B*, 105(40):9785–9792, 2001.

[84] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. A coarse grain model for phospholipid simulations. *Journal of Physical Chemistry B*, 105(19):4464–4470, 2001.

[85] W. Shinoda, R. Devane, and M. L. Klein. Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants. *Molecular Simulation*, 33(1-2):27–36, 2007.

[86] W. Shinoda, R. DeVane, and M. L. Klein. Coarse-grained molecular modeling of non-ionic surfactant self-assembly. *Soft Matter*, 4(12):2454–2462, 2008.

[87] B. L. Bhargava and M. L. Klein. Formation of micelles in aqueous solutions of a room temperature ionic liquid: a study using coarse grained molecular dynamics. *Molecular Physics*, 107(4-6):393–401, 2009.

[88] R. DeVane, W. Shinoda, P. B. Moore, and M. L. Klein. Transferable coarse grain nonbonded interaction model for amino acids. *Journal of Chemical Theory and Computation*, 5(8):2115–2124, 2009.

[89] K. R. Hadley and C. McCabe. On the investigation of coarse-grained models for water: balancing computational efficiency and the retention of structural properties. *Journal of Physical Chemistry B*, 114(13):4590–4599, 2010.

[90] M. Levitt and A. Warshel. Computer-simulation of protein folding. *Nature*, 253(5494):694–698, 1975.

[91] M. S. Shell. *Coarse graining and multiscale techniques*. Department of Chemical Engineering, University of California Santa Barbara, 2009.

[92] E. Brini, E. A. Algaer, P. Ganguly, C. L. Li, F. Rodriguez-Ropero, and N. F. A. van der Vegt. Systematic coarse-graining methods for soft matter simulations - A review. *Soft Matter*, 9(7):2108–2119, 2013.

[93] W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *J. Phys. Chem.*, 139:090901, 2013.

[94] J. A. Barker and D. Henderson. Pair and triplet interactions in argon. *Molecular Physics*, 17(6):579–592, 1969.

[95] G. Casanova, R. J. Dulla, D. A. Jonah, J. S. Rowlinson, and G. Saville. Effective pair potentials in fluids in the presence of three-body forces. *Molecular Physics*, 18(5):589–606, 1970.

[96] M. A. van der Hoef and P. A. Madden. Three-body dispersion contributions to the thermodynamic properties and effective pair interactions in liquid argon. *Journal of Chemical Physics*, 111(4):1520–1526, 1999.

[97] V. A. Kirkwood, J. G.and Lewinson and B. J. Alder. Radial distribution functions and the equation of state of fluids composed of molecules interacting according to the Lennard-Jones potential. *Journal of Chemical Physics*, 20(6):929–938, 1952.

[98] J. S. Rowlinson. Intermolecular potentials that are functions of thermodynamic variables. *Molecular Physics*, 52(3):567–572, 1984.

[99] L. Reatto and M. Tau. Integral-equation for the structure of fluids in presence of 3-body forces - krypton, argon, and models for complex mixtures. *Journal of Chemical Physics*, 86(11):6474–6485, 1987.

[100] R. L. Henderson. Uniqueness theorem for fluid pair correlation-functions. *Physics Letters A*, A 49(3):197–198, 1974.

[101] F. M´uller-Plathe. Coarse-graining in polymer simulation: From the atomistic to the mesoscopic scale and back. *Chemphyschem*, 3(9):754–769, 2002.

[102] P. Ganguly, D. Mukherji, C. Junghans, and N. F. A. van der Vegt. Kirkwood-Buff coarse-grained force fields for aqueous solutions. *Journal of Chemical Theory and Computation*, 8(5):1802–1807, 2012.

[103] J. Baschnagel, K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Muller-Plathe, M. Murat, W. Paul, S. Santos, U. W. Suter, and V. Tries. Bridging the gap between atomistic and coarse-grained models of polymers: Status

and perspectives. *Advances in Polymer Science: Viscoelasticity, Atomistic Models, Statistical Chemistry*, 152:41–156, 2000.

[104] A. Villa, C. Peter, and N. F. A. van der Vegt. Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Physical Chemistry Chemical Physics*, 11(12):2077–2086, 2009.

[105] M. Seo, S. Rauscher, R. Pomes, and D. P. Tieleman. Improving internal peptide dynamics in the coarse-grained MARTINI model: Toward large-scale simulations of amyloid- and elastin-like peptides. *Journal of Chemical Theory and Computation*, 8(5):1774–1785, 2012.

[106] D. Reith, M. Putz, and F. Muller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry*, 24(13):1624–1636, 2003.

[107] Z. J. Wang and M. Deserno. A systematically coarse-grained solvent-free model for quantitative phospholipid bilayer simulations. *Journal of Physical Chemistry B*, 114(34):11207–11220, 2010.

[108] Multiscale Molecular Methods in Applied Chemistry. *Multiscale molecular methods in applied chemistry*, volume 307. Springer, 2011.

[109] P. Majek and R. Elber. A coarse-grained potential for fold recognition and molecular dynamics simulations of proteins. *Proteins-Structure Function and Bioinformatics*, 76(4):822–836, 2009.

[110] T. Terakawa and S. Takada. Multiscale ensemble modeling of intrinsically disordered proteins: p53 n-terminal domain. *Biophysical Journal*, 101(6):1450–1458, 2011.

[111] G. Srinivas, X. L. Cheng, and J. C. Smith. A solvent-free coarse grain model for crystalline and amorphous cellulose fibrils. *Journal of Chemical Theory and Computation*, 7(8):2539–2548, 2011.

[112] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial-distribution functions - A reverse Monte-Carlo approach. *Physical Review E*, 52(4):3730–3737, 1995.

[113] A. Lyubartsev, A. Mirzoev, L. J. Chen, and A. Laaksonen. Systematic coarse-graining of molecular models by the Newton inversion method. *Faraday Discussions*, 144:43–56, 2010.

[114] A. P. Lyubartsev. Multiscale modeling of lipids and lipid bilayers. *European Biophysics Journal with Biophysics Letters*, 35(1):53–61, 2005.

[115] A. Savelyev and G. A. Papoian. Chemically accurate coarse graining of double-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 107(47):20340–20345, 2010.

[116] M. S. Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *Journal of Chemical Physics*, 129(14), 2008.

[117] A. Chaimovich and M. S. Shell. Coarse-graining errors and numerical optimization using a relative entropy framework. *Journal of Chemical Physics*, 134(9), 2011.

[118] F. Ercolessi and J. B. Adams. Interatomic potentials from 1st principles calculations : the force-matching method. *Europhysics Letters*, 26(8):583–588, 1994.

[119] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *Journal of Chemical Physics*, 120(23):10896–10913, 2004.

[120] W. G. Noid, J. W. Chu, G. S. Ayton, and G. A. Voth. Multiscale coarse-graining and structural correlations: Connections to liquid-state theory. *Journal of Physical Chemistry B*, 111(16):4116–4127, 2007.

[121] W. G. Noid, J. W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *Journal of Chemical Physics*, 128(24), 2008.

[122] W. G. Noid, P. Liu, Y. Wang, J. W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *Journal of Chemical Physics*, 128(24), 2008.

[123] J. W. Mullinax and W. G. Noid. A generalized-Yvon-Born-Green theory for determining coarse-grained interaction potentials. *Journal of Physical Chemistry C*, 114(12):5661–5674, 2010.

[124] M. Born and M. S. Green. *A General Kinetic Theory of Liquids*. Cambridge University Press, Cambridge, 1989.

[125] G. A. Voth and S. Izvekov. A multiscale coarse-graining method for biomolecular systems. *Journal of Physical Chemistry B*, 109(7):2469–2473, 2005.

[126] P. Liu, S. Izvekov, and G. A. Voth. Coarse-grained molecular dynamics of carbohydrates using a systematic multiscale approach. *Biophysical Journal*, pages 566A–566A, 2007.

[127] P. Liu, S. Izvekov, and G. A. Voth. Multiscale coarse-graining of monosaccharides. *Journal of Physical Chemistry B*, 111(39):11566–11575, 2007.

[128] Y. T. Wang, S. Izvekov, T. Y. Yan, and G. A. Voth. Multiscale coarse-graining of ionic liquids. *Journal of Physical Chemistry B*, 110(8):3564–3575, 2006.

[129] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth. Coarse-grained peptide modeling using a systematic multiscale approach. *Biophysical Journal*, 92(12):4289–4303, 2007.

[130] V. N. Maiorov and G. M. Crippen. Contact potential that recognizes the correct folding of globular proteins. *Journal of Molecular Biology*, 227(3):876–888, 1992.

[131] B. Smit, K. Esselink, P. A. J. Hilbers, N. M. Vanos, L. A. M. Rupert, and I. Szleifer. Computer-simulations of surfactant self-assembly. *Langmuir*, 9(1):9–11, 1993.

[132] G. Jimenez-Serratos, A. Gil-Villegas, C. Vega, and F. J. Blas. Monte Carlo simulation of flexible trimers: From square well chains to amphiphilic primitive models. *Journal of Chemical Physics*, 139(11), 2013.

[133] A. J. Crane and E. A. Müller. Global phase behaviour of polyphilic tapered dendrons. *Soft Matter*, 7(16):7465–7476, 2011.

[134] L. Onsager. The effects of shape on the interaction of colloidal particles. *Annals of the New York Academy of Sciences*, 51:627–659, 1949.

[135] M. A. Cotter. Hard spherocylinders in an anisotropic mean field: A simple model for a nematic liquid crystal. *Journal of Chemical Physics*, 66:1098–1106, 1977.

[136] S. Verga and G Jackson. Study of the pitch of fluids of electrostatically chiral anisotropic molecules: Mean-field theory and simulation. *Molecular Physics*, 104:3681–3691, 2006.

[137] S. C. McGrother, A. Gil-Villegas, and G. Jackson. The liquid-crystalline phase behaviour of hard spherocylinders with terminal point dipoles. *Journal of Physics-Condensed Matter*, 8(47):9649–9655, 1996.

[138] A. Gil-Villegas, S. C. McGrother, and G. Jackson. Chain and ring structures in smectic phases of molecules with transverse dipoles. *Chemical Physics Letters*, 269(5-6):441–447, 1997.

[139] S. C. McGrother, A. Gil-Villegas, and G. Jackson. The effect of dipolar interactions on the liquid crystalline phase transitions of hard spherocylinders with central longitudinal dipoles. *Molecular Physics*, 95(3):657–673, 1998.

[140] X. Periole and S. J. Marrink. The Martini coarse-grained force field. *Methods in Molecular Biology*, 924, 2012.

[141] S. J. Marrink, A. H. de Vries, and A. E. Mark. Coarse grained model for semi-quantitative lipid simulations. *Journal of Physical Chemistry B*, 108(2):750–760, 2004.

[142] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink. The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.

[143] S. J. Marrink, A. H. de Vries, T. A. Harroun, J. Katsaras, and S. R. Wassall. Cholesterol shows preference for the interior of polyunsaturated lipid. *Journal of the American Chemical Society*, 130(1):10, 2008.

[144] W. F. D. Bennett, J. L. MacCallum, M. J. Hinner, S. J. Marrink, and D. P. Tieleman. Molecular view of cholesterol flip-flop and chemical potential in different membrane environments. *Journal of the American Chemical Society*, 131(35):12714–12720, 2009.

[145] T. Vuorela, A. Catte, P. S. Niemela, A. Hall, M. T. Hyvonen, S. J. Marrink, M. Karttunen, and I. Vattulainen. Role of lipids in spheroidal high density lipoproteins. *Plos Computational Biology*, 6(10), 2010.

[146] H. Lee, A. H. de Vries, S. J. Marrink, and R. W. Pastor. A coarse-grained model for polyethylene oxide and polyethylene glycol: Conformation and hydrodynamics. *Journal of Physical Chemistry B*, 113(40):13186–13194, 2009.

[147] H. Lee and R. W. Pastor. Coarse-grained model for PEGylated lipids: Effect of PEGylation on the size and shape of self-assembled structures. *Journal of Physical Chemistry B*, 115(24):7830–7837, 2011.

[148] A. V. Sangwai and R. Sureshkumar. Coarse-grained Molecular Dynamics simulations of the sphere to rod transition in surfactant micelles. *Langmuir*, 27:6628–6638, 2011.

[149] M.; Velinova, D.; Sengupta, A. V.; Tadjer, and S. J. Marrink. Sphere-to-rod transitions of nonionic surfactant micelles in aqueous solution modeled by Molecular Dynamics simulations. *Langmuir*, 27:14071–14077, 2011.

[150] C. A. Lopez, A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hunenberger, and S. J. Marrink. Martini coarse-grained force field: Extension to carbohydrates. *Journal of Chemical Theory and Computation*, 5(12):3195–3210, 2009.

[151] P. J. Bond and M. S. P. Sansom. Insertion and assembly of membrane proteins via simulation. *Journal of the American Chemical Society*, 128(8):2697–2704, 2006.

[152] A. Y. Shih, A. Arkhipov, P. L. Freddolino, and K. Schulten. Coarse grained protein-lipid model with application to lipoprotein particles. *Journal of Physical Chemistry B*, 110(8):3674–3684, 2006.

[153] X. Periole, T. Huber, S. J. Marrink, and T. P. Sakmar. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. *Journal of the American Chemical Society*, 129(33):10126–10132, 2007.

[154] H. Lee and R. G. Larson. Molecular Dynamics simulations of PAMAM dendrimer-induced pore formation in DPPC bilayers with a coarse-grained model. *Journal of Physical Chemistry B*, 110(37):18204–18211, 2006.

[155] H. Lee and R. G. Larson. Coarse-grained Molecular Dynamics studies of the concentration and size dependence of fifth- and seventh-generation PAMam dendrimers on pore formation in dmpc bilayer. *Journal of Physical Chemistry B*, 112(26):7778–7784, 2008.

[156] G. Rossi, L. Monticelli, S. R. Puisto, I. Vattulainen, and T. Ala-Nissila. Coarse-graining polymers with the MARTINI force-field: polystyrene as a benchmark case. *Soft Matter*, 7(2):698–708, 2011.

[157] S. Khalid, P. J. Bond, J. Holyoake, R. W. Hawtin, and M. S. P. Sansom. DNA and lipid bilayers: self-assembly and insertion. *Journal of the Royal Society Interface*, 5:S241–S250, 2008.

[158] J. Wong-Ekkabut, S. Baoukina, W. Triampo, I. M. Tang, D. P. Tieleman, and L. Monticelli. Computer simulation study of fullerene translocation through lipid membranes. *Nature Nanotechnology*, 3(6):363–368, 2008.

[159] S. W. Chiu, H. L. Scott, and E. Jakobsson. A coarse-grained model based on Morse potential for water and n-alkanes. *Journal of Chemical Theory and Computation*, 6(3):851–863, 2010.

[160] M. L. Klein and W. Shinoda. Large-scale molecular dynamics simulations of self-assembling systems. *Science*, 321(5890):798–800, 2008.

[161] A. Jusufi, S. Sanders, M. L. Klein, and A. Z. Panagiotopoulos. Implicit-solvent models for micellization: Nonionic surfactants and temperature-dependent properties. *Journal of Physical Chemistry B*, 115(5):990–1001, 2011.

[162] M. Orsi, D. Y. Haubertin, W. E. Sanderson, and J. W. Essex. A quantitative coarse-grain model for lipid bilayers. *Journal of Physical Chemistry B*, 112(3):802–815, 2008.

[163] A. Vishnyakov, M. T. Lee, and A. V. Neimark. Prediction of the critical micelle concentration of nonionic surfactants by Dissipative Particle Dynamics simulations. *Journal of Physical Chemistry Letters*, 4(5):797–802, 2013.

[164] E. A. Müller and K. E. Gubbins. An equation of state for water from a simplified intermolecular potential. *Industrial and Engineering Chemistry Research*, 34(10):3662–3673, 1995.

[165] G. Stell, J. C. Rasaiah, and H. Narang. Thermodynamic perturbation-theory for simple polar fluids. 2. *Molecular Physics*, 27(5):1393–1414, 1974.

[166] G. Stell, J. C. Rasaiah, and H. Narang. Thermodynamic perturbation-theory for simple polar fluids. 1. *Molecular Physics*, 23(2):393, 1972.

[167] A. Lofti, J. Vrabec, and J. Fischer. Vapour liquid equilibria of the Lennard-Jones fluid from the NPT plus test particle method. *Molecular Physics*, 76(6):1319–1333, 1992.

[168] J. Vrabec, J. Stoll, and H. Hasse. A set of molecular models for symmetric quadrupolar fluids. *Journal of Physical Chemistry B*, 105(48):12126–12133, 2001.

[169] J. Stoll, J. Vrabec, and H. Hasse. Comprehensive study of the vapour-liquid equilibria of the pure two-centre Lennard-Jones plus pointdipole fluid. *Fluid Phase Equilibria*, 209(1):29–53, 2003.

[170] J. Vrabec, Y. L. Huang, and H. Hasse. Molecular models for 267 binary mixtures validated by vapor-liquid equilibria: A systematic approach. *Fluid Phase Equilibria*, 279(2):120–135, 2009.

[171] C. G. Gray and K. E. Gubbins. *Theory of Molecular Fluids*, volume I. Clarendon Press, Oxford, UK, 1984.

[172] J. Gross. An equation-of-state contribution for polar components: Quadrupolar molecules. *AIChE Journal*, 51(9):2556–2568, 2005.

[173] J. Y. Cui and J. R. Elliott. Phase diagrams for a multistep potential model of n-alkanes by discontinuous molecular dynamics and thermodynamic perturbation theory. *Journal of Chemical Physics*, 116(19):8625–8631, 2002.

[174] O. Unlu, N. H. Gray, Z. N. Gerek, and J. R. Elliott. Transferable step potentials for the straight-chain alkanes, alkenes, alkynes, ethers, and alcohols. *Industrial and Engineering Chemistry Research*, 43(7):1788–1793, 2004.

[175] A. Vahid and J. R. Elliott. Transferable intermolecular potentials for carboxylic acids and their phase behavior. *AIChE Journal*, 56(2):485–505, 2010.

[176] A. F. Ghobadi and J. R. Elliott. Adapting SAFT-$\gamma$ perturbation theory to site-based molecular dynamics simulation. I. Homogeneous fluids. *Journal of Chemical Physics*, 139(23), 2013.

[177] J. A. Barker and D. Henderson. Perturbation theory and equation of state for fluids ii. a successful theory of liquids. *Journal of Chemical Physics*, 47:4714–21, 1967.

[178] J. A. Barker and D. Henderson. What is liquid - understanding states of matter. *Reviews of Modern Physics*, 48(4):587–671, 1976.

[179] T. van Westen, T. J. H. Vlugt, and J. Gross. Determining force field parameters using a physically based equation of state. *Journal of Physical Chemistry B*, 115(24):7872–7880, 2011.

[180] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM - a program for macromolecular energy, minimisation, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[181] J. I. Siepmann, S. Karaborni, and B. Smit. Simulating the critical-behavior of complex fluids. *Nature*, 365(6444):330–332, 1993.

[182] M. G. Martin. Comparison of the AMBER, CHARMM, COMPASS, GROMOS, OPLS, TraPPE and UFF force fields for prediction of vapor-liquid coexistence curves and liquid densities. *Fluid Phase Equilibria*, 248(1):50–55, 2006. 089HL Times Cited:21 Cited References Count:37.

[183] J. R. Errington and A. Z. Panagiotopoulos. A new intermolecular potential model for the n-alkane homologous series. *Journal of Physical Chemistry B*, 103(30):6314–6322, 1999.

[184] E. A. Müller and A. Mejia. Comparison of united-atom potentials for the simulation of vapor-liquid equilibria and interfacial properties of long-chain n-alkanes up to n-C-100. *Journal of Physical Chemistry B*, 115(44):12822–12834, 2011.

[185] H. Okumura and F. Yonezawa. Liquid-vapor coexistence curves of several interatomic model potentials. *Journal of Chemical Physics*, 113(20):9162–9168, 2000.

[186] E. A. Müller and K. E. Gubbins. Simulation of hard triatomic and tetratomic molecules - a test of associating fluid theories. *Molecular Physics*, 80(4):957–976, 1993.

[187] S. Rahman, O. Lobanova, C. C. Braga, V. Raptis, E. A. Müller, A. Galindo, and G Jackson. SAFT-$\gamma$ Mie approach for the simulation of alkanes with coarse-grained force fields for the intra- and inter- molecular interactions. *in preparation*, 2014.

[188] W. G. Hoover. Canonical dynamics - equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[189] S. Nose. A unified formulation of the constant temperature Molecular Dynamics methods. *Journal of Chemical Physics*, 81(1):511–519, 1984.

[190] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005.

[191] G. Jackson and K. E. Gubbins. Mixtures of associating spherical and chain molecules. *Pure and Applied Chemistry*, 61(6):1021–1026, 1989.

[192] A. L. Archer, M. D. Amos, G. Jackson, and I. A. McLure. The theoretical prediction of the critical points of alkanes, perfluoroalkanes, and their mixtures using bonded hard-sphere (BHS) theory. *International Journal of Thermophysics*, 17(1):201–211, 1996.

[193] J. M. H. L. Sengers. Mean-field theories, their weaknesses and strength. *Fluid Phase Equilibria*, 158:3–17, 1999.

[194] E. Forte, F. Llovell, L. F. Vega, J. P. M. Trusler, and A. Galindo. Application of a renormalization-group treatment to the statistical associating fluid theory for potentials of variable range (SAFT-VR). *Journal of Chemical Physics*, 134(15), 2011.

[195] E. W. Lemmon, M. O. McLinden, and D. G. Friend. *Thermophysical Properties of Fluid Systems in NIST Chemistry WebBook, NIST Standard Reference Database Number 69 Eds. Linstrom, P.J.; Mallard, W.G. National Institute of Standards and Technology, Gaithersburg MD, 20899*, retrieved 2013. http://webbook.nist.gov.

[196] DECHEMA. *Gesellschaft für Chemische Technik und Biotechnologie e.V Frankfurt am Main, Germany*, retrieved 2013. https://cdsdt.dl.ac.uk/detherm.

[197] R. D. Dunlap, R. D. Bedford, J. C. Woodbrey, and S. D. Furrow. Liquid-vapor equilibrium for the system perfluoro - n-hexane - n-hexane. *Journal of the American Chemical Society*, 81(12):2927–2930, 1959.

[198] P. Morgado, J. Ben Lewis, C. M. C. Laginhas, L. F. G. Martins, C. McCabe, F. J. Blas, and E. J. M. Filipe. Systems involving hydrogenated and fluorinated chains: Volumetric properties of perfluoroalkanes and perfluoroalkylalkane surfactants. *Journal of Physical Chemistry B*, 115(50):15013–15023, 2011.

[199] P. J. R. Morgado. *Semifluorinated alkanes; Structure - properties relations.* PhD thesis, Universidade técnica de Lisboa, 2011.

[200] C. L. Sassen, A. Gonzalez Casielles, T. W. de Loos, and J. de Swaan Arons. The influence of pressure and temperature on the phase behaviour of the system H2O + C12 + C7E5 and relevant binary subsystems. *Fluid Phase Equilibria*, 72:173–187, 1991.

[201] A. Ben-Naim. *Molecular Theory of Water and Aqueous Solutions: Understanding water.* Number pt. 1 in Molecular Theory of Water and Aqueous Solutions. World Scientific, 2009.

[202] F. Paesani and G. A. Voth. The Properties of Water: Insights from Quantum Simulations. *Journal of Physical Chemistry B*, 113(17):5702–5719, 2009.

[203] M. D. Baer, C. J. Mundy, M. J. McGrath, I. F. W. Kuo, J. I. Siepmann, and D. J. Tobias. Re-examining the properties of the aqueous vapor-liquid interface using dispersion corrected density functional theory. *Journal of Chemical Physics*, 135(12), 2011.

[204] B. Santra, J. Klimes, A. Tkatchenko, D. Alfe, B. Slater, A. Michaelides, R. Car, and M. Scheffler. On the accuracy of van der Waals inclusive density-functional theory exchange-correlation functionals for ice at ambient and high pressures. *Journal of Chemical Physics*, 139(15), 2013.

[205] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones. What ice can teach us about water interactions: A critical comparison of the performance of different water models. *Faraday Discussions*, 141:251–276, 2009.

[206] I. Shvab and R. J. Sadus. Intermolecular potentials and the accurate prediction of the thermodynamic properties of water. *Journal of Chemical Physics*, 139(19), 2013.

[207] K. A. Dill, T. M. Truskett, V. Vlachy, and B. Hribar-Lee. Modeling water, the hydrophobic effect, and ion solvation. *Annual Review of Biophysics and Biomolecular Structure*, 34:173–199, 2005.

[208] M. S. Shell and A. Chaimovich. Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy. *Physical Chemistry Chemical Physics*, 11(12):1901–1915, 2009.

[209] K. R. Hadley and C. McCabe. Coarse-grained molecular models of water: a review. *Molecular Simulation*, 38(8-9):671–681, 2012.

[210] C. Knight and G. A. Voth. Coarse-graining away electronic structure: a rigorous route to accurate condensed phase interaction potentials. *Molecular Physics*, 110(9-10):935–944, 2012.

[211] A. Striolo. From interfacial water to macroscopic observables: A review. *Adsorption Science and Technology*, 29(3):211–258, 2011.

[212] S. L. Meadley and C. A. Angell. Water and its relatives: the stable, supercooled and particularly the stretched, regimes. *Proceedings of the International School of Physics "Enrico Fermi" Course CLXXXVII*, Water: Fundamentals as the basis for Understanding the Environment and Promoting Technology, 2014.

[213] O. Matsuoka, E. Clementi, and M. Yoshimine. Cl study of water dimer potential surface. *Journal of Chemical Physics*, 64(4):1351–1361, 1976.

[214] R. Car and M. Parrinello. Unified approach for Molecular-Dynamics and Density-Functional Theory. *Physical Review Letters*, 55(22):2471–2474, 1985.

[215] M. J. McGrath, J. I. Siepmann, I. F. W. Kuo, and C. J. Mundy. Vapor-liquid equilibria of water from first principles: comparison of density functionals and basis sets. *Molecular Physics*, 104(22-24):3619–3626, 2006.

[216] E. E. Santiso and K. E. Gubbins. Multi-scale molecular modeling of chemical reactivity. *Molecular Simulation*, 30(11-12):699–748, 2004.

[217] P. G. Karamertzanis, P. Raiteri, and A. Galindo. The use of anisotropic potentials in modeling water and free energies of hydration. *Journal of Chemical Theory and Computation*, 6(5):1590–1607, 2010.

[218] J. D. Bernal and R. H. Fowler. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *Journal of Chemical Physics*, 1:515, 1933.

[219] J. S. Rowlinson. The second virial coefficients of polar gases. *Transactions of the Faraday Society*, 45:974–984, 1949.

[220] J. S. Rowlinson. The lattice energy of ice and the second virial coefficient of water vapour. *Transactions of the Faraday Society*, 47:120–129, 1951.

[221] J. A. Pople. Molecular association in liquids. II. A theory of the structure of water. *Proceedings of the Royal Society A*, 205:163, 1951.

[222] N. Bjerrum. *Kgl. Danske Videnskab., Mat.-fis. Medd.*, 27:1, 1951.

[223] J. A. Barker and R. O. Watts. Structure of Water; a Monte Carlo Calculation. *Chemical Physics Letters*, 3(3):144–145, 1969.

[224] A. Ben-Naim and F. H. Stillinger. *Structure and Transport Processes in Water and Aqueous Solutions*. Aspects of the Statistical Mechanical Theory of Water. Wiley, 1972.

[225] F. H. Stilling and A. Rahman. Improved simulation of liquid water by Molecular-Dynamics. *Journal of Chemical Physics*, 60(4):1545–1557, 1974.

[226] F. H. Stillinger. *Theory and Molecular Models for Water*, volume 31 of *Advances in Chem. Phys.: Non-simple Liquids*. Wiley and Sons, Inc., New York, 1975.

[227] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, and J. Hermans. Interaction models for water in relation to protein hydration. In Bernard Pullman, editor, *Intermolecular Forces*, volume 14 of *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*, pages 331–342. Springer Netherlands, 1981.

[228] W. L. Jorgensen. Quantum and statistical mechanical studies of liquids .10. transferable intermolecular potential functions for water, alcohols, and ethers - application to liquid water. *Journal of the American Chemical Society*, 103(2):335–340, 1981.

[229] W. L. Jorgensen. Quantum and statistical mechanical studies of liquids .24. Revised TIPs for simulations of liquid water and aqueous-solutions. *Journal of Chemical Physics*, 77(8):4156–4163, 1982.

[230] M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *Journal of Chemical Physics*, 112(20):8910–8922, 2000.

[231] S. W. Rick. A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums. *Journal of Chemical Physics*, 120(13):6085–6093, 2004.

[232] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *Journal of Chemical Physics*, 120(20):9665–9678, 2004.

[233] J. L. F. Abascal, E. Sanz, R. G. Fernandez, and C. Vega. A potential model for the study of ices and amorphous water: TIP4P/Ice. *Journal of Chemical Physics*, 122(23), 2005.

[234] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *Journal of Chemical Physics*, 123(23), 2005.

[235] C. Vega and J. L. F. Abascal. Simulating water with rigid non-polarizable models: A general perspective. *Physical Chemistry Chemical Physics*, 13(44):19663–19688, 2011.

[236] Y. L. Huang, T. Merker, M. Heilig, H. Hasse, and J. Vrabec. Molecular modeling and simulation of vapor-liquid equilibria of ethylene oxide, ethylene glycol, and water as well as their binary mixtures. *Industrial and Engineering Chemistry Research*, 51(21):7428–7440, 2012.

[237] J. R. Errington and A. Z. Panagiotopoulos. Phase equilibria of the modified buckingham exponential-6 potential from hamiltonian scaling grand canonical monte carlo. *Journal of Chemical Physics*, 109(3):1093–1100, 1998.

[238] B. Guillot and Y. Guissani. How to build a better pair potential for water. *Journal of Chemical Physics*, 114(15):6720–6733, 2001.

[239] E. Sanz, C. Vega, J. L. F. Abascal, and L. G. MacDowell. Tracing the phase diagram of the four-site water potential (TIP4P). *Journal of Chemical Physics*, 121(2):1165–1166, 2004.

[240] E. Sanz, C. Vega, J. L. F. Abascal, and L. G. MacDowell. Phase diagram of water from computer simulation. *Physical Review Letters*, 92(25):255701–1–255101–4, 2004.

[241] P. G. Kusalik and I. M. Svishchev. The spatial structure in liquid water. *Science*, 265(5176):1219–1221, 1994.

[242] P. Jedlovszky and J. Richardi. Comparison of different water models from ambient to supercritical conditions: A Monte Carlo simulation and molecular Ornstein-Zernike study. *Journal of Chemical Physics*, 110(16):8019–8031, 1999.

[243] P. T. Kiss and A. Baranyai. Sources of the deficiencies in the popular SPC/E and TIP3P models of water. *Journal of Chemical Physics*, 134(5), 2011.

[244] A. Baranyai, A. Bartok, and A. A. Chialvo. On the performance of simple planar models of water in the vapor and the ice phases. *Physics and Chemistry of Ice*, pages 109–115, 2007.

[245] F. H. Stillinger and C. W. David. Polarization model for water and its ionic dissociation products. *Journal of Chemical Physics*, 69(4):1473–1484, 1978.

[246] P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn. Cooperative effects in simulated water. *Nature*, 282(5738):459–464, 1979.

[247] A. A. Chialvo and P. T. Cummings. Engineering a simple polarizable model for the molecular simulation of water applicable over wide ranges of state conditions. *Journal of Chemical Physics*, 105(18):8274–8281, 1996.

[248] P. Paricaud, M. Predota, A. A. Chialvo, and P. T. Cummings. From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a Gaussian charge polarizable model. *Journal of Chemical Physics*, 122(24), 2005.

[249] P. Kiss, M. Darvas, A. Baranyai, and P. Jedlovszky. Surface properties of the polarizable baranyai-kiss water model. *Journal of Chemical Physics*, 136(11), 2012.

[250] P. T. Kiss, P. Bertsyk, and A. Baranyai. Testing recent charge-on-spring type polarizable water models. I. Melting temperature and ice properties. *Journal of Chemical Physics*, 137(19), 2012.

[251] P. T.; Kiss and A. Baranyai. A systematic development of a polarizable potential of water. *The Journal of Chemical Physics*, 138(204507), 2013.

[252] S. W. Rick, S. J. Stuart, and B. J. Berne. Dynamical fluctuating charge force-fields - Application to liquid water. *Journal of Chemical Physics*, 101(7):6141–6156, 1994.

[253] B. Chen, J. H. Xing, and J. I. Siepmann. Development of polarizable water force fields for phase equilibrium calculations. *Journal of Physical Chemistry B*, 104(10):2391–2401, 2000.

[254] M. Sprik and M. L. Klein. A polarizable model for water using distributed charge sites. *Journal of Chemical Physics*, 89(12):7556–7560, 1988.

[255] I. M. Svishchev, P. G. Kusalik, J. Wang, and R. J. Boyd. Polarizable point-charge model for water: Results under normal and extreme conditions. *Journal of Chemical Physics*, 105(11):4742–4750, 1996.

[256] P. Y. Ren and J. W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *Journal of Physical Chemistry B*, 107(24):5933–5947, 2003.

[257] P. Y. Ren and J. W. Ponder. Temperature and pressure dependence of the amoeba water model. *Journal of Physical Chemistry B*, 108(35):13427–13437, 2004.

[258] J. W. Ponder, C. J. Wu, P. Y. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon. Current status of the AMOEBA polarizable force field. *Journal of Physical Chemistry B*, 114(8):2549–2564, 2010.

[259] P. J. van Maaren and D. van der Spoel. Molecular Dynamics simulations of water with novel shell-model potentials. *Journal of Physical Chemistry B*, 105(13):2618–2626, 2001.

[260] P. Bopp, G. Jancso, and K. Heinzinger. An improved potential for non-rigid water-molecules in the liquid-phase. *Chemical Physics Letters*, 98(2):129–133, 1983.

[261] Y. J. Wu, H. L. Tepper, and G. A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *Journal of Chemical Physics*, 124(2), 2006.

[262] G. Raabe and R. J. Sadus. Molecular dynamics simulation of the dielectric constant of water: The effect of bond flexibility. *Journal of Chemical Physics*, 134(23), 2011.

[263] I. Nezbeda. Simple short-ranged models of water and their application. a review. *Journal of Molecular Liquids*, 73-4:317–336, 1997.

[264] A. Ben-Naim. Statistical Mechanics of "water-like" particles in Two Dimensions, Part I: Interaction between two identical non-polar solute particles. *Journal of Chemical Physics*, 54:3682–2695, 1971.

[265] K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill. A simple model of water and the hydrophobic effect. *Journal of the American Chemical Society*, 120(13):3166–3175, 1998.

[266] T. M. Truskett, P. G. Debenedetti, S. Sastry, and S. Torquato. A single-bond approach to orientation-dependent interactions and its implications for liquid water. *Journal of Chemical Physics*, 111(6):2647–2656, 1999.

[267] E. A. Jagla. Phase behavior of a system of particles with core collapse. *Physical Review E*, 58(2):1478–1486, 1998.

[268] E. A. Jagla. Core-softened potentials and the anomalous properties of water. *Journal of Chemical Physics*, 111(19):8980–8986, 1999.

[269] G. Franzese, G. Malescio, A. Skibinsky, S. V. Buldyrev, and H. E. Stanley. Generic mechanism for generating a liquid-liquid phase transition. *Nature*, 409(6821):692–695, 2001.

[270] Z. Y. Yan, S. V. Buldyrev, N. Giovambattista, and H. E. Stanley. Structural order for one-scale and two-scale potentials. *Physical Review Letters*, 95(13), 2005.

[271] S. V. Buldyrev, P. Kumar, P. G. Debenedetti, P. J. Rossky, and H. E. Stanley. Water-like solvation thermodynamics in a spherically symmetric solvent model with two characteristic lengths. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51):20177–20182, 2007.

[272] P. H. Poole, F. Sciortino, T. Grande, H. E. Stanley, and C. A. Angell. Effect of hydrogen-bonds on the thermodynamic behavior of liquid water. *Physical Review Letters*, 73(12):1632–1635, 1994.

[273] A. B. de Oliveira, P. A. Netz, T. Colla, and M. C. Barbosa. Structural anomalies for a three dimensional isotropic core-softened potential. *Journal of Chemical Physics*, 125(12), 2006.

[274] A. B. de Oliveira, P. A. Netz, T. Colla, and M. C. Barbosa. Thermodynamic and dynamic anomalies for a three-dimensional isotropic core-softened potential. *Journal of Chemical Physics*, 124(8), 2006.

[275] Z. Y. Yan, S. V. Buldyrev, N. Giovambattista, and H. E. Stanley. Structural order for one-scale and two-scale potentials. *Physical Review Letters*, 95(13), 2005.

[276] A. Perera, A. Rispe, L. Zoranic, R. Mazighi, and F. Sokolic. Water-like structure with repulsive double-core interactions. *Molecular Physics*, 107(13):1349–1353, 2009.

[277] E. Salcedo, A. B. de Oliveira, N. M. Barraz, C. Chakravarty, and M. C. Barbosa. Core-softened fluids, water-like anomalies, and the liquid-liquid critical points. *Journal of Chemical Physics*, 135(4), 2011.

[278] T. Urbic. Two-dimensional core-softened model with water like properties. study by thermodynamic perturbation theory. *Condensed Matter Physics*, 16(4), 2013.

[279] M. Hus and T. Urbic. Core-softened fluids as a model for water and the hydrophobic effect. *Journal of Chemical Physics*, 139(11), 2013.

[280] M. Hus and T. Urbic. The hydrophobic effect in a simple isotropic water-like model: Monte carlo study. *Journal of Chemical Physics*, 140(14), 2014.

[281] T. Headgordon and F. H. Stillinger. An orientational perturbation-theory for pure liquid water. *Journal of Chemical Physics*, 98(4):3313–3327, 1993.

[282] V. Molinero and W. A. Goddard. M3B: A coarse grain force field for molecular simulations of malto-oligosaccharides and their water mixtures. *Journal of Physical Chemistry B*, 108(4):1414–1427, 2004.

[283] T. Head-Gordon, M. E. Johnson, and A. A. Louis. Representability problems for coarse-grained water potentials. *Journal of Chemical Physics*, 126(14):144509, 2007.

[284] S. Izvekov, J. M. J. Swanson, and G. A. Voth. Coarse-graining in interaction space: A systematic approach for replacing long-range electrostatics with short-range potentials. *Journal of Physical Chemistry B*, 112(15):4711–4724, 2008.

[285] H. Wang, C. Junghans, and K. Kremer. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *European Physical Journal E*, 28(2):221–229, 2009.

[286] V. Molinero and E. B. Moore. Water modeled as an intermediate element between carbon and silicon. *Journal of Physical Chemistry B*, 113(13):4008–4016, 2009.

[287] M. Winger, D. Trzesniak, R. Baron, and W. F. van Gunsteren. On using a too large integration time step in molecular dynamics simulations of coarse-grained molecular models. *Physical Chemistry Chemical Physics*, 11(12):1934–1941, 2009.

[288] M. U. Hammer, T. H. Anderson, A. Chaimovich, M. S. Shell, and J. Israelachvili. The search for the hydrophobic force law. *Faraday Discussions*, 146:299–308, 2010.

[289] Z. Wu, Q. A. Cui, and A. Yethiraj. A new coarse-grained model for water: The importance of electrostatic interactions. *Journal of Physical Chemistry B*, 114(32):10524–10529, 2010.

[290] L. Darre, M. R. Machado, P. D. Dans, F. E. Herrera, and S. Pantano. Another coarse grain model for aqueous solvation: WAT FOUR? *Journal of Chemical Theory and Computation*, 6(12):3793–3807, 2010.

[291] S. Riniker and W. F. Gunsteren. A simple, efficient polarizable coarse-grained water model for Molecular Dynamics simulations. *Journal of Chemical Physics*, 134(8):084110, 2011.

[292] W. F. D. Bennett and D. P. Tieleman. Water defect and pore formation in atomistic and coarse-grained lipid membranes: Pushing the limits of coarse graining. *Journal of Chemical Theory and Computation*, 7(9):2981–2988, 2011.

[293] L. H. Gao and W. H. Fang. Semi-bottom-up coarse graining of water based on microscopic simulations. *Journal of Chemical Physics*, 135(18), 2011.

[294] B. van Hoof, A. J. Markvoort, R. A. van Santen, and P. A. J. Hilbers. The CUMULUS coarse graining method: Transferable potentials for water and solutes. *Journal of Physical Chemistry B*, 115(33):10001–10012, 2011.

[295] S. Y. Mashayak and N. R. Aluru. Thermodynamic state-dependent structure-based coarse-graining of confined water. *Journal of Chemical Physics*, 137(21), 2012.

[296] H. C. Gonzalez, L. Darre, and S. Pantano. Transferable mixing of atomistic and coarse-grained water models. *Journal of Physical Chemistry B*, 117(46):14438–14448, 2013.

[297] F. Zipoli, T. Laino, S. Stolz, E. Martin, C. Winkelmann, and A. Curioni. Improved coarse-grained model for molecular-dynamics simulations of water nucleation. *Journal of Chemical Physics*, 139(9), 2013.

[298] J. C. Fogarty, S. W. Chiu, P. Kirby, E. Jakobsson, and S. A. Pandit. Automated optimization of water-water interaction parameters for a coarse-grained model. *Journal of Physical Chemistry B*, 118(6):1603–1611, 2014.

[299] M. Orsi. Comparative assessment of the ELBA coarse-grained model for water. *Molecular Physics*, 112(11):1566–1576, 2014.

[300] L. C. Jacobson, R. M. Kirby, and V. Molinero. How short is too short for the interactions of a water potential? Exploring the parameter space of a coarse-grained water model using uncertainty quantification. *The Journal of Physical Chemistry B*, 118:8190–8202, 2014.

[301] W. Smith. DL POLY-applications to molecular simulation II. *Molecular Simulation*, 32(12-13):933–933, 2006.

[302] J. G.; Kirkwood and F. P. Buff. The statistical mechanical theory of surface tension. *Journal of Chemical Physics*, 17:338–343, 1949.

[303] A. Trokhymchuk and J. Alejandre. Computer simulations of liquid/vapor interface in Lennard-Jones fluids: Some questions and answers. *Journal of Chemical Physics*, 111(18):8510–8523, 1999.

[304] G. Jackson, G. J. Gloor, F. J. Blas, and E. de Miguel. Test-area simulation method for the direct determination of the interfacial tension of systems with continuous or discontinuous potentials. *Journal of Chemical Physics*, 123(13), 2005.

[305] R. W. Zwanzig. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *Journal of Chemical Physics*, 22(8):1420–1426, 1954.

[306] H. Wang, C. Schutte, and P. W. Zhang. Error estimate of short-range force calculation in inhomogeneous molecular systems. *Physical Review E*, 86(2), 2012.

[307] N. Ramrattan. *Simulation and theoretical perspectives of the phase behaviour of solids, liquids, and gases using the Mie family of intermolecular potential.* PhD thesis, Imperial College London, 2013.

[308] G. J. Gloor, G. Jackson, F. J. Blas, E. M. del Rio, and E. de Miguel. An accurate density functional theory for the vapor-liquid interface of associating chain molecules based on the statistical associating fluid theory for potentials of variable range. *Journal of Chemical Physics*, 121(24):12740–12759, 2004.

[309] G. J. Gloor, G. Jackson, F. J. Blas, E. M. del Rio, and E. de Miguel. Prediction of the vapor-liquid interfacial tension of nonassociating and associating fluids with the saft-vr density functional theory. *Journal of Physical Chemistry C*, 111(43):15513–15522, 2007.

[310] F. Llovell, N. Mac Dowell, F. J. Bias, A. Galindo, and G. Jackson. Application of the saft-vr density functional theory to the prediction of the interfacial properties of mixtures of relevance to reservoir engineering. *Fluid Phase Equilibria*, 336:137–150, 2012.

[311] C. Herdes, O. Lobanova, E. E. Santiso, G. Jackson, J. Eastoe, and E. A. Müller. A molecular framework for modelling the surface tensions and surface excess of diluted surfactant solutions. *in preparation*, 2014.

[312] J. Hermans, H. J. C. Berendsen, W. F. Vangunsteren, and J. P. M. Postma. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23(8):1513–1518, 1984.

[313] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen. A systematic study of water models for molecular simulation: Derivation of water models optimized for use with a reaction field. *Journal of Chemical Physics*, 108(24):10220–10230, 1998.

[314] G. C. Boulougouris, I. G. Economou, and D. N. Theodorou. Engineering a molecular model for water phase equilibrium over a wide temperature range. *Journal of Physical Chemistry B*, 102(6):1029–1035, 1998.

[315] C. Vega, J. L. F. Abascal, and I. Nezbeda. Vapor-liquid equilibria from the triple point up to the critical point for the new generation of TIP4P-like models: TIP4P/Ew, TIP4P/2005, and TIP4P/ice. *Journal of Chemical Physics*, 125(3), 2006.

[316] N. F. A. van der Vegt and B. Hess. Hydration thermodynamic properties of amino acid analogues: A systematic comparison of biomolecular force fields and water models. *Journal of Physical Chemistry B*, 110(35):17616–17626, 2006.

[317] P. T. Kiss and A. Baranyai. Clusters of classical water models. *Journal of Chemical Physics*, 131(20), 2009.

[318] C. H. Twu, K. E. Gubbins, and C. G. Gray. Thermodynamics of mixtures of non-spherical molecules .3. Fluid phase-equilibria and critical loci. *Journal of Chemical Physics*, 64(12):5186–5197, 1976.

[319] K. E. Gubbins and C. H. Twu. Thermodynamics of polyatomic fluid mixtures .1. Theory. *Chemical Engineering Science*, 33(7):863–878, 1978.

[320] J. van der Steen, T. W. Deloos, and J. D. Arons. The volumetric analysis and prediction of liquid-liquid-vapor equilibria in certain carbon dioxide + n-alkane systems. *Fluid Phase Equilibria*, 51:353–367, 1989.

[321] G. Schneider, Z. Alwani, W. Heim, E. Horvath, and E. U. Franck. Phase equilibria and critical appearances in binary mixtures up to 1500 bar. *Chemie Ingenieur Technik*, 39(11):649, 1967.

[322] G. M. Schneider. Phase equilibria in fluid mixtures at high pressure. *Advances in Physical Chemistry*, 17(1), 1970.

[323] S. Takenouchi and J. Kennedy. The binary system H2O-CO2 at high temperatures and pressures. *American Journal of Science*, 262:1055–1074, 1964.

[324] K. Todheide and E. U. Franck. Das Zweiphasengebiet und die kritische Kurve im System Kohlendioxid-Wasser bis zu Drucken von 3500 bar. *Zeitschrift fuer Physikalische Chemie Neue Folge*, 37:387–401, 1963.

[325] E. Brunner. Fluid mixtures at high-pressures .9. phase-separation and critical phenomena in 23 (normal-alkane + water) mixtures. *Journal of Chemical Thermodynamics*, 22(4):335–353, 1990.

[326] R. L. Stevenson, D. S. Labracio, T. A. Beaton, and M. C. Thies. Fluid-phase equilibria and critical phenomena for the dodecane-water and squalane-water systems at elevated-temperatures and pressures. *Fluid Phase Equilibria*, 93:317–336, 1994.

[327] M. Cismondi, S. B. Rodriguez-Reartes, J. M. Milanesio, and M. S. Zabaloy. Phase equilibria of CO2+n-alkane binary systems in wide ranges of conditions: Development of predictive correlations based on cubic mixing rules. *Industrial and Engineering Chemistry Research*, 51(17):6232–6250, 2012.

[328] A. Galindo and F. J. Blas. Theoretical examination of the global fluid phase behavior and critical phenomena in carbon dioxide plus n-alkane binary mixtures. *Journal of Physical Chemistry B*, 106(17):4503–4515, 2002.

[329] F. J. Blas and A. Galindo. Study of the high pressure phase behaviour of CO2+n-alkane mixtures using the SAFT-VR approach with transferable parameters. *Fluid Phase Equilibria*, 194:501–509, 2002.

[330] J. Garcia, L. Lugo, and J. Fernandez. Phase equilibria, PVT behavior, and critical phenomena in carbon dioxide plus n-alkane mixtures using the perturbed-chain statistical associating fluid theory approach. *Industrial and Engineering Chemistry Research*, 43(26):8345–8353, 2004.

[331] A. Valtz, A. Chapoy, C. Coquelet, P. Paricaud, and D. Richon. Vapour-liquid equilibria in the carbon dioxide-water system, measurement and modelling from 278.2 to 318.2 K. *Fluid Phase Equilibria*, 226:333–344, 2004.

[332] M. C. dos Ramos, F. J. Blas, and A. Galindo. Phase equilibria, excess properties, and Henry's constants of the water plus carbon dioxide binary mixtures. *Journal of Physical Chemistry C*, 111(43):15924–15934, 2007.

[333] M. C. dos Ramos, F. J. Blas, and A. Galindo. Modelling the phase equilibria and excess properties of the water plus carbon dioxide binary mixture. *Fluid Phase Equilibria*, 261(1-2):359–365, 2007.

[334] N. Mac Dowell, F. Llovell, C. S. Adjiman, G. Jackson, and A. Galindo. Modeling the fluid phase behavior of carbon dioxide in aqueous solutions of monoethanolamine using transferable parameters with the SAFT-VR approach. *Industrial and Engineering Chemistry Research*, 49(4):1883–1899, 2010.

[335] F. Llovell and L. F. Vega. Global fluid phase equilibria and critical phenomena of selected mixtures using the crossover soft-SAFT equation. *Journal of Physical Chemistry B*, 110(3):1350–1362, 2006.

[336] E. Forte. *Measurement and Prediction of the Phase Behaviour of Carbon Dioxide, Alkane and Water Mixtures at Reservoir Conditions.* Phd thesis, Imperial College London, 2011.

[337] B. H. Patel, P. Paricaud, A. Galindo, and G. C. Maitland. Prediction of the salting-out effect of strong electrolytes on water plus alkane solutions. *Industrial and Engineering Chemistry Research*, 42(16):3809–3823, 2003.

[338] L. F. Vega, F. Llovell, and F. J. Blas. Capturing the solubility minima of n-alkanes in water by soft-SAFT. *Journal of Physical Chemistry B*, 113(21):7621–7630, 2009.

[339] A. Galindo, P. J. Whitehead, G. Jackson, and A. N. Burgess. Predicting the high-pressure phase equilibria of water plus n-alkanes using a simplified SAFT theory with transferable intermolecular interaction parameters. *Journal of Physical Chemistry*, 100(16):6781–6792, 1996.

[340] I. V. Yakoumis, G. M. Kontogeorgis, E. C. Voutsas, E. M. Hendriks, and D. P. Tassios. Prediction of phase equilibria in binary aqueous systems containing alkanes, cycloalkanes, and alkenes with the cubic-plus-association equation of state. *Industrial and Engineering Chemistry Research*, 37(10):4175–4182, 1998.

[341] E. C. Voutsas, G. C. Boulougouris, I. G. Economou, and D. P. Tassios. Water/hydrocarbon phase equilibria using the thermodynamic perturbation theory. *Industrial and Engineering Chemistry Research*, 39(3):797–804, 2000.

[342] X. S. Li and P. Englezos. Vapor-liquid equilibrium of systems containing alcohols, water, carbon dioxide and hydrocarbons using SAFT. *Fluid Phase Equilibria*, 224(1):111–118, 2004.

[343] M. B. Oliveira, J. A. P. Coutinho, and A. J. Queimada. Mutual solubilities of hydrocarbons and water with the CPA EOS. *Fluid Phase Equilibria*, 258(1):58–66, 2007.

[344] C. P. Emborsky, K. R. Cox, and W. G. Chapman. Correlation and prediction of water content in alkanes using a molecular theory. *Industrial and Engineering Chemistry Research*, 50(13):7791–7799, 2011.

[345] V. Papaioannou, C. S. Adjiman, G. Jackson, and A. Galindo. Simultaneous prediction of vapour-liquid and liquid-liquid equilibria (VLE and LLE) of aqueous mixtures with the SAFT-gamma group contribution approach. *Fluid Phase Equilibria*, 306(1):82–96, 2011.

[346] G. M. Kontogeorgis, M. L. Michelsen, G. K. Folas, S. Derawi, N. von Solms, and E. H. Stenby. Ten years with the CPA (Cubic-Plus-Association) equation of state. Part 2. Cross-associating and multicomponent systems. *Industrial and Engineering Chemistry Research*, 45(14):4869–4878, 2006.

[347] J. J. Potoff and J. I. Siepmann. Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen. *AIChE Journal*, 47(7):1676–1682, 2001.

[348] E. Johansson, K. Bolton, D. N. Theodorou, and P. Ahlstrom. Monte Carlo simulations of equilibrium solubilities and structure of water in n-alkanes and polyethylene. *Journal of Chemical Physics*, 126(22), 2007.

[349] A. L. Ferguson, P. G. Debenedetti, and A. Z. Panagiotopoulos. Solubility and molecular conformations of n-alkane chains in water. *Journal of Physical Chemistry B*, 113(18):6405–6414, 2009.

[350] Y. Liu, A. Z. Panagiotopoulos, and P. G. Debenedetti. Monte Carlo simulations of high-pressure phase equilibria of CO(2)-H(2)O mixtures. *Journal of Physical Chemistry B*, 115(20):6629–6635, 2011.

[351] C. G. Aimoli, E. J. Maginn, and C. R. A. Abreu. Force field comparison and thermodynamic property calculation of supercritical CO2 and CH4 using molecular dynamics simulations. *Fluid Phase Equilibria*, 368:80–90, 2014.

[352] R. W. Wiebe and V. L. Gaddy. The solubility of carbon dioxide in water at various temperatures from 12 to 40 ° and at pressures to 500 atmospheres: Critical phenomena. *Journal of the American Chemical Society*, 62:815–817, 1940.

[353] R. W. Wiebe and V. L. Gaddy. Vapor phase composition of carbon dioxide-water mixtures at various temperatures and at pressures to 700 atmospheres. *Journal of the American Chemical Society*, 63:475–477, 1941.

[354] M. B. King, A. Mubarak, J. D. Kim, and T. R. Bott. The mutual solubilities of water with supercritical and liquid carbon-dioxide. *Journal of Supercritical Fluids*, 5(4):296–302, 1992.

[355] C. R. Coan and A. D. King. Solubility of water in compressed carbon dioxide, nitrous oxide, and ethane - evidence for hydration of carbon dioxide and nitrous oxide in gas phase. *Journal of the American Chemical Society*, 93(8):1857, 1971.

[356] K. Y. Song and R. Kobayashi. Water content of CO2 in equilibrium with liquid water and/or hydrates. *SPE Formation Evaluation*, 2:500–508, 1987.

[357] P. C. Gillepsie and G. M. Wilson. Vapor-liquid and liquid- liquid equilibria: Water-methane, water-carbon dioxide, water- hydrogen sulfide, water-npentane, water-methane-npentane. *Gas Processors Association*, Tulsa:Research Report RR–48, 1982.

[358] N. Spycher, K. Pruess, and J. Ennis-King. CO2-H2O mixtures in the geological sequestration of CO2. I. Assessment and calculation of mutual solubilities from 12 to 100 degrees C and up to 600 bar. *Geochimica Et Cosmochimica Acta*, 67(16):3015–3031, 2003.

[359] A.Yu. Namiot, V.G. Skripka, and Yu.G. Lotter. Phase equilibria in hydrocarbon - water systems at high temperatures. *Deposited Doc. VINITI*, 76:1213, 1976.

[360] R. G. Sultanov and V. G. Skripka. Solubility of water in n-alkanes at elevated temperatures and pressures. *Deposited Doc. VINITI*, 72:4386, 1972.

[361] R. G. Sultanov and V. G. Skripka. The solubility of water in n-hexane, cyclohexane and benzene at elevated temperatures and pressures. *Deposited Doc. VINITI*, 73:5347, 1973.

[362] J. G. Liu, Z. F. Qin, G. F. Wang, X. L. Hou, and J. G. Wang. Critical properties of binary and ternary mixtures of hexane plus methanol, hexane plus carbon dioxide, methanol plus carbon dioxide, and hexane plus carbon dioxide plus methanol. *Journal of Chemical and Engineering Data*, 48(6):1610–1613, 2003.

[363] H. Tanaka, Y. Yamaki, and M. Kato. Solubility of carbon-dioxide in pentadecane, hexadecane, and pentadecane plus hexadecane. *Journal of Chemical and Engineering Data*, 38(3):386–388, 1993.

[364] E. Forte, A. Galindo, and J. P. M. Trusler. Experimental and Molecular Modeling Study of the Three-Phase Behavior of (n-Decane plus Carbon Dioxide plus Water)

at Reservoir Conditions. *Journal of Physical Chemistry B*, 115(49):14591–14609, 2011.

[365] C. Tanford. *The hydrophobic effect: formation of micelles and biological membranes.* Wiley Interscience Publications. Wiley, 1980.

[366] M.J. Rosen and J. T. Kunjappu. *Surfactants and Interfacial Phenomena.* Wiley, 2012.

[367] R. Goetz and R. Lipowsky. Computer simulations of bilayer membranes: Self-assembly and interfacial tension. *Journal of Chemical Physics*, 108(17):7397–7409, 1998.

[368] S. Fujiwara, T. Itoh, M. Hashimoto, and R. Horiuchi. Molecular Dynamics simulation of amphiphilic molecules in solution: Micelle formation and dynamic coexistence. *Journal of Chemical Physics*, 130(14):144901, 2009.

[369] A. S. Keys, C. R. Iacovella, and S. C. Glotzer. Characterizing complex particle morphologies through shape matching: Descriptors, applications, and algorithms. *Journal of Computational Physics*, 230(17):6438–6463, 2011.

[370] H. A. Karimi-Varzaneh, N. F. A. van der Vegt, F. Müller-Plathe, and P. Carbone. How good are coarse-grained polymer models? A comparison for atactic polystyrene. *Chemphyschem*, 13(15):3428–3439, 2012.

[371] P. Christine, L. Delle Site, and K. Kremer. Classical simulations from the atomistic to the mesoscale and back: coarse graining an azobenzene liquid crystal. *Soft Matter*, 4:859–869, 2008.

[372] G. Megariotis, A. Vyrkou, A. Leygue, and D. N. Theodorou. Systematic coarse graining of 4-cyano-4'-pentylbiphenyl. *Industrial and Engineering Chemistry Research*, 50(2):546–556, 2011.

[373] S. Izvekov and G. A. Voth. Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *Journal of Chemical Theory and Computation*, 2(3):637–648, 2006.

[374] Q. Shi, S. Izvekov, and G. A. Voth. Mixed atomistic and coarse-grained molecular dynamics: Simulation of a membrane-bound ion channel. *Journal of Physical Chemistry B*, 110(31):15045–15048, 2006.

[375] W. Shinoda, R. DeVane, and M. L. Klein. Zwitterionic lipid assemblies: Molecular Dynamics studies of monolayers, bilayers, and vesicles using a new coarse grain force field. *Journal of Physical Chemistry B*, 114(20):6836–6849, 2010.

[376] B. L. Bhargava and M. L. Klein. Formation of micelles in aqueous solutions of a room temperature ionic liquid: a study using coarse grained molecular dynamics. *Molecular Physics*, 107(4-6):393–401, 2009.

[377] C. C. Chen. Molecular thermodynamic model for Gibbs energy of mixing of nonionic surfactant solutions. *AIChE Journal*, 42(11):3231–3240, 1996.

[378] C. C. Chen. A segment-based local composition model for the Gibbs energy of polymer-solutions. *Fluid Phase Equilibria*, 83:301–312, 1993.

[379] X. S. Li, J. F. Lu, Y. G. Li, and J. C. Liu. Studies on UNIQUAC and SAFT equations for nonionic surfactant solutions. *Fluid Phase Equilibria*, 153(2):215–229, 1998.

[380] X. S. Li, J. F. Lu, and Y. G. Li. Study on ionic surfactant solutions by SAFT equation incorporated with MSA. *Fluid Phase Equilibria*, 168(1):107–123, 2000.

[381] B. L. Larsen, P. Rasmussen, and A. Fredenslund. A modified UNIFAC group-contribution model for prediction of phase-equilibria and heats of mixing. *Industrial and Engineering Chemistry Research*, 26(11):2274–2286, 1987.

[382] E. C. Voutsas, M. V. Flores, N. Spiliotis, G. Bell, P. J. Halling, and D. P. Tassios. Prediction of critical micelle concentrations of nonionic surfactants in aqueous and nonaqueous solvents with UNIFAC. *Industrial and Engineering Chemistry Research*, 40(10):2362–2366, 2001.

[383] M. V. Flores, E. C. Voutsas, N. Spiliotis, G. M. Eccleston, G. Bell, D. P. Tassios, and P. J. Halling. Critical micelle concentrations of nonionic surfactants in organic solvents: Approximate prediction with UNIFAC. *Journal of Colloid and Interface Science*, 240(1):277–283, 2001.

[384] D. S. Abrams and J. M. Prausnitz. Statistical thermodynamics of liquid-mixtures - New expression for excess Gibbs energy of partly or completely miscible systems. *AIChE Journal*, 21(1):116–128, 1975.

[385] E. S. J. Rudolph, J. H. Langeveld, T. W. de Loos, and J. D. Arons. Phase behaviour of water plus nonionic surfactant systems: experiments and modelling applying the Peng-Robinson equation of state with the Wong-Sandler mixing rules and the UNIQUAC g(E)-model. *Fluid Phase Equilibria*, 173(1):81–96, 2000.

[386] M. N. Garcia-Lisbona, A. Galindo, G. Jackson, and A. N. Burgess. Predicting the high-pressure phase equilibria of binary aqueous solutions of 1-butanol, n-butoxyethanol and n-decylpentaoxyethylene ether (C10E5) using the SAFT-HS approach. *Molecular Physics*, 93(1):57–71, 1998.

[387] M. N. Garcia-Lisbona, A. Galindo, G. Jackson, and A. N. Burgess. An examination of the cloud curves of liquid-liquid immiscibility in aqueous solutions of alkyl polyoxyethylene surfactants using the SAFT-HS approach with transferable parameters. *Journal of the American Chemical Society*, 120(17):4191–4199, 1998.

[388] G. N. I. Clark, A. Galindo, G. Jackson, S. Rogers, and A. N. Burgess. Modeling and understanding closed-loop liquid - liquid immiscibility in aqueous solutions of poly(ethylene glycol) using the SAFT-VR approach with transferable parameters. *Macromolecules*, 41(17):6582–6595, 2008.

[389] S. O. Nielsen, C. F. Lopez, G. Srinivas, and M. L. Klein. Coarse grain models and the computer simulation of soft materials. *Journal of Physics-Condensed Matter*, 16(15):R481–R512, 2004.

[390] S. Zeppieri, J. Rodriguez, and A. L. L. de Ramos. Interfacial tension of alkane plus water systems. *Journal of Chemical and Engineering Data*, 46(5):1086–1088, 2001.

[391] R. Pool and P. G. Bolhuis. Accurate free energies of micelle formation. *Journal of Physical Chemistry B*, 109(14):6650–6657, 2005.

[392] R. Pool and P. G. Bolhuis. Prediction of an autocatalytic replication mechanism for micelle formation. *Physical Review Letters*, 97(1):018302, 2006.

[393] M. Sammalkorpi, M. Karttunen, and M. Haataja. Micelle fission through surface instability and formation of an interdigitating stalk. *Journal of the American Chemical Society*, 130(52):17977–17980, 2008.

[394] D. N. LeBard, B. G. Levine, R. DeVane, W. Shinoda, and M. L. Klein. Premicelles and monomer exchange in aqueous surfactant solutions above and below the critical micelle concentration. *Chemical Physics Letters*, 522:38–42, 2012.

[395] B. Andersson and G. Olofsson. Calorimetric study of non-ionic surfactants - enthalpies and heat-capacity changes for micelle formation in water of C8E4 and Triton X-100 and micelle size of C8E4. *Journal of the Chemical Society-Faraday Transactions I*, 84(11):4087–4095, 1988.

[396] M. Zulauf, K. Weckstrom, J. B. Hayter, V. Degiorgio, and M. Corti. Neutron-scattering study of micelle structure in isotropic aqueous-solutions of poly(oxyethylene) amphiphiles. *Journal of Physical Chemistry*, 89(15):3411–3417, 1985.

[397] Balmbra R. R., Clunie J. S., J. M. Corkill, and Goodman J. F. Variations in the micelle size of non-ionic detergents. *Transactions of the Faraday Society*, 60:979–985, 1964.

[398] K. Imanishi and Y. Einaga. Effects of hydrophilic chain length on the characteristics of the micelles of pentaoxyethylene n-decyl C10E5 and hexaoxyethylene n-decyl C10E6 ethers. *Journal of Physical Chemistry B*, 109(15):7574–7581, 2005.

[399] R. K. Mahajan and D. Nandni. Micellization and phase behavior of binary mixtures of anionic and nonionic surfactants in aqueous media. *Industrial and Engineering Chemistry Research*, 51(8):3338–3349, 2012.

[400] S. Y. Kwon and M. W. Kim. Structure and growth control of a nonionic surfactant micelle by adding a phospholipid. *Langmuir*, 17(26):8016–8023, 2001.

[401] A. Jusufi. Molecular simulations of self-assembly processes of amphiphiles in dilute solutions: The challenge for quantitative modelling. *Molecular Physics*, 111(21):3182–3192, 2013.

[402] A. G. Daful, V. A. Baulin, J. B. Avalos, and A. D. Mackie. Accurate critical micelle concentrations from a microscopic surfactant model. *Journal of Physical Chemistry B*, 115(13):3434–3443, 2011.

[403] D.R. Karsa and Royal Society of Chemistry (Great Britain). *Industrial applications of surfactants IV*. Number v. 4 in Royal Society of Chemistry: Special publication. Royal Society of Chemistry, 1999.

[404] T. Sottmann, R. Strey, and S. H. Chen. A small-angle neutron scattering study of nonionic surfactant molecules at the water-oil interface: Area per molecule, microemulsion domain size, and rigidity. *Journal of Chemical Physics*, 106(15):6483–6491, 1997.

[405] U. R. M. Kjellin, P. M. Claesson, and P. Linse. Surface properties of tetra(ethylene oxide) dodecyl amide compared with poly(ethylene oxide) surfactants. 1. Effect of the headgroup on adsorption. *Langmuir*, 18(18):6745–6753, 2002.

[406] L. Liggieri, M. Ferrari, A. Massa, and F. Ravera. Molecular reorientation in the adsorption of some CiEj at the water-air interface. *Colloids and Surfaces a-Physicochemical and Engineering Aspects*, 156(1-3):455–463, 1999.

[407] J. Eastoe, J. S. Dalton, P. G. A. Rogueda, E. R. Crooks, A. R. Pitt, and E. A. Simister. Dynamic surface tensions of nonionic surfactant solutions. *Journal of Colloid and Interface Science*, 188(2), 1997.

[408] P. V. Coveney and J. A. D. Wattis. Analysis of a generalized Becker-Döring model of self-reproducing micelles. *Proc. R. Soc. Lond. A*, 452:2079–2102, 1996.

[409] J. A. D. Wattis, C. D. Bolton, and P. V. Coveney. The Becker-Döring equations with exponentially size-dependent rate coeffients. *Proc. R. Soc. London, Ser. A*, 1996.

[410] S. J. Marrink, D. P. Tieleman, and A. E. Mark. Molecular dynamics simulation of the kinetics of spontaneous micelle formation. *Journal of Physical Chemistry B*, 104(51):12165–12173, 2000.

[411] M. Jorge. Molecular Dynamics simulation of self-assembly of n-decyltrimethylammonium bromide micelles. *Langmuir*, 24(11):5714–5725, 2008.

[412] R. Eppenga and D. Frenkel. Monte Carlo study of the isotropic and nematic phases of infinitely thin hard platelets. *Molecular Physics*, 52(6):1303–1334, 1984.

[413] E. M. Curtis and C. K. Hall. Molecular Dynamics Simulations of DPPC Bilayers Using "LIME", a New Coarse-Grained Model. *Journal of Physical Chemistry B*, 117(17):5019–5030, 2013.

[414] A. G. Fournial, Y. Zhu, V. Molinier, G. Vermeersch, J. M. Aubry, and N. Azaroual. Aqueous phase behavior of tetraethylene glycol decanoyl ester (C9COE4) and ether (C10E4) investigated by nuclear magnetic resonance spectroscopic techniques. *Langmuir*, 23(23):11443–11450, 2007.

[415] C. Stubenrauch, S. Burauer, R. Strey, and C. Schmidt. A new approach to lamellar phases (L-alpha) in water - non-ionic surfactant systems. *Liquid Crystals*, 31(1):39–53, 2004.

[416] G. N. I. Clark, A. J. Haslam, A. Galindo, and G. Jackson. Developing optimal Wertheim-like models of water for use in Statistical Associating Fluid Theory (SAFT) and related approaches. *Molecular Physics*, 104(22-24):3561–3581, 2006.

[417] G. C. Shearman, A. I. I. Tyler, N. J. Brooks, R. H. Templer, O. Ces, R. V. Law, and J. M. Seddon. Ordered micellar and inverse micellar lyotropic phases. *Liquid Crystals*, 37(6-7):679–694, 2010.

[418] L. J. Chen, S. Y. Lin, C. C. Huang, and E. M. Chen. Temperature dependence of critical micelle concentration of polyoxyethylenated non-ionic surfactants. *Colloids and Surfaces a-Physicochemical and Engineering Aspects*, 135(1-3):175–181, 1998.

[419] R. M. Hill. Superspreading. *Current Opinion in Colloid & Interface Science*, 3(3):247–254, 1998.

[420] F. M. Menger and C. A. Littau. Gemini surfactants - a new class of self-assembling molecules. *Journal of the American Chemical Society*, 115(22):10083–10090, 1993.

[421] R. Zana. Dimeric (gemini) surfactants: Effect of the spacer group on the association behavior in aqueous solution. *Journal of Colloid and Interface Science*, 248(2):203–220, 2002.

[422] W. Shinoda, R. DeVane, and M. L. Klein. Coarse-grained force field for ionic surfactants. *Soft Matter*, 7(13):6178–6186, 2011.

[423] P. A. Winsor. Hydrotropy, solubilisation and related emulsification processes. *Transactions of the Faraday Society*, 44:376–398, 1948.

[424] E. G. Schwarz and W. G. Reid. Surface-active agents-their behavior and industrial use. *Industrial & Engineering Chemistry*, 56(9):26–31, 1964.

[425] W.H. Press. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, 2007.

[426] T. Boublik. Hard-sphere equation of state. *Journal of Chemical Physics*, 53(1):471, 1970.

[427] G. A. Mansoori, N. F. Carnahan, K. E. Starling, and T. W. Leland. Equilibrium thermodynamic properties of mixture of hard spheres. *Journal of Chemical Physics*, 54(4):1523, 1971.

[428] I. Nezbeda and G. A. Iglesias-Silva. Primitive Model of Water. 3. Analytic theoretical results with anomalies for the thermodynamic properties. *Molecular Physics*, 69(4):767–774, 1990.

[429] B. J. Zhang. Calculating thermodynamic properties from perturbation theory - I. An analytic representation of square-well potential hard-sphere perturbation theory. *Fluid Phase Equilibria*, 154(1):1–10, 1999.

[430] T. Boublik. Background correlation-functions in the hard-sphere systems. *Molecular Physics*, 59(4):775–793, 1986.

[431] R. W. Hochney. The potential calculation and some applications. *Methods of Computational Physics*, 9:136–211, 1970.

[432] A. Georgiadis, F. Llovell, A. Bismarck, F. J. Blas, A. Galindo, G. C. Maitland, J. P. M. Trusler, and G. Jackson. Interfacial tension measurements and modelling of (carbon dioxide plus n-alkane) and (carbon dioxide plus water) binary mixtures at elevated pressures and temperatures. *Journal of Supercritical Fluids*, 55(2):743–754, 2010.